Grenze International Journal of Engineering and Technology, June Issue

AI based Detection of Depression and Anxiety through Social Media Analysis

Shashi Bhushan[1], Rajesh Kumar Pathak[2], Divya Rastogi[3], Mahima Shanker Pandey[4] and Vishvendra Pal Singh Nagar[5]
[1,5]Dept. of CSA, SSCSE, Assistant Professor, Sharda University, Greater Noida, India
Email: Shashi.cse45@gmail.com, Vishvendra.nagar@gmail.com
[2]Assistant Professor, Galgotia College of Engineering & Technology, India
Email: Rajeshkumar.pathak@gmail.com
[3]School of Computing, Grpahic Era Hill University, Bhimtal, India
Email: Divya.rastogi03@rediffmail.com
[4]Associate Professor, Galgotia College of Engineering & Technology, India
Email: mahimashanker@gmail.com

***Abstract*— The cases of depression and anxiety have been rising drastically since COVID-19 pandemic. There are quite a number of individuals who are left to be undiagnosed despite the progressive advancement of diagnostics because of the stigma, lack of knowledge and insufficient care amenities. A psychological insight is achieved with subtlety through the use of social media since the pages in the social media are typically full of personal feelings and problems posted there. The proposed research problem is how to find the symptoms of depression and anxiety at an early stage in social media content with the assistance of artificial intelligence (AI) and machine learning (ML), in particular, sentiment analysis and natural language processing (NLP). It is used to analyze the linguistic, emotional and behavioral patterns and contrast the findings with the deep learning models (e.g., LSTM, CNN, BERT) with the traditional machine learning models (e.g., SVM, Random Forest, Logistic Regression). Its results will have to show that the sentiment analysis using AI could be applied as a scalable, real-time, and non-invasive sentiment analysis tool to complement the conventional diagnostics. The significance of privacy, transparency and ethics are also noticed by the paper and presented as an efficient mental health provider and digital health project.**

***Keywords*— Natural language processing, machine learning, social media, depression and anxiety detection, sentiment analysis, and artificial intelligence.**

I. INTRODUCTION

Mental health conditions have been established to be among the most crucial health issues concerning the world in connection to the twenty first century. Anxiety and depression are two of the most prevalent and have a significant negative outcome with regard to social functioning, productivity and quality of life. The WHO predicts the figure to be more than 280 million individuals across the world that is traumatized by depression and over 264 million people are impacted by panic ailment. Inadequate educational and career performance, social pressure to maintain a specific lifestyle, heighten the vulnerability to physical conditions and, under the worst-case scenarios evolution of suicidal thoughts and behaviors, are all closely related to them.

Grenze ID: 01.GIJET.12.2.206_20

Stigmatization is the cause of the mental illness burden in the world, due to the ignorance and inaccessibility to qualified mental health professionals, which is more so in low and middle-income countries [1].

The traditional diagnostic methods of mental health disorders using self-reporting questionnaires, psychometric tests, and structured clinical interviews are crucial in the practice of managing mental health. However, these approaches usually consume numerous resources, time, and are scalable. Furthermore, people may postpone seeing a specialist because of the fear of social stigma and do not have a diagnosis or even do not pay mental health issues attention. Other researchers have considered alternative since these limitations provide mechanisms of the early surveillance and surveillance especially on the digital front where the citizens are free to become whoever they want to be[2].

The social media has proved to be an effective tool of capturing human feelings, thoughts and actions in real-time. People constantly post their everyday miseries, the personal experience and internal pain on social networks like Facebook, Instagram, Reddit, and Twitter. Stress, loneliness, and despair posts have random emotional post can be a sign of another underlying mental health issue. A case illustration is when De Choudhury et al. (2013) established that depressed Twitter users tended to use more first-person pronouns, words of negative emotions, and hopelessness than their counterparts who were control users[18]. Otherwise, in the conditions of the COVID-19 pandemic, certain studies claimed that social media expressions of anxiety and depression rose, and it gained a reflection of the psychological effect of lockdowns and health ambiguity.

