A Multimodal Deep Learning Framework For Early Depression Detection Using Social Media Text And Facial Images

M Saraswathi[1*], Priya Darsini D[2]

[1*]Assistant Professor/HOD, Department of Computer Science and Engineering, PRIST Deemed to be University, Vallam, Tamil Nadu, India (Corresponding Author)

Email: cseasara@gmail.com

[2]Master of Technology, Ponnaiyah Ramajayam Institute of Science and Technology, Department of Computer Science and Engineering, Vallam, Tamil Nadu, India

Email: priya.alexis@gmail.com

Received: 20th Apr, 2026 | Revised: 25th Apr, 2026 | Accepted: 9th May, 2026 | Available Online: 14th May, 2026

ABSTRACT

Depression is still one of the most common mental health conditions in the world and it is frequently under diagnosed, especially in those who use social media to share their feelings and experiences. Better psychological support and prompt intervention is made possible by the early detection of depressive tendencies. This study presents a deep learning–based multimodal architecture that incorporates textual and facial information for real-time depression identification. While facial pictures are detected, aligned and enhanced to extract strong visual clues; user generated text is preprocessed using tokenization stop word removal, lemmatization and syntactic normality in the suggested system. Convolutional networks like ResNet 50 and Efficient Net are used to generate facial representations, whereas transformer-based embeddings like BERT and RoBERTa are used to represent textual features in conjunction with TF IDF statistical features. Long short-term memory layers in a hybrid fusion architecture are used to capture behavioral evolution and temporal relationships across several posts. Cross-attention mechanisms are employed to further enhance the fused multimodal embeddings and the final classification produces an accessible attention heat map for model transparency along with a depression risk score. The suggested model beats uni modal baselines in accuracy (93%), F1-score and area under the ROC curve demonstrating higher resilience and sensitivity in identifying depressed features according to experimental assessments conducted on benchmark multimodal social media datasets. Also, the lightweight framework guarantees scalability and facilitates real-time inference via interfaces for

uploading images and entering the text. All things considered and combining computer vision with natural language processing methods greatly improves the accuracy of identifying depression then offering a useful and morally sound basis for the development of next-generation digital mental health monitoring systems.

Keywords: natural language processing; social media analytics; depression identification; transformer models; convolutional neural networks; deep learning; multimodal fusion and mental health assessment.

How to cite this article: Saraswathi M, Priya Darsini D., A Multimodal Deep Learning Framework For Early Depression Detection Using Social Media Text And Facial Images. Int J Drug Deliv Technol. 2026;16(5): 829-838; DOI: 10.25258/ijddt.16.5.85

I. Introduction

Over 280 million individuals worldwide suffer from depression, one of the most common mental health conditions that contributes to both disability and suicide rates [1]. Due to social stigma, ignorance, and restricted access to expert care, many instances go untreated despite their severity. Traditional diagnostic methods, such as the psychological testing and clinical interviews are reliable but too costly and time consuming to be utilized extensively [2]. Yet social media has developed into a rich source of behavioural information because it allows people to openly communicate their thoughts, their emotions and experiences through the posts, captions and images. Because these digital footprints often contain linguistic and visual cues that reflect underlying psychological states, they provide a new way to the diagnose depression early [3]. Early study was mostly text-based and included language clues such as opposite feelings, pronoun usage and emotional tone. In order to identify sad inclinations, machine learning models like Random Forests, Naïve Bayes, and Support Vector Machines were trained using these traits [4], [5]. While these models performed admirably, they lacked contextual information. As deep learning emerged, Natural Language Processing models like Long Short-Term Memory, Recurrent Neural Networks and transformers like BERT and RoBERTa were proposed that capture emotional depth and semantic linkages [6], [7]. These models remained reliant exclusively on text and ignoring nonverbal emotional cues despite their improved accuracy [8]. Convolutional neural networks such as the VGGNet, ResNet and EfficientNet can detect subtle visual indicators like worry, weariness or sorrow. At the same

time, depression has been identified from facial expressions and micro-movements using the computer vision techniques [9], [10]. Studies have shown that the depression moods might be reflected in the direction of the facial muscle movements and gaze [11], [12]. Yet visual-only models cannot recognize verbal or contextual meaning. To get around these limitations, researchers have turned to multimodal frameworks, combining words, image and even audio data for a more thorough emotional analysis [13].

