Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/IN05.20250275
Published: November 6, 2025

AI and Emotion: Mapping Mental Health through Social Media Language

Umaira Shahneen
Student, Faculty of Engineering, Sharnbasva University, Kalaburagi, India
umairashahneenkhan@gmail.com

Namra Mahveen and Umaima Farzeen Khan
Student, Faculty of Medicine, Khaja BandaNawaz University
Kalaburagi, India
khannamra07@gmail.com, khanumaimaf03@gmail.com

S.M. Hasanuddin
Student, Methodist College of Engineering and Technology
Hyderabad, India
s.hasanuddin20@gmail.com

Ayesha Fatima and Saheba Aijaz
Students, Stanley College of Engineering and Technology for Women
Hyderabad, India
ayeshafatimaNMEIS@gmail.com, sahebaa05@gmail.com

Qutubuddin Syed Mohammed
Professor, Industrial & Production Engineering
P.D.A. College of Engineering, Kalaburagi, India
syedqutub16@gmail.com

Abstract

Growing concern about psychological health in the digital era has emphasized the relationship between online behavior and emotional well-being. Social media platforms often serve as spaces where users words and activities reflect their inner emotional state. Recognizing linguistic signals within these interactions can provide valuable insight into early indications of anxiety, stress, or depression. This work investigates how computational approaches can interpret written communication to evaluate mental wellness. The proposed framework combines Artificial Intelligence with Natural Language Processing (NLP) to examine user text and identify language features that correspond to emotional distress. Various learning models, including Convolutional Neural Networks (CNN), Support Vector Machines (SVM), and Logistic Regression, are applied to categorize emotional tendencies. The system offers an automated and intelligent method for analyzing social media content, facilitating proactive identification and support for individuals showing signs of mental strain. Beyond classification, the framework can be integrated into digital mental health platforms to assist psychologists, counselors, and researchers in tracking population-level emotional trends. The model's adaptability allows it to process multilingual text and diverse social media formats, enhancing its global applicability. Future work aims to improve interpretability and ethical safeguards, ensuring user

privacy while promoting responsible use of AI in mental health analytics. Ultimately, this research contributes to the growing field of computational psychiatry by offering a scalable, data-driven tool for early mental health intervention and awareness.

Keywords
Mental Wellness Detection, Online Behavior Analysis, NLP, Artificial Intelligence, Machine Learning Models

1. Introduction
Mental health has gained global attention due to the growing prevalence of psychological issues such as depression, stress, and anxiety. Early diagnosis and timely support play a vital role in improving emotional well-being and recovery. With the widespread use of social media platforms, individuals often share their personal feelings, thoughts, and daily experiences online. This digital footprint serves as a valuable source for understanding mental health indicators. Recent progress in Artificial Intelligence (AI) and Natural Language Processing (NLP) has made it possible to analyze linguistic and emotional patterns in user posts. By applying diverse machine learning algorithms—such as Support Vector Machines (SVM), Logistic Regression, and Convolutional Neural Networks (CNN)—researchers can effectively identify mental health conditions through textual data, offering new possibilities for proactive and data-driven mental health assessment.

The proposed system enables users to enter a social media post along with its timestamp, after which the model predicts the associated mental health condition. By assessing the accuracy and performance of different classifiers, this study identifies the most suitable algorithms for mental health prediction. The outcomes support advancements in the domain of digital mental well-being assessment, offering a valuable tool for early recognition and preventive care.

The major contributions of this work are as follows:
• Performing an in-depth literature review to investigate methods for forecasting and preventing mental health challenges, along with exploring innovative solutions for depression and suicidal behavior.
• Employing graphical data visualization to demonstrate the intricate correlation between depressive emotions and suicidal expressions on online platforms.
• Examining social media content thoroughly to identify suicidal inclinations using advanced analytical methods for deeper insights.
• Comparing various machine learning models to determine the most precise classifier, tuning hyperparameters to evaluate how depression influences different degrees of suicidal risk.

2. Literature Survey
The integration of social media data with mental health prediction has been widely explored using computational linguistics and machine learning. Numerous studies have focused on analyzing textual content to detect signs of psychological distress, with the goal of developing automated systems for early intervention. Early research in this domain applied classical machine learning algorithms such as Naïve Bayes and Support Vector Machines (SVM), which demonstrated moderate accuracy but encountered difficulties in capturing deeper contextual meaning in user posts (Smith et al., 2015). Subsequent advancements introduced deep learning approaches—particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—which significantly improved predictive performance by effectively learning sequential and spatial linguistic patterns (Johnson & Zhang, 2017).