Now, it is possible to perform mass analysis of such user-generated data with the development of the artificial intelligence (AI) and specifically machine learning (ML) and natural language processing (NLP). Sentiment analysis is a natural language processing that can be employed to classify statements on the attitudinal or emotional trait that is manifested in texts about the topic. Adding it to the machine learning algorithms, it becomes possible to automatically discover the signs, associated with anxiety and depression, in bulk amounts of social media information. The transformer based and recurrent neural networks have demonstrated deep learning models that are superb at context, semantics and subtle nuances of emotion in a written text[3].

Despite all these developments, there are various impediments that are present. The social media information is unstructured, situational based, and informal. Having said that, it is challenging to identify it, simply because of the differences in the language of the cultures, the prevalence of the use of slang or sarcasm, and the closeness of the qualities of the manifestation of stress and anxiety to that of sadness. Furthermore, a lot of useful AI models are black boxes and this raise concerns on their interpretability and medical reliability. Other significant obstacles to mass adoption include other ethical concerns like informed consent, privacy and responsible use of open psychological information.

II. Research Objectives and Problem Statement

It has been achieved through the ever-increasing reliance on the social media as an avenue of self-expression hence revealing signs of mental distress. Nonetheless, individuals do not receive professional help, and the diagnostic paradigm is not universally applicable as of now due to the stigma or the available help sufficiently scalable. This is the motivator of the current rush of the automated, faithful and ethically sound methods of early detection [5]. This research aims to

- Identify the instances of depression and anxiety on the premise of social media posting (NLP, sentiment analysis.
- Determine Language, emotional, and behavioral signs that indicate signs of psychological distressing.
- Compare the deep learning performance (i.e. LSTM, CNN, BERT) with the traditional ML models (i.e. SVM, RF, Logistic Regression).
- Research the topic of explainable artificial intelligence (XAI) to make model results more understandable and reliable.

III. Proposed Methodology

The proposed methodology is developed to identify depression and anxiety from social media posts using AI-driven sentiment analysis. The paper is groundbreaking because it will address a gap that has been recognized in numerous other previously published articles in terms of anxiety and depression identification. It utilizes transformer-based embeddings (BERT), integrated with classification layers to categorize text into mental health statuses. The overall workflow involve preprocessing of data, extracting the fetures, training the model and evalutaing the peroformance of the model. Further, compared to other models that consider solely on the performance, the one is an improvement as it incorporates explainable AI and explicit ethical safeguards.

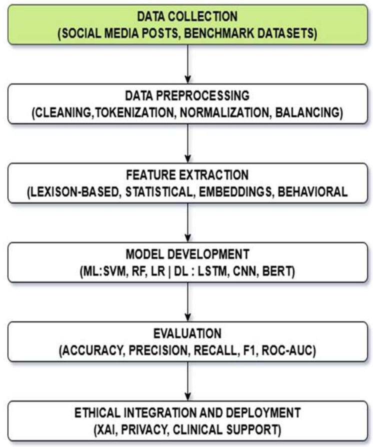

Figure 1. Research Methodology Framework for AI-Based Detection of Anxiety and Depression

The dataset used in this study was downloaded from Kaggle and contains text posts collected from Reddit, Twitter, and other social media platforms. Each post is labeled with a corresponding mental health category.

TABLE I. DESCRIPTION OF THE KAGGLE DATASET USED FOR ANXIETY AND DEPRESSION DETECTION

Attribute	Description
Source	Reddit, Twitter, Facebook
No. of Records	24,900
Input Field	Text posts/comments
Target Labels	7 mental health statuses
Selected Labels	Depression, Anxiety, Normal

For this study, only the classes Depression, Anxiety, and Normal were retained for binary and multiclass classification tasks.