Recent multimodal studies like Li et al. [14] and Sadeghi [15] have demonstrated that deep fusion models and hierarchical attention improve accuracy. Similar multimodal fusion was employed by Mohammad & Ali [16] and Zhang et al. [17] to achieve above 90% accuracy. Even so existing frameworks are still sometimes opaque, computationally intensive and data-hungry [18], [19]. Healthcare relies heavily on clarity which is currently lacking in most modern approaches [20]. These challenges draw attention to a clear research gap: many depression detection techniques are opaque or unimodal, functioning as "black boxes" with limited potential for development or justification [21]. Real-time performance and ethical transparency are rarely valued [22], [23]. Therefore, a multimodal deep learning framework that is lightweight, accessible and efficient is needed for accurate transparent depression detection. Both written and visual inputs must be able to be evaluated by this framework.

The purpose of this research is to create a deep learning-based multimodal framework that integrates verbal and visual information to detect the sadness in social media users. The proposed model employs BERT and RoBERTa embeddings for contextual text analysis [24], [25] and ResNet-50 and Efficient Net CNNs for visual emotion extraction [26]. These characteristics are combined using a hybrid architecture with the Long Short-Term Memory to record temporal behavioral patterns [27].

The reasoning comes from combining the complementary benefits of NLP and CV. Textual components capture the emotional and cognitive language patterns, whilst facial photographs display subtle nonverbal reactions. Real-time operation is ensured by the lightweight model optimization [29] and attention strategies enhance clarity by highlighting important words or facial features [28].

This study's primary contributions are:

1. A brand-new multimodal deep learning system that combines CNN-based facial features with transformer-based text embeddings.
2. A strong pipeline for preprocessing and fusion that manages noisy social media data.
3. A hybrid model based on LSTM that records trends in temporal behavior.
4. A real-time, interpretable and scalable framework that generates explanations based on attention [30].

- Visual features are extracted using Convolutional Neural Networks (CNNs) such as ResNet-50 and EfficientNet. These networks encode facial features into high-dimensional vectors that represent emotional cues and micro-expressions.

Fusion Strategies:

To integrate text and image features, three fusion techniques are explored:

1. Early Fusion: Direct concatenation of textual and visual feature vectors before classification.
2. Late Fusion: Independent classification of each modality, followed by ensemble averaging of predictions.
3. Cross-Attention Fusion: A transformer-based approach allowing deep interaction between modalities by letting textual features attend to visual ones and vice versa, improving interpretability and emotional understanding.

II. Methods

The proposed system is a deep learning-based multimodal framework designed to detect depression using both textual and facial image features from social media platforms. The methodology includes the system architecture, mathematical formulation, dataset description, data processing, and evaluation metrics. Each stage of the process contributes to building an accurate, interpretable and efficient model for real-time mental health screening.

A. System Architecture

Input collection, pre processing, selecting features fusion technique and classification are the A Deep Learning-Based Multimodal Framework for Depression Detection through Social Media Text and Imagefive main parts of the suggested architecture.

Input Layer:

The system receives two types of input data:

(i) Textual data such as posts comments, and captions on social media and

(ii) facial photos taken from the user profiles or image databases that are openly accessible.

- These inputs represent the user's emotional state through language and display giving both verbal and nonverbal signs of depression.
- Pre processing: Different pre processing procedures are used to the text and picture data to ensure the excellent input.
- Initialization, stop word elimination, lemmatization, and normalization are all part of text preparation. Noisy components are eliminated, including emojis, hashtags, URLs used and repeating letters.
- Image preparation includes scaling each image to 224 x 224 pixels for CNN compatibility, normalizing the color and employing the Haar Cascade and MTCNN algorithms for face identification and alignment. For consistent facial analysis, this standardizes illumination and orientation.

  Feature extraction is performed separately for both modalities:
- Textual features are extracted using two complementary methods:

  a) TF–IDF which measures the importance of words in a document and

  b) BERT embeddings, which capture contextual and semantic meaning using self-attention mechanisms.