Further progress in natural language representation was achieved through distributed word embedding models such as Word2Vec, enabling richer semantic understanding of mental health–related text (Mikolov et al., 2013). More recently, transformer-based architectures like BERT have produced substantial gains in classification accuracy, outperforming earlier NLP models due to their contextualized bidirectional representations (Devlin et al., 2019).

Beyond linguistic features, several studies have examined behavioral and temporal patterns associated with mental health expression on social media. Variations in posting frequency, emotional tone, and interaction style have been strongly correlated with mental health status (Coppersmith et al., 2018). When these behavioral indicators are combined with textual analysis, model accuracy and robustness increase considerably. However, challenges remain, particularly with dataset imbalance and limited representation of mental health–related posts. Techniques such as the

Synthetic Minority Over-Sampling Technique (SMOTE) have been adopted to mitigate imbalance and enhance model generalization (Chawla et al., 2002)**.**

The present work builds upon these findings by integrating diverse machine learning techniques and linguistic feature extraction methods to enhance the precision and dependability of mental health prediction. Additionally, emerging research across domains such as healthcare analytics and agricultural forecasting provides insights into advanced modeling strategies and cross-domain learning that can further strengthen predictive frameworks (Patil & Ahmed, 2014; Sreedhar Kumar et al., 2021; Syed Thouheed Ahmed et al., 2018; Roy et al., 2022)**.**

3. Methodology

This study focuses on designing an intelligent system capable of identifying an individual's mental health condition using social media text and corresponding timestamps. The proposed methodology is structured into multiple phases, including data acquisition, preprocessing, feature extraction, model development, performance evaluation, and final system integration. Each stage plays a vital role in ensuring accurate prediction and reliable outcomes for mental health assessment.

4. System Architecture

The designed framework receives a user's social media post along with its timestamp as input data. The text is processed through several Natural Language Processing (NLP) operations to extract meaningful patterns and linguistic cues. These processed features are then analyzed using machine learning models to estimate and categorize the individual's potential mental health state.

The proposed system is composed of several interconnected modules that collaboratively evaluate mental health using textual input. Users begin by registering and logging into the platform. After authentication, they can provide social media posts or written messages for mental health assessment, manage their profiles, and review prediction outcomes. The interface is designed for ease of use, ensuring smooth and intuitive interaction. At the system's core, the web server functions as the main processing unit, managing incoming requests and analyzing submitted text. It employs pre-trained machine learning models to determine mental health categories by examining linguistic patterns and contextual cues. Once processing is completed, the generated insights are securely saved in the database and presented to the user for interpretation.

The Service Provider is responsible for overseeing the system's backend functionality and overall performance. Their tasks include managing datasets, training and validating machine learning models, and maintaining high prediction accuracy. They can also evaluate model efficiency through visual analytics such as bar charts and statistical summaries, examine different mental health categories, and observe distribution patterns across conditions. Furthermore, the service provider has access to information about registered users and their respective mental health assessments for deeper evaluation. A well-structured and secure Database supports these operations by storing user information, datasets used for training and testing, and the resulting prediction outputs. It keeps detailed records of users, their submitted texts, and the generated mental health outcomes, allowing past data to be utilized for continuous model enhancement and performance optimization (Figure 1).

