Mental Health Detection Using Social Media Data

P.Sreedevi [1*], Selva Kumar S [2*], Siddharath S [3*] , Sivakesavan C [4*]

[1*] Assistant Professor, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.
[2*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India
[3*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.
[4*] UG Scholar, Department of Computer Science and Engineering, V.S.B College of Engineering Technical campus, kinathukadavu, Coimbatore, Tamilnadu, India.

***Abstract*—** People can now freely express their feelings, thoughts, and everyday experiences online thanks to the growing use of social media platforms. These user-generated texts frequently include insightful psychological clues that can aid in the early detection of mental health issues like stress, anxiety, and depression. However, informal language, emojis, slang expressions, and abbreviations make it difficult to analyze social media content and lessen the efficacy of conventional text analysis techniques. This study suggests an intelligent mental health detection system that uses social media data to address this problem. Natural Language Processing (NLP) preprocessing techniques, such as text cleaning, tokenization, stop-word removal, emoji normalization, and slang word replacement, are used in the Python implementation of the suggested system. Psychological indicators are obtained by extracting textual and emotional features from the processed data. After that, user posts are categorized into mental health categories like depression, anxiety, stress, and normal condition using machine learning and deep learning models. Standard metrics like accuracy, precision, recall, and F1-score are used to assess the model's performance. According to experimental results, the suggested method outperforms simple text classification techniques in terms of prediction performance and successfully identifies mental health risks from informal social media text. By effectively evaluating vast amounts of data, the system facilitates ongoing monitoring and offers early warning signs of possible mental health issues. In order to help families, counselors, and the medical professionals provide prompt intervention and preventive care, the suggested framework can be incorporated into mental health support systems. The potential of intelligent social media analysis as a scalable and affordable way to raise awareness of mental health issues and facilitate early detection in contemporary digital environments is generally highlighted by this work**.**

Keywords— Depression Detection, Machine Learning, Deep Learning, Natural Language Processing, Social Media Analysis, Mental Health Detection.

How to cite this article: Sreedevi P, Selva Kumar S, Siddharath S, Sivakesavan C. Mental health detection using social media data. Int J Drug Deliv Technol. 2026;16(74s): 2043-2048. DOI: 10.25258/ijddt.16.74s.164.

INTRODUCTION

Mental health disorders such as depression, anxiety, and stress are among the major concerns facing the world, affecting millions of lives, including daily activities, productivity, and quality of life [1]. In recent times, the popularity of social media sites such as Twitter, Facebook, Instagram, and online forums has increased rapidly. This has provided an opportunity to understand human behaviour and emotional expressions through digital communication [2], [3]. Most users express their feelings, opinions, and personal experiences on these sites, which are likely to have psychological cues to identify mental health disorders [4],In light of the growing availability of digital data, the possibility of automatically analyzing social media content and identifying mental health conditions has emerged as a significant field of study in both healthcare and data science communities [7]. Typically, the conventional approach to the diagnosis of mental health conditions is through interviews and psychological questionnaires completed by mental health professionals [8]. Although these practices are reliable, it is a time-consuming and expert-intensive task that requires the voluntary participation of individuals, which may lead to delays in early detection and treatment of mental health conditions [9]. In recent times, a number of researchers have sought to explore the possibility of developing automated systems that can analyze behavioral patterns to support early mental health detection and monitoring [10].

With the improvement of Natural Language Processing (NLP) and machine learning technology, researchers have used several techniques to process the text data collected from social media platforms [11]. Initially, researchers focused on sentiment analysis techniques to classify text data into positive, negative, and neutral sentiment using the emotional polarity of text [12]. However, sentiment analysis is not sufficient to identify the complex psychological patterns present in text data because people use casual words, abbreviations, emojis, and slang in their social media posts [13], [14].

To improve the accuracy of text analysis, various machine learning models have been proposed for processing social media data. Word representation techniques such as Word2Vec and GloVe have been widely used to transform textual information into numerical vectors that preserve

semantic relationships between words [15], [16]. These techniques have significantly improved the performance of machine learning algorithms in text classification and sentiment analysis tasks. Furthermore, deep learning models such as Long Short-Term Memory (LSTM) networks have demonstrated promising results in analyzing sequential text data and identifying emotional patterns in user posts [17].