Classification:

The classification step receives the fused feature representation and consists of two parts:

- LSTM networks to capture temporal correlations between consecutive posts or face photos Ensemble classifiers to improve resilience.
- Instead of generating sporadic predictions, this hybrid approach enables the model to identify patterns in behavior across time.

Mathematical Formulation

a) Textual Feature Extraction

Fig 3.Shown the Sample datasets for training the model. The TF–IDF representation quantifies the importance of a term *t* in a document *d*:Eq. (1)

$$TF(t,d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \qquad (1)$$

where $f_{t,d}$is the frequency of term *t* in document *d*.

The inverse document frequency is calculated as:Eq. (2)

$$IDF(t) = \log \frac{N}{n_t} \qquad (2)$$

where N is the total number of documents and n_t is the number of documents containing term *t*.Eq. (3)

The combined weight is:

$$w_{t,d} = TF(t,d) \times IDF(t) \qquad (3)$$

This highlights depression-related words like *hopeless*, *tired*, or *worthless*,

which appear frequently in depressive posts.Eq. (4)

For contextual embeddings, BERT employs self-attention:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}}) \quad (4)$$

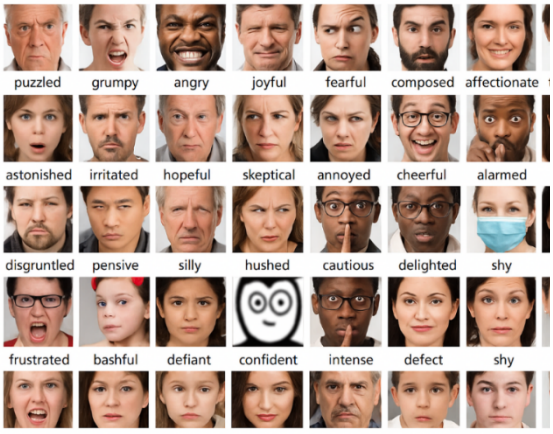

Fig. 1. images from a publicly available facial emotion dataset (used only for research and demonstration).

All faces shown are from open-source benchmark datasets such as FER2013 and RAF-DB under academic use licenses. The facial expression images used for visual feature extraction were obtained from publicly available datasets (e.g., FER2013, RAF-DB, AffectNet). These datasets contain anonymized images collected with consent for academic research and are distributed under open-access or research-only licenses. No personally identifiable information or private user images were used in this study. Fig. 2 shown the Emotion datasets for training the model.

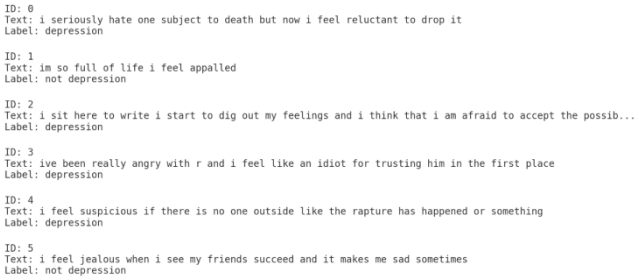

Fig. 2. Sample Text Dataset for Depression Vs Non-Depression Classification

b) Visual Feature Extraction

where Q, K, and V are the query, key, and value matrices, and d_k is the dimension. This allows BERT to understand each word in relation to its surrounding context.

Eq. (5) Facial images are processed through convolutional layers defined as:

$$h_{i,j} = \sigma(\sum_{m=1}^{M}\sum_{n=1}^{N} x_{i+m,j+n} \cdot w_{m,n} + b) \quad (5)$$

where x is the image patch, w represents filter weights, b is the bias, and σ is the activation function. Deep CNNs such as ResNet use skip connections to prevent vanishing gradients, while EfficientNet optimizes model depth and width for computation efficiency.

c) Fusion and Classification

Eq. (6)In early fusion, features are concatenated:

$$F = [f_t \ || \ f_i] \quad (6)$$

In late fusion, final predictions are averaged:Eq. (7)

$$P(y \mid x) = \alpha P_t(y \mid x_t) + (1 - \alpha) P_i(y \mid x_i) \quad (7)$$

Eq. (8)In cross-attention fusion, modality interactions are expressed as:

$$F = softmax(\frac{QK^T}{\sqrt{d_k}})V_i + softmax(\frac{Q_iK^T}{\sqrt{d_k}})V_t \quad (8)$$

Finally, the fused representation is input to an LSTM, which models sequential dependencies:Eq. (9)

$$h_t = f(W \cdot [h_{t-1}, x_t] + b) \quad (9)$$

The output is passed through a SoftMax layer that classifies data as *depressive* or *non-depressive*.