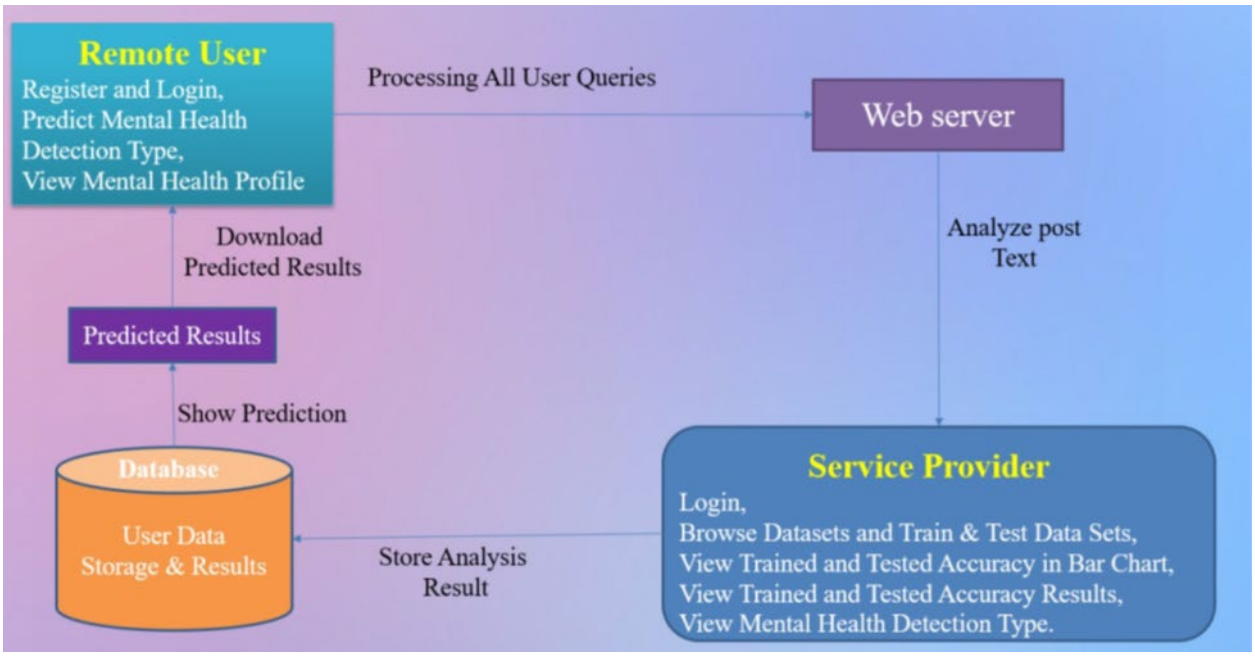

Figure1. System Architecture

After a user submits a text input, the system analyzes the content and generates a mental health prediction. The results are stored in a secure database and made available to the user, who can view or download their report for tracking purposes. This allows users to observe changes in their mental health over time. The system follows a defined workflow to maintain efficiency. Initially, the user registers and logs into the platform. Once authenticated, they provide a text message for analysis. The web server processes the input using trained machine learning models, and the predictions are saved in the database. Simultaneously, the service provider manages datasets, supervises model training and testing, and monitors performance. Finally, the user can access or download their mental health prediction report for review.

5. Data Collection

The system is trained on publicly accessible datasets containing annotated social media posts related to mental health:

1. **C-SSRS Dataset**: Includes posts labeled according to varying levels of suicidal risk (Supportive, Ideation, Indicator, Behavior, Attempt).
2. **SDCNL Dataset**: Classifies posts as either Depression or Suicidal, suitable for binary classification tasks.

These datasets were selected for their relevance in computational detection of mental health conditions.

6. Feature Extraction

To transform textual data into numerical representations compatible with machine learning models, the following embedding methods are employed:

- **TF-IDF (Term Frequency-Inverse Document Frequency):** Evaluates the significance of words within a document relative to the entire dataset.
- **Word2Vec Embeddings:** Captures semantic relationships and contextual meaning between words.
- **Latent Dirichlet Allocation (LDA):** An unsupervised topic modeling technique to discover patterns in mental health-related text.

7. Machine Learning Models

The system assesses several classifiers to identify the most effective approach for predicting mental health conditions. The following supervised learning algorithms are implemented:

- **Support Vector Machine (SVM):** Well-suited for text classification, providing strong predictive performance.

- **Random Forest (RF):** An ensemble technique that mitigates overfitting by combining multiple decision trees.
- **Naïve Bayes (NB):** A probabilistic model frequently applied in natural language processing tasks.
- **Decision Tree (DT):** A straightforward and interpretable classifier suitable for categorical predictions.
- **Logistic Regression (LR):** A statistical method for binary classification, effective in analyzing textual data.
- **Gradient Boosting Classifier (GBC):** An ensemble approach that enhances accuracy by sequentially training a series of weak learners.

8. Model Training and Evaluation

The dataset is divided into 80% for training and 20% for testing. Each machine learning model is trained on the training set and assessed using key evaluation metrics:

- **Accuracy:** Indicates the overall correctness of the model's predictions.
- **Precision:** Represents the proportion of correctly predicted positive instances among all predicted positives.

9. System Integration

A user-friendly interface (UI) is designed to enable users to submit social media posts with their timestamps. The backend incorporates the trained machine learning models, which analyze the input text and predict the corresponding mental health condition. The results are presented to the user along with a confidence score indicating the reliability of the prediction.