Along with textual analysis, researchers have also explored different techniques for identifying mental health patterns from online behavioral data. Topic modeling approaches have been used to detect depression-related language patterns in social media posts [18]. Similarly, studies have shown that linguistic patterns extracted from Facebook posts can predict depression among users when compared with medical records [19]. Online communities and social discussion platforms have also been analyzed to identify indicators of mental illness severity among users

information with visual features obtained from social media images [21], [22].

Although significant progress has been made, several challenges still exist in accurately identifying mental health conditions using social media data. Social media texts usually comprise informal language, symbols, and abbreviations that make the analysis process difficult for traditional machine learning algorithms [23]. In addition, many studies focus on individual classification approaches rather than combining multiple techniques to improve classification accuracy [24]. These limitations highlight the need for advanced preprocessing techniques and intelligent machine learning models for effective analysis.

Many existing approaches also face challenges in handling noisy and unstructured social media data. The presence of mixed languages, emojis, and irregular sentence structures reduces the performance of conventional models. Furthermore, current systems often fail to capture deep contextual meaning and long-term dependencies in textual data. Therefore, there is a need for robust hybrid models that combine preprocessing techniques with deep learning methods. These approaches can improve the overall performance and reliability of mental health detection systems [25].

The rest of this paper is organized as follows. Section II discusses the related work on mental health detection using machine learning techniques. Section III explains the proposed methodology and system design. Section IV discusses the experimental results and performance analysis. Finally, Section V concludes the paper and suggests future work.

II. RELATED WORKS

De Choudhury et al. investigated the use of social media data for predicting depression by analyzing linguistic patterns and behavioral signals present in user posts. Their study demonstrated that emotional expressions and online activity patterns can reveal important psychological indicators of depression. The research showed that analyzing user-generated textual content can help identify early mental health conditions among individuals [1]. However, the study mainly focuses on behavioral analysis and does not incorporate advanced machine learning techniques for improving prediction accuracy [2].

Pang et al. introduced opinion mining and sentiment analysis techniques for identifying emotional polarity in textual data. Their research focused on classifying text into positive, negative, or neutral sentiments to understand user emotions in online communication. These methods became widely used for analyzing emotional expressions in social media posts and discussion forums [3]. However, sentiment analysis methods often fail to capture deeper psychological meanings and complex mental health indicators in user-generated content [4].

Orabi et al. developed deep learning-based approaches for identifying depression in Twitter users by examining textual patterns in social media posts. Their method leveraged neural network models to learn complex emotional and linguistic features associated with depressive behavior. The study demonstrated that deep learning techniques can automatically extract meaningful patterns from large-scale textual data without manual feature engineering. These models were ableto capture contextual and sequential relationships in user posts, leading to improved classification performance compared to traditional machine learning methods [5]. However, these approaches depend heavily on large annotated datasets and may face challenges in handling noisy, informal, and unstructured social media text, which can affect the overall prediction accuracy [6].

Manivannan et al. explored the application of deep learning techniques combined with big data analytics for sentiment classification in textual datasets. Their research demonstrated that deep learning models can analyze large-scale textual data and identify emotional patterns effectively [7]. However, the study mainly focuses on sentiment classification and does not specifically address mental health condition detection in social media data [8].

Mikolov et al. introduced the Word2Vec model for generating vector representations of words based on contextual relationships within text data. This method significantly improved text representation in natural language processing tasks and enabled machine learning algorithms to understand semantic relationships between words [8]. However, Word2Vec models may struggle to capture global word relationships in large datasets and may not fully represent contextual information [9].

Pathak et al. investigated machine learning techniques for analyzing user feedback data and improving system performance. Their research demonstrated that machine learning algorithms can effectively extract meaningful insights from large datasets and improve decision-making processes [10]. However, the study mainly focuses on system optimization and does not explore the application of these techniques for detecting mental health conditions from textual data [11].

Hochreiter et al. introduced Long Short-Term Memory (LSTM) networks for analyzing sequential data in deep learning applications. LSTM models can capture long-term dependencies within text sequences and identify contextual relationships between words [11]. However, traditional LSTM networks may face limitations when processing very long text sequences or complex contextual information present in social media posts [12].