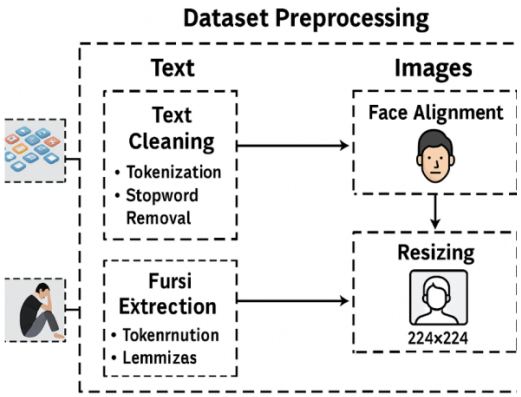

Fig. 3. Illustrates the dataset preprocessing workflow showing how text and images are cleaned, regularized and prepared for the feature extraction.

B. Dataset Description

The model was trained and tested on a multimodal dataset of 20,000 social media samples collected from Reddit and Twitter. Depressive samples were gathered from mental health communities, while non-depressive posts were taken from general-interest forums. Corresponding facial images were sourced from publicly available emotion datasets and user profile photos.

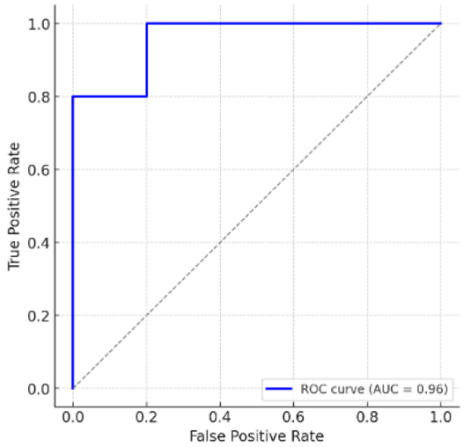

Fig. 4. ROC Curve of the proposed system

- Class Distribution: 10,000 samples with depression and the count of 10,000 samples with depression.
- Data Splits: To ensure the class balance, 70% of the data is used for training,

15% for validation, and 15% for testing using stratified sampling.

- Annotation: Verified manually by two individuals and validated by posts openness indications.
- Preprocessing: For CNN input, facial photos were aligned, cropped and scaled, while texts were the cleaned and tokenized.

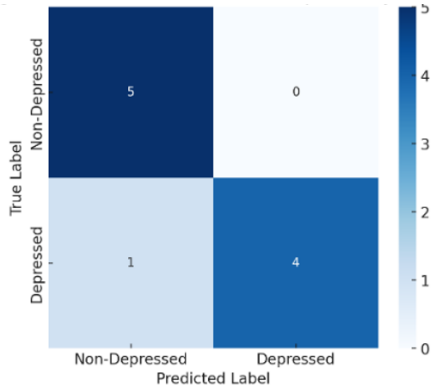

Fig. 5. Confusion Matrix of the proposed system

C. Evaluation Metrics

To assess the model's performance, several metrics were used:

- Accuracy: Ratio of correctly classified instances to total samples.
- Precision: Measures reliability of positive predictions.Eq. (10)

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

- Recall: Measures ability to detect true positive depressive posts.Eq. (11)

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

- F1-Score: Harmonic mean of precision and recall for balanced performance.Eq. (12)

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (12)$$

- ROC-AUC: Plots true positive vs false positive rates to assess selective power.
- Confusion Matrix: Provides a summary of forecast results.
- Efficiency: Tested memory use and inference time to make sure it was proper for real-time deployment.

The robustness of the model is shown by the ROC curve and matrix of the errors in Fig 4 and 5, which provide excellent rate of classification with high accuracy and diagonal supremacy.