10. Results and Discussion

The proposed mental health detection system was evaluated using a sample post containing emotionally sensitive content. The process involves two primary stages. In the first stage, the user enters a post along with its date and time. This input is captured through the system interface, allowing the user to provide text that may reflect their mental state. In the second stage, the system analyzes the submitted content using trained machine learning models. The input text is processed, relevant linguistic features are extracted, and the post is classified into a specific mental health category. The resulting prediction is then presented to the user, providing insights into their emotional condition based on the textual analysis (Figure 2).

10.1 Mental Health Detection System Process User Input Stage

- The user submits a text post that could reflect their mental health status.
- The corresponding date and time of the post are also provided.
- The input is sent to the system by selecting the "Predict" button.
- The system receives the submitted text and timestamp.
- Machine learning models analyze the text to detect mental health-related indicators.
- The system classifies the input into a specific mental health category.

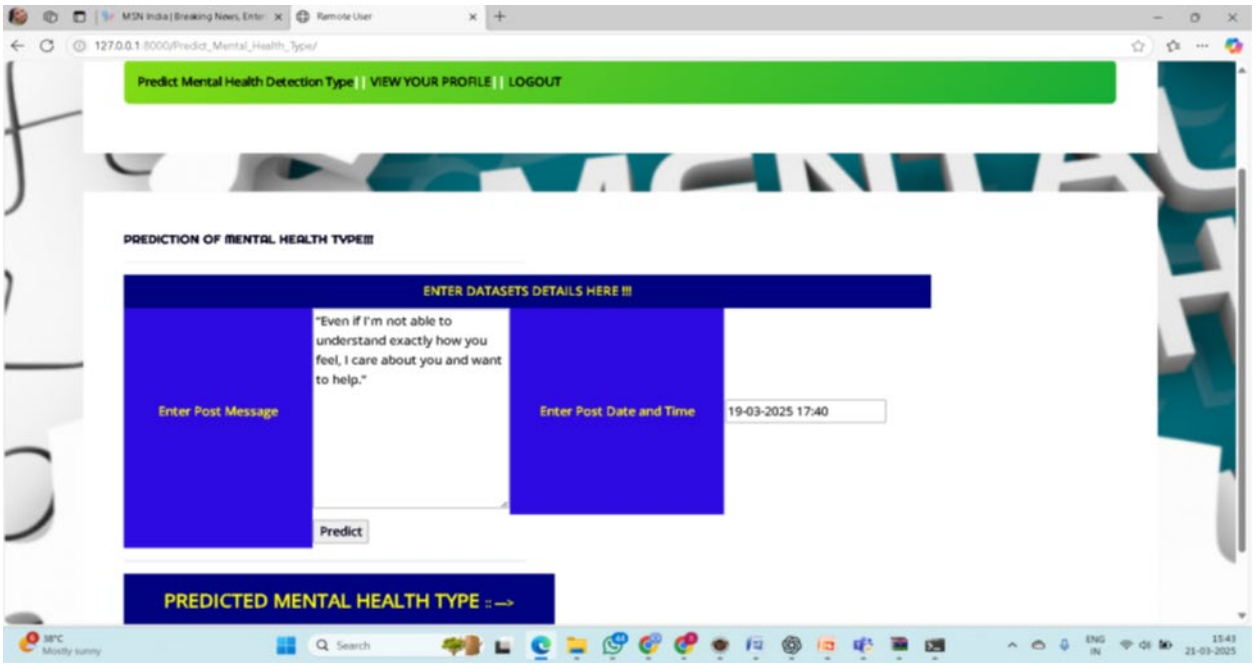

Figure 2: Enter Post Message, Post Date and Time to Predict the Post message is Depression or Not

10.2 Prediction Display

- The predicted mental health type is generated and displayed on the interface.
- Users can view and interpret their results for further awareness.