Devlin et al. proposed the Bidirectional Encoder Representations from Transformers (BERT) model for improving natural language understanding tasks. BERT analyzes contextual relationships between words by processing text in both directions and provides more accurate

language representations [13]. However, the model requires significant computational resources and large datasets for training, which may limit its application in resource-constrained environments [14].

Resnik et al. applied supervised topic modeling techniques to identify depression-related language patterns in social media posts. Their research analyzed linguistic features and emotional indicators associated with psychological distress [16]. However, topic modeling techniques may fail to capture complex contextual relationships between words present in informal social media language [17].

Chancellor et al. studied online communities to analyze mental illness severity using computational techniques. Their work focused on identifying behavioral and linguistic patterns in user discussions related to psychological distress [18]. However, these approaches mainly rely on textual features and do not incorporate multimodal information such as images or behavioral activity patterns [19].

Shen et al. proposed a multimodal framework for detecting depression by combining textual features with machine learning techniques. Their research demonstrated that integrating multiple data sources can improve the accuracy of depression detection systems [19]. However, the approach requires complex data preprocessing and may be difficult to implement in real-world applications [20].

Preotiuc-Pietro et al. analyzed emotional valence and arousal patterns in Facebook posts to understand psychological conditions of users. Their research demonstrated that emotional patterns present in online posts can reveal mental health indicators [21]. However, these models mainly focus on emotional analysis and do not integrate advanced deep learning architectures for improved prediction accuracy [22].

Reece et al. analyzed Instagram images to identify visual indicators associated with depression among users. Their study demonstrated that visual features such as brightness, color tone, and image composition can reveal emotional states of individuals [23]. However, the approach relies heavily on visual data and may not capture textual expressions of psychological conditions [24].

Cao et al. conducted a systematic review of machine learning techniques used for detecting mental illness using social media data. Their study summarized various approaches including natural language processing, deep learning, and sentiment analysis methods used in psychological analysis. However, many existing systems still face challenges in accurately analyzing informal language, slang, and emojis commonly present in social media text [25]. Although considerable progress has been achieved in the early detection of mental health conditions by utilizing social media data, there are a number of challenges associated with the accurate analysis of informal textual data. The social media data includes a number of slang terms, abbreviations, and emojis, which makes it a noisy data set. Therefore, it is necessary to apply intelligent deep learning techniques to improve the accuracy of mental health condition detection systems.

III. PROPOSED METHODOLOGY

The proposed system can detect mental health conditions by using the text data available on social media platforms. The proposed system can combine Natural Language Processing, machine learning, and deep learning techniques to analyze the user-generated content on social media platforms to detect mental health conditions. The overall workflow of the proposed mental health detection system can be described by the following stages: data collection, text pre-processing, feature extraction, classification, and performance evaluation.

The overall working process of the proposed mental health detection system is depicted in Fig. 1

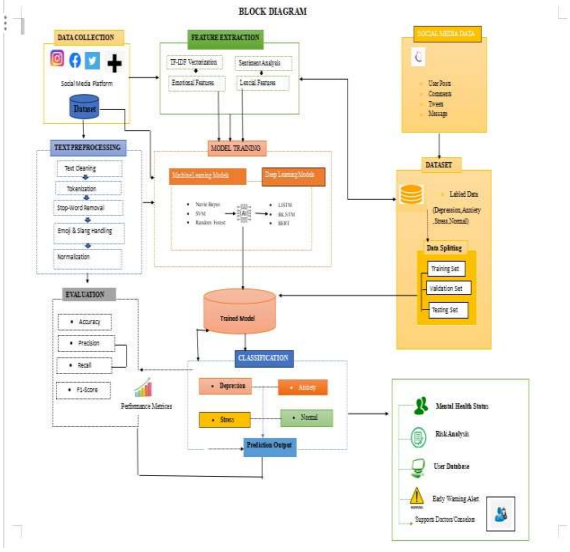

Fig. 1. System Architecture of Mental Health Detection Using Social Media Data

A. Data Collection

The first phase of the proposed system includes the collection of data in the form of text from different social media platforms. Data such as users' posts, tweets, comments, messages, etc., are collected from different sources such as Twitter, Reddit, etc., or from publicly available datasets on Kaggle. These datasets include emotional expressions that can result in mental health conditions. The collected dataset is then categorized into different categories such as depression, anxiety, stress, normal condition, etc.

