Depression Detection Technology Based on Multimodal Social Media Data: A Comprehensive Review and Future Outlook on Text and Image Fusion Analysis

Hongyi Pu *

International College, Chongqing University of Posts and Telecommunications, Chongqing, China

* Corresponding Author Email: znjcd@ldy.edu.rs

Abstract. Depression is a serious mental disorder, now ranking as the third leading cause of disability worldwide. It profoundly impacts an individual's physical and mental well-being, while also imposing substantial burdens on families, workplaces, and communities. Consequently, early and accurate detection of depression is of paramount importance. Traditional diagnostic methods, which rely on clinical interviews and self-report scales, often fall short due to their subjective nature and reliance on patient cooperation. However, recent advancements in machine learning offer promising new avenues for the objective diagnosis and early intervention of depression. Social media, with its vast troves of user-generated content, serves as a rich database where depressive symptoms can be subtly manifested in users' posts. This paper provides a comprehensive overview of the current methods for detecting depression through social media, focusing on text-image-based multimodal recognition techniques. In addition, the paper identifies existing limitations in these approaches and offers insights into future research directions, aiming to enhance the accuracy and applicability of depression detection technologies. *Contribution of This Paper :* This work contributes to the field by offering an in-depth analysis of the emerging multimodal methods used for depression detection through social media platforms. It identifies the strengths and weaknesses of current approaches, providing a critical assessment that informs future research. Additionally, the paper expands the scope of traditional depression diagnostics by incorporating advanced machine learning techniques, thereby laying the groundwork for more objective and early intervention strategies.

Keywords: Emotion Recognition, Multimodal Fusion, Depression Assessment.

1. Introduction

Research Background: Depression, also known as major depressive disorder, is a prevalent mental health condition characterized by persistent sadness, loss of interest or pleasure, insomnia, low self-esteem, difficulty concentrating, and changes in appetite and sleep patterns. It significantly impacts an individual's normal behavioral function and mental state, with severity ranging from mild, short-term episodes to severe, long-lasting conditions. According to the World Health Organization (WHO), approximately 1.2 billion people worldwide are affected by depression and anxiety, with around 280 million suffering specifically from depression. Adolescents, in particular, present a complex and diverse range of depressive symptoms compared to adults, including emotional instability, irritability, and various somatic complaints [1, 2]. Early detection and diagnosis of depression are crucial for effective management and prevention of severe outcomes.

Development of Depression Diagnosis: Traditional methods for diagnosing depression rely heavily on subjective self-reports and observations, utilizing tools such as the Self-Rating Depression Scale (SDS), Patient Health Questionnaire-9 (PHQ-9), and the Hamilton Depression Rating Scale (HAM-D) [3]. However, these methods often lack precise data and are particularly challenging in adolescent populations, where symptoms may be atypical, and reliable information from parents or caregivers may be difficult to obtain. Additionally, the shortage of child psychiatry specialists further complicates accurate diagnosis and treatment. In response to these challenges, researchers have begun exploring innovative approaches, including the use of social media data for early detection. Social media platforms, with their widespread use among adolescents, provide a unique opportunity to analyze user-generated content for signs of depression, offering a new direction in the early detection and diagnosis of this disorder [4, 5].

This Work: This paper aims to explore the use of multimodal methods, particularly the analysis of text and images from social media, in recognizing and diagnosing depression among adolescents. The second section presents the basic methods for multimodal depression recognition, focusing on the integration of deep learning and machine learning techniques. The third section reviews and discusses relevant research findings from both domestic and international studies, highlighting current applications and their effectiveness. The fourth section addresses future research directions and the challenges associated with these emerging technologies. Finally, the fifth section concludes with a summary of the paper’s findings and implications for the field of depression diagnosis.

2. Methodology

2.1. Processing and Analysis of Text Data

Social media, as a vast database, contains large quantities of various types of data, such as text data (including the text edited by users in posts, the text displayed in user profiles, etc.), image data (including images posted by users, profile photos, etc.), audio data (such as audio played by users on online streaming music websites or apps, and audio information posted by users), and other massive data. This data provides a foundation for models to effectively identify depression.

As shown in Figure 1, the general approach researchers use to detect depression by extracting text features involves first acquiring text data from social media databases, then preprocessing the data, extracting features, selecting appropriate features, training the model using different algorithms, and finally evaluating the model.