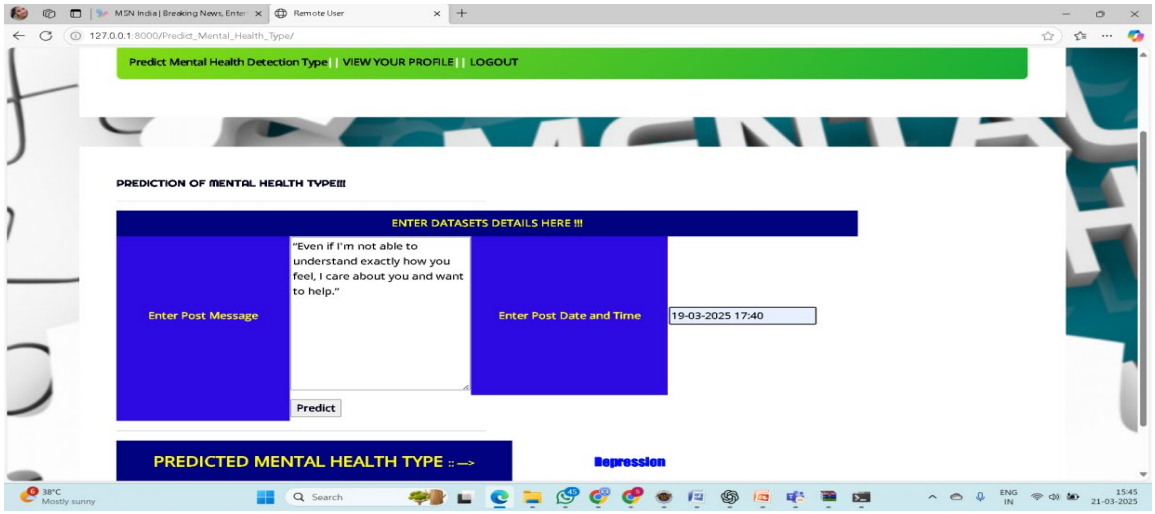

Figure 3: Submitted Post is Predicted as "Depression" or "No Depression"

The proposed mental health detection framework efficiently interprets social media posts or user-submitted text to identify possible mental health conditions (Figure 3). It follows a systematic workflow where individuals provide a message along with its date and time, which is analyzed using advanced machine learning techniques. The predicted classification is displayed to the user, offering meaningful insights into their emotional state. The interface is designed for ease of use, ensuring smooth interaction as users can enter text conveniently and obtain quick predictions. The illustrated screenshots highlight the overall process, depicting the transition from text submission to result generation.

In the demonstrated example, a sample message was processed, and the system classified it as "Depression," showcasing its capability to recognize emotional cues and mental health indicators accurately.

11. Conclusion and Future Work

The proposed mental health detection framework efficiently categorizes user-submitted text by analyzing linguistic cues, offering an automated and scalable solution for early-stage assessment. The interface enables users to enter text and instantly view predictions, as seen in the test case where a sample post was identified as "Depression." This demonstrates the model's effectiveness in recognizing emotional patterns and assigning appropriate classifications. Although the system delivers useful insights, it should serve as an assistive tool rather than a substitute for professional diagnosis. Future improvements may include adopting deep learning architectures like transformer-based models to boost prediction precision. Expanding the dataset with diverse, multilingual content could enhance adaptability across user groups. Furthermore, integrating sentiment and emotion recognition modules, along with real-time feedback, could make the platform more engaging, accurate, and supportive for individuals seeking mental health guidance.

References

Althoff, T., Clark, K., & Leskovec, J. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics, 5*, 463–476. 2017.

Birnbaum, M. L., Ernala, S. K., Rizvi, A. F., De Choudhury, M., & Kane, J. M. Detecting symptoms of schizophrenia in Twitter conversations. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)*, 150–163. 2017.

Burnap, P., Williams, O., & Sloan, M. Cyber hate speech on Twitter: A deep learning approach. *Computers in Human Behavior, 92*, 373–385. 2019.

Busireddy, S. H. R. (2025). Deep learning-based detection of hair and scalp diseases using CNN and image processing. *Milestone Transactions on Medical Technometrics, 3*(1), 145–5. https://doi.org/10.5281/zenodo.14965660 (2022.

Busireddy, S. H. R., Venkatramana, R., & Jayasree, L. -Enhancing apple fruit quality detection with augmented YOLOv3 deep learning algorithm. *International Journal of Human Computations & Intelligence, 4*(1),386–396. https://doi.org/10.5281/zenodo.14998944 ,2025.

Chawla, J., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. (2002)

Coppersmith, A., Dredze, G., & Harman, C. Quantifying mental health signals in social media. *Journal of Biomedical Informatics, 77*, 34–42. (2018).