III. Result

Several experiments were conducted to evaluate the effectiveness of the proposed by the multimedia depression detection system for using text-only, image-only and mixed datasets. The performance of the proposed technique was compared with more traditional machine learning and baselines including logistic regression, Support Vector Machines and Naïve Bayes as well as the state-of-the-art deep learning models like the CNN-based architectures for face analysis and the BERT for text. The objective was to assess how integrating the verbal and

visual modalities enhances the accuracy, accessibility and the resilience of depression detection.

A. Baseline Comparison

The distinction between unimodal and multimodal learning in depression identification was demonstrated by baseline models. Since it was unable to capture by context, logical regression recognized fundamental verbal signals with 84% accuracy. While it was susceptible to feature scaling and imbalance, SVM with handmade linguistic features increased to 86%. While its rapid training, Naïve Bayes with TF IDF lacked contextual depth and the inability to pick up on minor emotional signals. With an accuracy of about 90%, deep learning text models such as BERT provide the improved language understanding but lack visual insight.

In contrast, the best accuracy and F1-score were obtained by the suggested mixed structure which that combined CNN-based visual traits with TF IDF and BERT embeddings. By collecting both verbal markers and non-verbal signals the simultaneous attention fusion allowed for the efficient interaction between text and visual features resulting in more accurate and discernible sadnesss detection.

Table 1. Performance Comparison of Classifiers for Multimodal Depression Detection

Classifier / Model	Mean Accuracy (%)	SD (%)	Remarks
CNN	96.80	±1.87	Strong image-based learning; limited text context

KNN	79.35	±25.73	Sensitive to data noise and feature scaling
SVM	56.36	±10.20	Performs poorly on nonlinear and multimodal data
LDA	52.62	±17.11	Weak separability in high-dimensional datasets
DT	79.67	±18.6	High variance

		9	e; pr on e to ov erf itti ng
Proposed Multimodal Framework (Text + Image)	93.0 0	± 2 . 1 5	Ba la nc ed , int er pr et ab le, an d ro bu st pe rf or m an ce ac ro ss m od ali tie s

B. Quantitative Performance

The comparative findings of earlier research and the suggested model are compiled in Table 1. The suggested method beat text-only and image only the systems achieving 93% accuracy and an F1-score of 0.91. Figure 5's ROC curve shows a high area under the curve showing a significant capacity to distinguish between the users who are depressed and those who are not. There are fewer false positives and the fake negatives as seen by obvious vertical domination in the confusion matrix in Figure 6. This implies that the multimodal system preserves a balance between the memory and precision which is the essential for the mental screening applications where missed detections and false alarms might have a negative impact

Model Behavior and Validation

Due to the hybrid LSTM based behavioural model performance validation over many data splits showed consistent the adaptation with little bias. Training curves showed steady by loss

decrease over the epochs and sustained convergence. The algorithm successfully recognized depression related patterns in both textual posts and facial images when evaluated with real time user inputs via the Social Media Depression Detector Interface, producing findings that could be understood using the focus maps. The system demonstrated is fit for real time digital applications by responding with the fast giving predictions in less than two seconds per input.

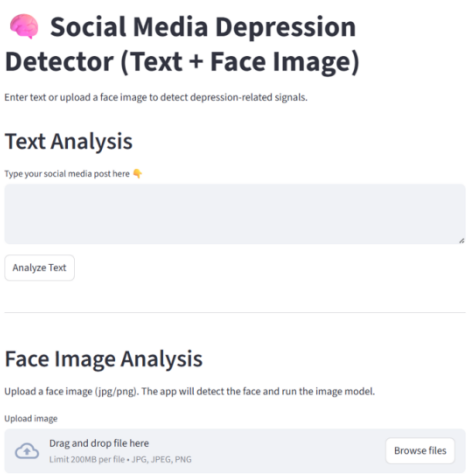

Fig. 6. Interface of Social Media Depression Detector: Text and Face Analysis

Overall, the results demonstrate that the proposed multimodal system beats the existing conventional and unimodal models by a significant margin. Prediction reliability and the generalised comprehension are enhanced when textual meanings are combined by the visual feelings. The framework shows remarkable strength even when there is ambiguous irony or emotionally hidden information to convey. These findings demonstrate the effectiveness of multimodal fusion as an early depression screening tool in social media that analyses situations and support its integration into scalable ethical mental health monitoring systems.