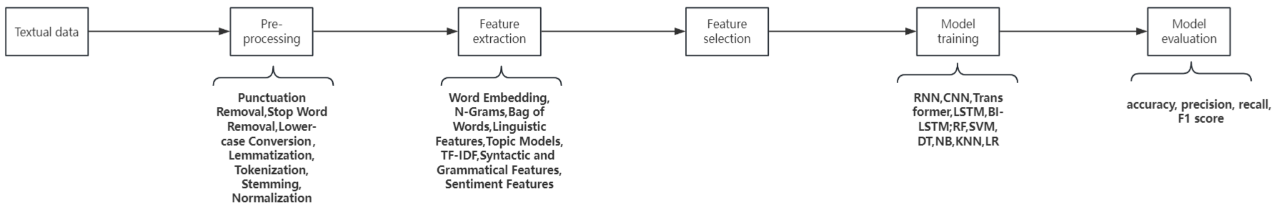

Figure 1. Data extraction process (Photo credit: Original).

2.1.1. Data preprocessing.

Text data usually contains a lot of noise, such as punctuation marks, special characters, HTML tags, etc. Preprocessing involves removing these irrelevant details to improve data quality, converting all characters to lowercase, removing extra spaces, eliminating URLs, removing stop words, and correcting spelling errors. Specific text preprocessing methods include:

Removing punctuation: e.g., “Hey,girl!” → “Hey girl”
Converting to lowercase: e.g., “Hey girl” → “hey girl”
Removing stop words: e.g., “this is Tom” → “Tom”
Tokenization: e.g., “hello world” → [“hey”, “girl”]
Lemmatization: e.g., “better” → “good”
Stemming: e.g., “beating” → “beat”
Spelling correction: e.g., “waht” → “what”

These preprocessing steps can significantly improve the quality and consistency of text data, making subsequent model training and prediction more accurate and efficient.

2.1.2. Feature extraction.

Feature extraction is the process of obtaining information from raw data that reflects depressive symptoms. It involves transforming text into numerical representations that capture the semantics and structure of the text, aiding the model in identifying and detecting depression. Commonly used methods for text feature extraction include: Word Embedding (e.g., Word2Vec, GloVe, BERT, and

FastText), N-Grams, Bag of Words, Topic Models, TF-IDF, Linguistic Features, Syntactic and Grammatical Features, Sentiment Features [6].

Satpute et al. used the TF-IDF method for feature extraction, assigning lower weights to common words and higher weights to words often emerge in a few documents to improve extraction accuracy [7]. Yao applied an N-gram model based on posts from depressed patients to convert segmented text into numerical representations of term or token frequency, demonstrating that the text posted by potential depressed users resembles language patterns unique to depression [8]. Abbas et al. proposed a new Bert-RF method for text feature extraction, where the BERT transformer model generates embeddings for each token in the text, and these latent features are included as inputs in a Random Forest model to generate probabilistic features for training a depression detection model [9]. Ahmad Wani et al. used the Word2Vec model to extract relationships between words within a corpus, mapping words into a continuous vector space to make them closer to each other, capturing relationships between sentences and textual context to extract sentiment features from the text [10]. In study, researchers used Word2Vec, GloVe, and BERT for text feature extraction and paired them with CNN, LSTM, CNN-LSTM, LSTM-CNN, and BERT classifier models for the Twitter dataset, finding that the BERT model was the optimal word embedding method for predicting depression, with significant advantages [11]. Less commonly used methods include Topic Modeling, where LDA represents documents as distributions of latent topics to uncover the underlying thematic structure in the text; Syntactic and Grammatical Features, which parse the syntactic structure of sentences and extract parse trees or dependency relations; and Sentiment Features, which use sentiment lexicons or sentiment analysis models to extract sentiment polarity and sentiment intensity from the text.

2.1.3. Feature selection.

Feature selection is a process used to choose the features that contribute most to the model's predictive performance. By selecting the most effective features and discarding redundant ones, the risk of overfitting can be reduced, and the model's complexity can be decreased, improving classification speed and efficiency. Main methods for feature selection including Filter Methods (e.g., Variance Threshold, Pearson Correlation Coefficient), Wrapper Methods (e.g., Forward Selection, Backward Elimination), Embedded Methods (e.g., Lasso Regression, Decision Trees, Random Forests). Similarly, some key function methods like SelectKBest, MRMR, and Boruta have also been used in the feature selection process for depression detection.