B. Text Preprocessing

Generally, the text in social media is noisy and unstructured in nature. Thus, a series of operations, known as text preprocessing, is carried out on the text data to clean it prior to analysis. The operations involved in the preprocessing stage are cleaning, tokenization, removal of stop words, normalization, and processing emojis and slang words.

C. Feature Extraction

Once the preprocessing is done, important features are extracted using Natural Language Processing techniques. Techniques used include TF-IDF vectorization, sentiment analysis, and feature extraction. These features are used to represent the user posts in a numerical form. They contain emotional content.

D. Model Training

In this stage, the machine learning algorithms, as well as deep learning algorithms, can be applied for training the model. The algorithms that can be applied for training the model include Naïve Bayes, Support Vector Machine, Random Forest, LSTM, BiLSTM, BERT, etc. The data can be split into training, testing, and validation sets for better performance of the model.

E. Classification

The model, upon being trained, is capable of classifying he given text into different mental health states, such as depression, anxiety, stress, or a normal emotional state. The classification is done based on the patterns learned during the training process, and the mental health state is predicted accordingly.

F. Performance Evaluation

The performance of the proposed system is measured by using metrics such as accuracy, precision, recall, and F1-score. These metrics measure how well the model performs in detecting mental health conditions using text data obtained from social media.

MENTAL HEALTH DETECTION ALGORITHM

1. Collect social media text data from platforms such as Twitter, Reddit, or publicly available datasets from Kaggle.
2. Perform text preprocessing by removing URLs, punctuation marks, and irrelevant symbols from the collected data.
3. Apply tokenization to split the text into individual words or tokens.
4. Remove stop-words that do not contribute meaningful information for analysis.
5. Normalize the text by converting emojis and slang expressions into standard textual representations.
6. Extract features from the processed text using Natural Language Processing techniques such as TF-IDF and sentiment analysis.
7. Divide the dataset into training and testing sets for model training and validation.
8. Train machine learning and deep learning models such as Naïve Bayes, Support Vector Machine (SVM), Random Forest, LSTM, BiLSTM, and BERT.
9. Use the trained model to classify the input text into mental health categories such as depression, anxiety, stress, or normal emotional state.
10. Evaluate the model performance using evaluation metrics such as accuracy, precision, recall, and F1-score.

IV. PERFORMANCE ANALYSIS

The performance of the proposed mental health detection system is evaluated using various classification metrics. The evaluation metrics for the proposed mental health detection system help determine how effectively the model can identify mental health conditions from the social media text data. The dataset has been divided into training, testing, and validation sets to evaluate the performance of the proposed machine learning and deep learning models. The evaluation of the proposed models has been performed to determine the accuracy and reliability of the predictions. These evaluation measures provide a comprehensive understanding of the model's effectiveness in detecting mental health conditions from social media data.

A. Accuracy

Accuracy is the measure of the overall correctness of the model's predictions. Accuracy is the ratio of the number of correct predictions made by the model over the total number of instances in the dataset.

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (1)$$

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

The higher the accuracy of the model, the better it is at predicting mental health conditions.

B. Precision

Precision represents the number of correctly predicted positive data points divided by the total number of data points correctly predicted to be positive.

$$\text{Precision} = TP / (TP + FP) \quad (2)$$

Where,

- TP – True Positive
- FP – False Positive

High precision implies that the model has fewer false positives.

C. Recall

Recall, also known as sensitivity, is a measure that calculates the ratio of actual positive instances that are correctly classified by the model.

$$\text{Recall} = TP / (TP + FN) \quad (3)$$

Where,

- TP = True Positive
- FN = False Negative

D. 1-Score

The F1-score is the harmonic mean of precision and recall. It gives a balanced view of the performance of the model.

$$\text{F1 Score} = 2 * (\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

E. Model Accuracy Comparison

Machine learning and deep learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), Random Forest, Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and BERT are used to analyze the social media text dataset for detecting mental health conditions. These algorithms learn patterns from the textual features extracted through natural language processing techniques. Traditional machine learning algorithms mainly rely on statistical features and word frequency for classification. In contrast, deep learning models can capture contextual relationships and sequential dependencies in text data. Fig. 2 presents the accuracy comparison of these algorithms, where

deep learning models demonstrate higher accuracy and better performance than traditional machine learning algorithms. This demonstrates the effectiveness of deep learning techniques in analyzing complex social media text data for mental health detection.