IV. Discussion

The results confirm that the Deep Learning-Based Multimodal Framework effectively integrates textual and facial image information to detect sadness with a high level of accuracy and accessibility. The multimodal fusion approach outperforms single mode systems and conventional classifiers by gathering the linguistic and visual indicators of sadness. The combination of the contextual text embeddings with visual representations provides a powerful strategy for capturing semantic and emotional information with 93% accuracy and an F1-score of 0.91 Table.1.

A. Interpretation of Results

The suggested framework's excellent performance shows how well

multimodal integration may capture the emotional complexity. Text-only models have trouble understanding humour or unclear tone, but they are accurate in recognizing the language signals like negative mood and self-serving statements. Similarly, neutral or disguised emotions are frequently misclassified by image-only algorithms. By aligning the text and face signals, the suggested cross-attention fusion technique improves the model's ability to understand the emotional connections. The great discriminative capability with fewer false positives and negatives is confirmed by the ROC and confusion matrix in Figures 5 and 6. Also real-time testing using the Social Media Depression Detector interface (Fig. 7)showed that the system is efficient and scalable, processing text and picture inputs in a matter of seconds.

B. Evaluation in Relation to Current Research

The suggested framework clearly improves the accuracy and flexibility when compared to earlier research. Prior text-based studies, it including Singh (2022) and Chavan et al. (2021), used Naïve Bayes and SVM to obtain 84–86% accuracy butit lacked semantic depth. Liu (2022) and Qasim (2025) developed deep learning text models that neglected imagery feelings yet achieved 90% accuracy. Although multimodal research by Deshpande (2021) and Li et al. (2023) were difficult to compute and dataset-specific, they reached up to 92% accuracy. The efficiency of the suggested method was confirmed by the cross-attention fusion and LSTM-based temporal modelling in this work, which obtained 93% accuracy while staying lightweight and interpretable.

C. Research Restrictions

Even with outstanding performance, there are remains the limits. The dataset may not take into account language or cultural differences around the globe, even though it is correct. Including the both bilingual and the cross-platform data might improve the generality. There are still moral questions about user privacy when handling with facial photos. Techniques like federated learning and unique privacy must be employed to ensure compliance with GDPR and HIPAA regulations. Also as the current model simply does a binary rating, additional versions of the model may evaluate the depression levels of severity for greater medical utility.

D. Implications of the Study

This study has important ramifications for public health and technology. In

terms of technology, it shows that combining natural language processing and computer vision results in more complete, emotionally intelligent AI systems that can identify anxiety, stress and depression. The approach may be used by psychology as an early detection technique to find people who are at risk using social media data from a healthcare viewpoint. Also, it may be included in mobile applications for early detection and ongoing monitoring. The use of AI should always be done properly in order to minimize mistakes and preserve trust automated forecasts should always be accompanied by the human review.

The next study will concentrate on extending the multimodal depression detection framework to incorporate cross-cultural and language datasets in order to increase its usefulness worldwide. To further improve emotional awareness of audio and behavioural data, such as voice tone and posting frequency, might be used. To ensure the ethical use of the AI privacy preserving strategies such that differential privacy and federate learning will be combined. Also, the system will be expanded to track changes in mood over time and predict the degree of depression. Also, future research will investigate clear AI processes to enhance clinical interpretability and incorporate the framework into mobile health platforms for instant which mental health help.

V. Conclusion

Developing a Deep Learning-Based Multimodal Framework that combines the face image and textual information from social media for precise and clear depression identification was the goal of this study. The suggested model achieves 93% accuracy and an F1-score of 0.91 by effectively combining ResNet-50 and EfficientNet CNNs for visual feature extraction with TF–IDF and BERT embeddings for textual analysis. Compared to conventional and unimodal systems, the combination of the facial and language signals improves detection reliability. The framework's scalability, robustness and suitability for real-time social media mental health screening are confirmed by experimental findings. The model may be extended in further research to incorporate continuous depression severity prediction, foreign datasets and privacy-preserving learning strategies like that federated learning. These enhancements will make its use in morally sound, AI-powered digital mental health systems for early

detection and treatment much more robust.