A. Data Collection

The dataset was compiled from publicly available sources, including Kaggle, containing 52,681 labeled posts across multiple classes such as Anxiety, Depression, and Neutral. Each entry includes a statement (text post) and status (label). The data represent short texts and emotional reflections extracted from open-access social media platforms like Twitter and Reddit, in line with ethical guidelines ensuring anonymity and privacy.

B. Data Preprocessing

The derivation of the features is based on the patterns that indicate psychological distress is received. Various feature layers will be considered to make it stronger: Since raw social media text is noisy(slang, URLs, emojis, etc.), the following preprocessing pipeline was applied:
Steps performed:

- Removal of punctuation, URLs, emojis, and extra spaces
- Tokenization and lowercasing
- Label encoding for target classes
- Train–test split (80:20 ratio)
- Tokenization using BERT tokenizer

Mathematically, cleaned input text xix_ixi is transformed into embeddings as:

$$\chi_i' = BERTTokenizer(Cleaned\_Text_i)$$

C. Feature Extraction

Statistical Features (TF-IDF)
TF-IDF measures term importance within a document:

$$TF - IDF(t,d) = TF(t,d) \times \log\left(\frac{N}{DF(t)}\right)$$

Where:
- $TF(t, d)$ = frequency of term t in document d
- $DF(t)$ = number of documents containing t
- N = total number of documents

Embedding-Based Features
Contextual embeddings (Word2Vec, GloVe, BERT) capture semantic and psychological cues:

$$E = f_{BERT}(T) = \sum_{i=1}^{n} nw_i \, h_i$$

D. Model Development

Machine Learning Models
Baseline ML models were trained using TF-IDF features:
- Support Vector Machine (SVM)
- Random Forest (RF)
- Logistic Regression

Deep Learning Models
LSTM Model
Captures sequential dependencies in text:

$$h_t = f(W_h \cdot h_{t-1} + W_x \cdot x_t + b)$$

BERT Fine-Tuning
Fine-tuned using the pre-trained bert-base-uncased model for classification:

$$y = Soft\max(W \cdot [CLS]BERT + b)$$

$$Soft\max(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

Evaluation Metrics
We have evaluated the models using Accuracy, Precision, Recall, F1-score.
Accuracy= $TP + TN/TP + TN + FP + FN$
Precision=$TP/TP + FP$
Recall=$TP/TP + FN$
F1 score $= 2 \times \Pr ecision \times Recall / \Pr ecision + Recall$

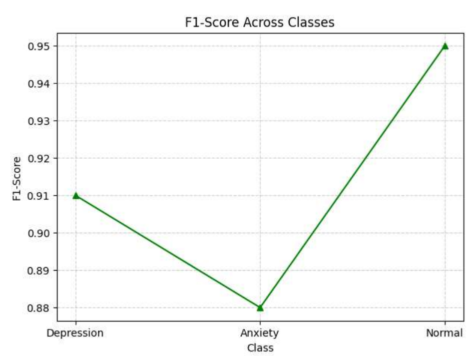

Figure 2. Model Performance Based on F1-Score   Figure 3. Model Performance Based on Recall

IV. Literature Review

The problem of mental health detection, which is a branch of artificial intelligence (AI) has been an actively developing interdisciplinary domain of data science, psychology, and computational linguistics. As the user-generated contents on the social media sites continue to expand rapidly, including Facebook, Twitter, and

Reddit, scholars have discovered that the social media offers a viable and conveniently accessible source of emotional and behavioral indicators. Research has discovered that digital footprints characterized by depressed or anxious individuals are usually overt, including more frequent use of first-person boxes, more frequent discussion of sadness, inconsistent posting, or a negative sentence polarity. Artificial intelligence (AI) based systems have been created to automatically detect these cues using natural language processing (NLP) and sentiment[13].