TABLE I.I Comparative Analysis

Evaluation Parameter	Traditional Approach	Proposed System
Accuracy	Limited accuracy due to feature-based methods	Improved accuracy using deep learning models
Precision	Less precise in identifying mental health patterns	Better precision with contextual understanding
Recall	May miss important psychological signals	Higher recall by capturing emotional patterns
F1-Score	Moderate balance between precision and recall	Better balance due to advanced models

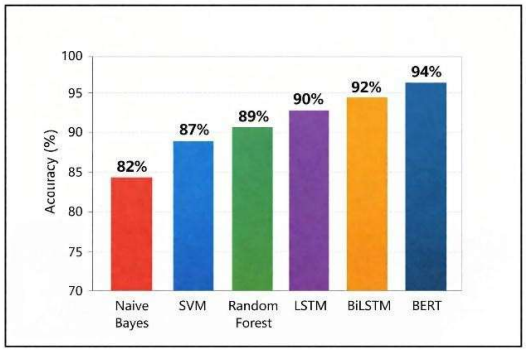

Fig. 2. Accuracy Comparison of Machine Learning and Deep Learning Models

F. Performance Evaluation

The proposed mental health detection system is evaluated using precision, recall, and F1-score. Fig. 3 presents the performance results of the model. The results show that the model achieves high precision, recall, and F1-score values, indicating effective detection of mental health conditions from social media text data.

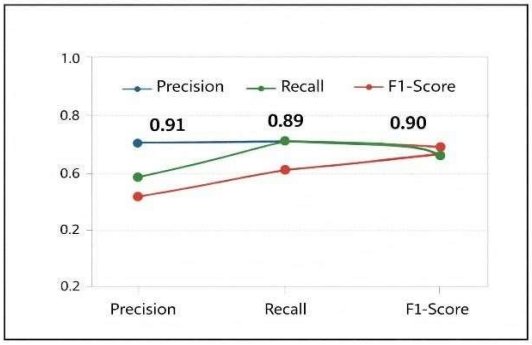

Fig. 3. Performance Evaluation Metrics of the Proposed ModelPrecision, recall, and F1-score are used to further evaluate the proposed system. Fig. 4 shows the precision and recall values. The results indicate balanced performance in detecting mental health issues from social media data.

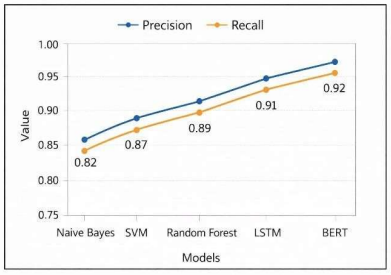

Fig. 4. Precision and Recall Comparison of the Proposed Model.

V. CONCLUSION

In this paper,a mental health detection system has been proposed by utilizing the information available in the social media domain. The proposed system analyzes the user-generated text available in the social media domain by utilizing Natural Language Processing techniques along with machine learning and deep learning algorithms to detect mental health conditions such as depression, anxiety, stress, and normal emotional states. The proposed framework has different phases such as data collection, text pre-processing, feature extraction, model building, classification, and performance evaluation. Different machine learning as well as deep learning models can be utilized to analyze the text-based information to predict the mental health state of the user. The experimental results show that the proposed deep learning models such as LSTM, BiLSTM, and BERT perform better in terms of accuracy when compared to other machine learning models. From the evaluation metrics such as accuracy, precision, recall, and F1-score, it can be concluded that the proposed system can effectively detect mental health patterns using social media text. This system can be used as a tool for the early identification of mental health issues and can be helpful for healthcare professionals. In future work, this proposed system can be improved by incorporating larger datasets, multilingual text analysis, and advanced deep learning models to further enhance prediction accuracy and reliability.

REFERENCES

[1] De Choudhury, Munmun., "Predicting depression via social media." Proceedings of the international AAAI conference on web and social media, Vol. 7. No. 1, 2013.