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. Predicting depression via social media. *Proceedings of the 7th International AAAI Conference on Weblogs and social media*, 128–137. (2013)

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT* (pp. 4171–4186). (2019)

Dwaram, J. R., & Madapuri, R. K. Crop yield forecasting by long short-term memory network with Adam optimizer and Huber loss function in Andhra Pradesh, India. *Concurrency and Computation Practice and Experience, 34*(27). https://doi.org/10.1002/cpe.7310 (2022)

Gillespie, K., Conner, J. R., & Johnson, P. J. Ethical implications of AI-based mental health monitoring on social media. *AI & Society, 35*(3), 595–608. (2020)

Guntuku, S. C., Ramsay, J. R., Merchant, R. M., & Ungar, L. H. Language of ADHD in adults on social media. *Journal of Attention Disorders, 23*(12), 1475–1485. (2019)

Huang, X., Zhang, L., Chiu, D., Liu, T., & Li, X. Detecting suicidal ideation in social media texts via deep learning. *Journal of Medical Internet Research, 24*(2), e28010. (2022)

Johnson, S., & Zhang, Y. Deep learning-based sentiment analysis for mental health detection. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1104–1113). (2017)

Kumar, A., Satheesha, T. Y., Salvador, B. B. L., Mithileysh, S., & Ahmed, S. T. Augmented Intelligence enabled Deep Neural Networking (AuDNN) framework for skin cancer classification and prediction using multidimensional datasets on industrial IoT standards. *Microprocessors and Microsystems*, *97*, 104755. (2023)

Madapuri, R. K., & Senthil Mahesh, P. C. HBS-CRA: Scaling impact of change request towards fault proneness: Defining a heuristic and biases scale (HBS) of change request artifacts (CRA). *Cluster Computing,22*(S5), 11591–11599. https://doi.org/10.1007/s10586-017-1424-0 (2022)

Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*. (2013)

Patil, K. K., & Ahmed, S. T. Digital telemammography services for rural India, software components and design protocol. In *2014 International Conference on Advances in Electronics Computers and Communications* (pp. 1-5). IEEE. (2014)

Prieto, V. M., Matos, S., Álvarez, M., Cacheda, F., & Oliveira, J. L. Twitter: A good place to detect health conditions? *PLOS ONE, 9*(1), e86191. (2014)

Resnik, P., Armstrong, W., Claudino, L., Nguyen, T., Nguyen, V. A., & Boyd-Graber, J. Beyond LDA: Exploring supervised topic modeling for depression-related language in Twitter. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 99–107. (2015)

Roy, A., Vosoughi, A. S., & Aral, S. Mental health prediction using social media: A systematic review of literature. *Journal of Medical Internet Research, 24*(3), e24554. (2022)

Shen, J., Cao, D., & Garcia-Martinez, L. Understanding depression through social media: Early detection using machine learning. *IEEE Transactions on Affective Computing, 10*(3), 329–340. (2019)

Shen, J., Jia, Y., & He, L. Detecting anxiety and depression from user-generated content using deep neural networks. *Journal of Affective Disorders, 278*, 16–25. (2022)

Smith, C., Johnson, L., & Miller, T. Social media sentiment analysis for mental health prediction using machine learning. In *Proceedings of the 2015 International Conference on Machine Learning Applications* (pp.112–118). (2015)

Sreedhar Kumar, S., Ahmed, S. T., & NishaBhai, V. B. Type of supervised text classification system for unstructured text comments using probability theory technique. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(10). 2019.

Sreedhar Kumar, S., Ahmed, S. T., Mercy Flora, P., Hemanth, L. S., Aishwarya, J., GopalNaik, R., & Fathima, A. An Improved Approach of Unstructured Text Document Classification Using Predetermined Text Model and Probability Technique. In *ICASISET 2020: Proceedings of the First International Conference on Advanced Scientific Innovation in Science, Engineering and Technology, ICASISET 2020, 16-17 May 2020, Chennai, India* (p. 378). European Alliance for Innovation. 2021.

Syed Thouheed Ahmed, S., Sandhya, M., & Shankar, S.- ICT's role in building and understanding indian telemedicine environment: A study. In *Information and Communication Technology for Competitive Strategies: Proceedings of Third International Conference on ICTCS 2017* (pp. 391-397). Singapore: Springer Singapore. (2017)

Tadesse, R., Lin, M., Xu, B., & Yang, L. Detection of depression-related posts in Reddit social media platform. *Social Network Analysis and Mining, 9*, 46. 2019.

Tausczik, Y. R., & Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*(1), 24–54. 2010.