Research in this field has also resulted in the more complex machine learning models and, more so, deep learning and transformer-based architecture replacements of simple lexicon-based sentiment classifiers in the last ten years. Sentiment scores (which were often used to establish an indication of mental distress) were mostly determined using the predominant lexicons like LIWC, NRC and VADER. These approaches might be deduced; however, they were flaws regarded to the context and minor variations in the linguistics. The next step, machine-learning algorithms like the Random Forest, logistic regression and Support Vector Machines (SVM) were a potent constructive input to the classification task. These models also relied on the handcrafted models, which included n-grams, TF-IDF and lexicon-derived sentiment vectors, which would be more precise[14].

The invention of the deep learning technology has radically changed the paradigm of the mental health detection research. The local patterns and sequential dependence of text were separated by CNNs, long short-term memory (LSTM) models and recurrent neural networks (RNNs). These models have proven more effective than conventional baselines but it was often not understood, giving rise to explainable AI (XAI) in mental health use cases.

Recent transformers like the BERT, RoBERTa and DeBERTa massively changed the field by using contextual embeddings to elicit fine linguistic cues, and make them state of the art accurate. The common criticism of such models is that they are usually black boxed and also they make huge use of processing units.

The existence of bias and absence of fairness in AI-based detectors of depression and anxiety has also been brought to limelight by recent research. Studies indicate that the frameworks trained on corrupted sets experience failure in the minority groups and it can cause systemic injustices. Further, the privacy, accountability, and transparency have caused the debate due to the ethical aspects of the use of private social media data analysis without sufficient permission. To address these gaps, there should be appropriate structures to be adopted to a practical clinical practice that should use high accuracy and explainability, equity and ethical protection.

The findings of 2020-2025 are summarized in Table I which also outlines methods, findings, and limitations of studies.

TABLE II. SUMMARY OF RELATED STUDIES ON AI-BASED MENTAL HEALTH DETECTION

Author(s)	Year	Method / Model Used	Key Findings	Limitations / Issues Identified
Aldkheel	2023	Transformer embeddings + lexicon features	Hybrid approach improved interpretability while maintaining high accuracy	Limited dataset diversity; focused only on English
Basile et al.	2023	Multimodal fusion (text + metadata)	Improved prediction by combining linguistic cues with behavioral data	Ethical concerns regarding use of sensitive metadata
Ogunleye et al.	2024	Sentence-BERT embeddings + XGBoost	Achieved competitive results with interpretable outputs	Performance reliant on high-quality labeled datasets
Mansoor & Ansari	2024	ML/DL-based social media monitoring system	Validated real-world feasibility of AI-powered depression detection	Privacy and consent issues unaddressed
Reuters AI Audit	2024	Bias analysis on commercial depression AI models	Found systematic underperformance for Black Americans and minority groups	Did not propose bias mitigation methods
Allam et al.	2025	AI-driven public health surveillance	Demonstrated scalable real-time monitoring of depression and suicidal ideation	Raised surveillance and consent-related ethical risks
Ahmed et al.	2025	Scoping review of ML/DL for mental health	Highlighted transformer-based dominance; F1-scores up to ~0.90	Identified lack of anxiety-specific research and fairness evaluations

V. RESULT AND DISCUSSION

The anticipated results of the current research indicate the superior efficiency of the artificial intelligence (AI) and natural language processing (NLP) of identifying depression and anxiety cases through the analysis of social media sentiments. The research is anticipated to provide a comparative analysis of the methods in their relative accuracy, interpretability, and robustness through the transformer-based architectures, deep learning (DL) networks and the traditional machine learning (ML) algorithms.