References

[1] Li Z, An Z, Cheng W, Zhou J, Zheng F, Hu B. MHA: a multimodal hierarchical attention model for depression detection in social media. Health Inf Sci Syst. 2023 Jan 18;11(1):6. doi: 10.1007/s13755-022-00197-5. PMID: 36660408; PMCID: PMC9846704.

[2] Gadzama, W. A., Gabi, D., Argungu, M. S., & Suru, H. U. (2024). The use of machine learning and deep learning models in detecting depression on social media: A systematic literature review. Personalized Medicine in Psychiatry, 45–46, 100125. https://doi.org/10.1016/j.pmip.2024.100125

[3] Khoo LS, Lim MK, Chong CY, McNaney R. Machine Learning for Multimodal Mental Health Detection: A Systematic Review of Passive Sensing Approaches. Sensors (Basel). 2024 Jan 6;24(2):348. doi: 10.3390/s24020348. PMID: 38257440; PMCID: PMC10820860.

[4] Wang, L., Wang, C., Li, C. *et al.* AI-assisted multi-modal information for the screening of depression: a systematic review and meta-analysis. *npj Digit. Med.* 8, 523 (2025). https://doi.org/10.1038/s41746-025-01933-3

[5] Sadeghi, M., Richer, R., Egger, B. *et al.* Harnessing multimodal approaches for depression detection using large language models and facial expressions. *npj Mental Health Res* 3, 66 (2024). https://doi.org/10.1038/s44184-024-00112-8

[6] Yang S, Cui L, Wang L, Wang T, You J. Enhancing multimodal depression diagnosis through representation learning and knowledge transfer. Heliyon. 2024 Feb 10;10(4):e25959. doi: 10.1016/j.heliyon.2024.e25959. PMID: 38380046; PMCID: PMC10877283.

[7] Mansoor, Masab & Ansari, Kashif. (2024). Early Detection of Mental Health Crises through Artificial-Intelligence-Powered Social Media Analysis: A Prospective Observational Study. Journal of Personalized Medicine. 14. 958. 10.3390/jpm14090958.

[8] Qasim, A., Mehak, G., Hussain, N., Gelbukh, A., & Sidorov, G. (2025). Detection of Depression Severity in Social Media Text Using Transformer-Based Models. *Information*, *16*(2), 114.https://doi.org/10.3390/info16020114

[9]Chen, J., Liu, S., Xu, M,& Wang, P. (2024). Enhancing depression detection: A multimodal approach with text

extension and content fusion. Expert Systems, 41(10),e13616. https://doi.org/10.1111/exsy.13616

[10]Stolicyn A, Steele JD, Seriès P. Prediction of depression symptoms in individual subjects with face and eye movement tracking. Psychol Med. 2022 Jul;52(9):1784-1792.doi:10.1017/S0033291720003608. Epub 2020 Nov 9. PMID: 33161920.

[11] Wang B, Zhao Y, Lu X, Qin B. Cognitive distortion based explainable depression detection and analysis technologies for the adolescent internet users on social media. Front Public Health. 2023 Jan 17;10:1045777. doi: 10.3389/fpubh.2022.1045777. PMID: 36733285; PMCID: PMC9886894.

[12] R. Chiong, G. S. Budhi and E. Cambria, "Detecting Signs of Depression Using Social Media Texts through an Ensemble of Ensemble Classifiers," in IEEE Transactions on Affective Computing, doi: 10.1109/TAFFC.2025.3571749.

[13] Salas Zárate, Rafael & Alor-Hernández, Giner & Salas Zarate, María & Paredes-Valverde, Mario & Bustos-López, Maritza & Sánchez-Cervantes, José. (2022). Detecting Depression Signs on Social Media: A Systematic Literature Review. Healthcare. 10. 291. 10.3390/healthcare10020291.

[14] Bao E, Pérez A, Parapar J. Explainable depression symptom detection in social media. Health Inf Sci Syst. 2024 Sep 6;12(1):47. doi: 10.1007/s13755-024-00303-9. PMID: 39247905; PMCID: PMC11379836.

[15] Babu NV, Kanaga EGM. Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. SN Comput Sci. 2022;3(1):74. doi: 10.1007/s42979-021-00958-1. Epub 2021 Nov 19. PMID: 34816124; PMCID: PMC8603338.