[2] Guntuku, Sharath Chandra.,"Understanding and measuring psychological stress using social media." Proceedings of the International AAAI Conference on Web and Social Media.,Vol. 11. No. 1, 2017.

[3] Pang, Bo, and Lillian Lee., "Opinion mining and sentiment analysis." Foundations and Trends in Information Retrieval,Vol. 2. No. 1–2, 2008.

[4] Gowda, D., Irshad, M.A., Mary, P.A., Prasad, Y.D., Prasad, K.D.V.

and Tarigonda, H., “Machine Learning-driven Quality Control

Strategies for Industrial IoT Applications,” 2024 4th Asian Conference on Innovation in Technology (ASIANCON), IEEE, pp. 1–7, 2024.

[5] Orabi, Ahmed H., "Deep learning for depression detection of Twitter users." Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology, 2018.

[6] Trotzek, Marcel, Sven Koitka, and Christoph Friedrich., "Utilizing neural networks and linguistic metadata for early detection of depression." CLEF eRisk Workshop, 2018.

[7] Manivannan, K., Suresh, T. and Parthiban, M., “Big data analytics assisted arithmetic optimization with deep learning model for sentiment classification,” International Journal of Engineering Trends and Technology, 2023.

[8] Mikolov, Tomas.,"Efficient estimation of word representations in vector space." Proceedings of the International Conference on Learning Representations, 2013.

[9] Pennington, Jeffrey., Richard Socher, and Christopher Manning. "GloVe: Global vectors for word representation." Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2014.

[10] Pathak, D., Gowda, D., Manivannan, K., “Advanced Machine Learning Approaches to Evaluate User Feedback on Virtual Assistants for System Optimization,” IEEE, 2024.

[11] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural Computation.,Vol. 9. No. 8, 1997.

[12] Mary, P.A., Manivannan, K., “Leveraging Hybrid Neural Networks for Misinformation Detection in Online Social Media,” IEEE, 2025.

[13] Devlin, Jacob., "BERT: Pre-training of deep bidirectional transformers for language understanding." Proceedings of NAACL-HLT, 2019.

[14] Ji, Shaoxiong., "MentalBERT: Publicly available pretrained language models for mental healthcare." IEEE Transactions on Artificial Intelligence,Vol. 3. No. 5, 2022.

[15] Tehlan, A., Moharana, A.,“Role of artificial intelligence and machine learning for enabling IoT-enabled healthcare systems,” 2023.

[16] Resnik, Philip.,"Beyond LDA: Exploring supervised topic modeling for depression-related language." Proceedings of CLPsych Workshop, 2015.

[17] Tadesse, M.M., Lin, H., Xu, B. and Yang, L., “Detection of Depression-Related Posts in Reddit Social Media Forum,” IEEE Access, vol. 7, pp. 44883–44893, 2019.

[18] Chancellor, Stevie., "Quantifying and predicting mental illness severity in online communities." Proceedings of the ACM Conference on Computer Supported Cooperative Work, 2016.

[19] Shen, Guangyao., "Depression detection via harvesting social media: A multimodal dictionary learning solution." Proceedings of IJCAI, 2017.

[20] Tehlan, A., Moharana, A., Chintapalli, G., Mary, A.P. d Behera, S., “Role of Artificial Intelligence and Machine Learning for Enabling IoT-enabled Healthcare Systems,” Research and Applications Towards Mathematics and Computer Science, vol. 2, no. 2, pp. 46–58, 2023.

[21] Preotiuc-Pietro, Daniel., "Modelling valence and arousal in Facebook posts." Proceedings of ACL Workshop on Computational Linguistics and Clinical Psychology, 2016.

[22] Saha, Koustuv.,"A computational study of mental health awareness campaigns on social media." Proceedings of the International AAAI Conference on Web and Social Media, 2019.

[23] Reece, Andrew G., and Christopher M. Danforth. "Instagram photos reveal predictive markers of depression." EPJ Data Science,Vol. 6. No. 1, 2017.

[24] Chancellor, Stevie, and Munmun De Choudhury., "Methods in predictive techniques for mental health status on social media." ACM Computing Surveys, Vol. 53. No. 1, 2020.

[25] Cao, Yue.,"Machine learning approaches for mental illness detection on social media: A systematic review." Artificial Intelligence Review,Vol. 55. No. 6, 2022.