Firstly, transformer-based models, including BERT, RoBERTa, and DeBERTa are likely to reach the highest results compared to the rest of the models that were explored. The current research was founded on the earlier study that confirmed that the state-of-the-art models with regard to transformers can indeed detect hidden linguistic and contextual features as evidenced by F1-scores of 0.82-0.90 levels when it comes to detecting depression. In a similar vein,
SBERT (sentenceBERT) with ensemble models, such as XGBoost models, have also shown good results, with lower accuracy competition, and higher interpretability results. The inferences of the findings are that the embeddings of contexts can probably be superior compared to simple statistical characterizations.
Second, the old machine learning models, including the Support Vector Machine (SVM), Random Forest (RF) and Logistic Regression (LR) need to be comparable in terms of the base level performance with accuracy scores falling in the range of 70-78 percent and F1-scores falling in the range of 0.65-0.75. The ability of such models to be explained, the low cost and efficiency of the computation also come in handy when compared with other models.
Thirdly, the performance variance will be anticipated between the detection of the process of the anxiety and depression. They also hope to obtain a specific 5-15 percentage points higher in F1-scores of depression classification in comparison with that of anxiety classification, which will meet the scoping review. Marking of uncertainty, avoidance diction, or stressful vocabulary was more linguistically explicit and depression signatures, use of words of negative affect, use of first person pronouns, and hopelessness expressions were more linguistically explicit.
As a result, he or she will not be able to spot all the apprehensive users because anxious models may be less recollected.
Fourth, feature engineering is sure to take a central role in performance results. As in contrast to statistical characteristics (TF-IDF), lexicon-based characteristics (e.g. sentence polarity and NRC emotion types) are likely to be easier to interpret.
However, the most likely significant ones are embedding-based capabilities (Word2Vec, GloVe, BERT) which integrate both semantic and contextual dependencies. It is anticipated that the behavioral characteristics (e.g. frequency of posting, time activity) should add to the earlier identification of finding the difference between the chronic conditions and the emotional moods short-term follow-ups.
The last one is fairness and subgroup analysis, which will expose the dissimilarity in model performance. The existing literature has identified that language and cultural disparity in manifestations of symptoms make some models inapplicable in their use when working with posts left by minority groups and, specifically, black Americans. Consequently, this is bound to bring significant questions of unfairness in the subgroup performance that ought to be eliminated before actual implementation though the overall accuracy may be satisfactory[16] .
All this suggests the process of transformer-based and hybrid models, as the most suitable ones to identify the presence of anxiety and depression with the assistance of AI, and the traditional machine learning models can be helpful in this regard. Neither has fairness been that important already.

VI. CONCLUSION

The increasing prevalence of anxiety and depression in the digital era highlights the urgent need for scalable, data-driven mental health solutions. Social media, once a space for casual interaction, has evolved into a digital reflection of human emotions and thoughts. This study demonstrates how artificial intelligence can interpret these expressions to identify early signs of psychological distress through sentiment analysis and advanced learning models.
By combining machine learning and deep learning techniques with lexicon-based and embedding-based features, the proposed framework enhances accuracy, scalability, and interpretability while maintaining ethical transparency and clinical complementarity.
Beyond individual detection, the research underscores the potential of AI-driven sentiment analysis in public health monitoring and policymaking. Real-time analysis of social media data can reveal mental health trends during crises, enabling timely interventions and informed decisions.
Integrating such systems into healthcare infrastructures can democratize access to mental health support, especially in resource-limited regions. The paper concludes by emphasizing the powerful intersection of technology, psychology, and ethics—positioning AI not merely as a computational tool but as a compassionate and responsible mechanism to reshape how society perceives, detects, and responds to mental health challenges in the digital age.

REFERENCES

[1] A. De and S. Mishra, "Augmented intelligence in mental health care: Sentiment analysis and emotion detection with health care perspective," Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis, pp. 205–235, 2022.

[2] P. Chakriswaran, D. R. Vincent, K. Srinivasan, V. Sharma, C.-Y. Chang, and D. G. Reina, "Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues," Applied Sciences, vol. 9, no. 24, p. 5462, 2019.