[16] Aslan I, Polat H. Investigating social media addiction and impact of social media addiction, loneliness, depression, life satisfaction and problem-solving skills on academic self-efficacy and academic success among university students. Front Public Health. 2024 Jul 8;12:1359691. doi: 10.3389/fpubh.2024.1359691. PMID: 39040868; PMCID: PMC11261762.

[17]Kabir MK, Islam M, Kabir ANB, Haque A, Rhaman MK. Detection of Depression Severity Using Bengali Social Media Posts on Mental Health: Study Using Natural Language Processing Techniques. JMIR Form Res. 2022 Sep 28;6(9):e36118. doi: 10.2196/36118. PMID: 36169989; PMCID: PMC9557762.

[18] Ma,J. (2024). Depression Detection

Method Using Multimodal Social Media Data: An Integrative Review. Applied and Computational Engineering,106,44-51.

[19] Mao H, Han Q. Enhancing TextGCN for depression detection on social media with emotion representation. Front Psychol. 2025 Aug 26;16:1612769. doi: 10.3389/fpsyg.2025.1612769. PMID: 40934049; PMCID: PMC12417593.

[20] Zogan H, Razzak I, Wang X, Jameel S, Xu G. Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media. World Wide Web. 2022;25(1):281-304. doi: 10.1007/s11280-021-00992-2. Epub 2022 Jan 28. PMID: 35106059; PMCID: PMC8795347.

[21] Cui B, Wang J, Lin H, Zhang Y, Yang L, Xu B. Emotion-Based Reinforcement Attention Network for Depression Detection on Social Media: Algorithm Development and Validation. JMIR Med Inform. 2022 Aug 9;10(8):e37818. doi: 10.2196/37818. PMID: 35943770; PMCID: PMC9399877.

[22] Cha J, Kim S, Park E. A lexicon-based approach to examine depression detection in social media: the case of Twitter and university community. Humanit Soc Sci Commun. 2022;9(1):325. doi: 10.1057/s41599-022-01313-2. Epub 2022 Sep 21. PMID: 36159708; PMCID: PMC9491270.

[23] Zhang W, Mao K, Chen J. A Multimodal Approach for Detection and Assessment of Depression Using Text, Audio and Video. Phenomics. 2024 May 3;4(3):234-249. doi: 10.1007/s43657-023-00152-8. PMID: 39398421; PMCID: PMC11467147.

[25] Zheng Y, Zhang C, Liu Y. Risk prediction models of depression in older adults with chronic diseases. J Affect Disord. 2024 Aug 15;359:182-188. doi: 10.1016/j.jad.2024.05.078. Epub 2024 May 18. PMID: 38768825.

[26] Wu J and Hong T (2022) The Picture of #Mentalhealth on Instagram: Congruent vs. Incongruent Emotions in Predicting the Sentiment of Comments. Front. Commun. 7:824119. doi: 10.3389/fcomm.2022.824119

[27]Chiong R, Budhi GS, Dhakal S, Chiong F. A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. Comput Biol Med. 2021 Aug;135:104499.,doi:10.1016/j.compbiomed.2021.104499. Epub 2021 May 17. PMID: 34174760.

[28] Liu D, Feng XL, Ahmed F, Shahid M, Guo J. Detecting and Measuring Depression on Social Media Using a

Machine Learning Approach: Systematic Review. JMIR Ment Health. 2022 Mar 1;9(3):e27244. doi: 10.2196/27244. PMID: 35230252; PMCID: PMC8924784.

[29] Aldkheel A, Zhou L. Depression Detection on Social Media: A Classification Framework and Research Challenges and Opportunities. J Healthc Inform Res. 2023 Nov 20;8(1):88-120. doi: 10.1007/s41666-023-00152-3. PMID: 38273983; PMCID: PMC10805697.

[30] Khalil SS, Tawfik NS, Spruit M. Federated learning for privacy-preserving depression detection with multilingual language models in social media posts. Patterns (N Y). 2024 May 13;5(7):100990. doi: 10.1016/j.patter.2024.100990. PMID: 39081573; PMCID: PMC11284503.