[3] H. Allam, C. Davison, F. Kalota, E. Lazaros, and D. Hua, "AI-driven mental health surveillance: Identifying suicidal ideation through machine learning techniques," Big Data and Cognitive Computing, vol. 9, no. 1, p. 16, 2025.

[4] M. A. Mansoor and K. H. Ansari, "Early detection of mental health crises through artificial-intelligence-powered social media analysis: A prospective observational study," Journal of Personalized Medicine, vol. 14, no. 9, p. 958, 2024.

[5] A. H. Zuberi et al., "Machine learning-based sentiment analysis for suicide prevention and mental health monitoring in educational institutions," Journal of Neonatal Surgery, vol. 14, no. 5s, 2025.

[6] L. AlSumait, A. AlFarhan, and H. AlHeneidi, "AI framework for DRIVE model-based mental health detection in text: A case study on how coping strategies are expressed during COVID-19," PeerJ Computer Science, vol. 11, p. e2828, 2025.

[7] K. Kasaudhan, "AI-driven tools for detecting and monitoring mental health conditions through behaviour patterns," International Journal of Preventive Medicine and Health (IJPMH), vol. 5, no. 3, pp. 14–19, 2025.

[8] N. V. Babu and E. G. M. Kanaga, "Sentiment analysis in social media data for depression detection using artificial intelligence: A review," SN Computer Science, vol. 3, no. 1, p. 74, 2022.

[9] A. Ahmed, S. Aziz, C. T. Toro, M. Alzubaidi, S. Irshaidat, H. A. Serhan, … and M. Househ, "Machine learning models to detect anxiety and depression through social media: A scoping review," Computer Methods and Programs in Biomedicine Update, vol. 2, p. 100066, 2022.

[10] M. L. Joshi and N. Kanoongo, "Depression detection using emotional artificial intelligence and machine learning: A closer review," Materials Today: Proceedings, vol. 58, pp. 217–226, 2022.

[11] D. K. Choudhary, A. Gupta, S. Singh, T. Abhinav, and T. Agrawal, "Sentiment analysis for depression detection using artificial intelligence," in 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT), May 2024, pp. 1–5.

[12] A. VarastehNezhad, R. Tavasoli, S. Elyasi, M. LotfiNia, and H. Farbeh, "AI in Mental Health: Emotional and Sentiment Analysis of Large Language Models' Responses to Depression, Anxiety, and Stress Queries," arXiv preprint arXiv:2508.11285, 2025.

[13] B. Bharadwaj, S. Nayak, and P. K. Panigrahi, "Sentiment analysis for identifying depression through social media texts using machine learning technique," Big Data and Computing Visions, vol. 5, no. 2, pp. 102–118, 2025.

[14] D. M. Mami and T. R. Xuan, "Artificial Intelligence in Mental Health: Detecting Depression and Anxiety Using Social Media Data," International Neuropsychiatric Disease Journal, vol. 22, no. 4, pp. 111–123, 2025.

[15] M. A. Mansoor and K. H. Ansari, "Early detection of mental health crises through artificial-intelligence-powered social media analysis: A prospective observational study," Journal of Personalized Medicine, vol. 14, no. 9, p. 958, 2024.

[16] D. Owen, A. J. Lynham, S. E. Smart, A. F. Pardiñas, and J. Camacho Collados, "AI for analyzing mental health disorders among social media users: Quarter-century narrative review of progress and challenges," Journal of Medical Internet Research, vol. 26, p. e59225, 2024.

[17] T. T. Elakya, "A Comprehensive Review on Artificial Intelligence based Depression Detection through Social Media Data Analysis," A Comprehensive Review on Artificial Intelligence based Depression Detection through Social Media Data Analysis, vol. 15, no. 84, pp. 1–9, 2024.

[18] De Choudhury, M., Gamon, M., Counts, S. and Horvitz, E., 2013. Predicting depression via social media. In Proceedings of the international AAAI conference on web and social media (Vol. 7, No. 1, pp. 128-137).