2.1.4. Model training.

Model training is a process of inputting data into the model through algorithms, letting the model automatically recognize and study the patterns and rules within the data. Researchers are now using social networks such as Twitter and Reddit to classify and identify depression, requiring complex machine learning methods for accurate detection. Commonly used machine learning methods in current depression detection research include CNN, Transformer, LSTM, Bi-LSTM, RF, SVM , DT , NB, LR and others [12-14].

Many researchers combine two or more machine learning methods to optimize model performance. Xu et al. used a two-layer LSTM model to recognize social media text. The two-layer LSTM model can extract deep features from the text, enriching the representation of the text vectors as the output of the neural network. Then, the pre-trained word embeddings from BERT are used as input into the two-layer LSTM model for sentiment classification, thereby achieving the classification of depressive texts. Gupta et al. collected a dataset of approximately 200,000 posts from Reddit and classified the dataset using CNN, LSTM, and LSTM-CNN models [15]. Experimental results showed that the hybrid LSTM-CNN model achieved an accuracy of 93.98%, a recall of 0.937, an F1 score of 0.940, and a precision of 0.942, which were higher than those of the other two models, indicating that the hybrid model is more effective for text-based depression detection. Prasad et al. combined CNN and BiLSTM and proposed a CNN-BiLSTM model, using the processed data from the Sentiment_140 dataset, which consists of 1.6 million tweets, as input for training, and compared the results with SVM and LSTM based on the same dataset. According to accuracy, recall, precision, and F1 score, they found that the CNN-BiLSTM model perform better the other two models [16].

2.1.5. Model evaluation.

Model evaluation is one of the critical steps in depression detection. Through model evaluation, we can judge the model's performance and make improvements, selecting the appropriate model for depression detection. The goal of evaluation is to make sure the result of the new data performed in the model is the same as the result of the training data, thereby validating the model's generalization ability. Common model evaluation methods include confusion matrix, cross-validation, ROC and AUC, MSE and RMSE, among others.

Accuracy, Precision, Recall, and F1 Score each reflect the model's performance, and the closer these values are to 1, the better and more effective the model is. The confusion matrix method is very common in depression detection [17]. The matrix is composed of four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), showing how the model's predicted results overlap with the actual results. The values of these metrics can be calculated using the following formulas to see if the model performed well.

$$\text{Accuracy} = \frac{|\mathrm{TP+TN}|}{|\mathrm{FP+FN+TP+TN}|} \tag{1}$$

$$\text{Precision} = \frac{|\mathrm{TP}|}{|\mathrm{FN+TP}|} \tag{2}$$

$$\text{Recall} = \frac{|\mathrm{TP}|}{|\mathrm{FP+TP}|} \tag{3}$$

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

2.2. Processing and Analysis of Image Data

Visual characteristics of images, such as color themes, brightness, contrast, saturation, and sharpness, can reflect users' emotions. Research found that depressed users often post images with darker color schemes and without themselves in the photos to show resistance and suppression of positive emotions [18]. In a study on using image recognition to identify depression on social media, Maxim et al. extracted color features from images by converting the images from the original RGB color space to other color spaces (HSV, XYZ, LAB) and calculated the mean and standard deviation to capture the color features of the entire set. Reece et al. extracted the number of faces in photos as one of the typical image features from 43,950 Instagram photos, achieving good results [19]. Ignatiev et al. detected 80 types of objects from user avatars, received images, and albums on the VKontakte social media network, and calculated the frequency of each object as image features to identify depression [20].

Similar to text features, the general approach researchers take to detect depression by extracting image features involves first obtaining image data from social media databases, preprocessing the image data, extracting features, combining them with text features, training the model with various algorithms, and finally evaluating the model.

Common preprocessing methods for image data include Resizing, Denoising, Segmentation, and Normalization, with the goal of making the processed images easier for subsequent feature extraction. For visual feature extraction, deep learning models such as CNN are the most commonly used methods. Shivangi et al. used the VGG16 network, pre-trained on the ImageNet dataset, to detect users' profile photos, posted images, and public personal albums on Instagram, achieving good predictive results [21]. Salunkhe et al. used the CLIP model to extract image features, embedding and aligning text and image features to capture both visual and textual characteristics [22]. Hu et al. used an attention-based RNN model to decode the content within images, generate implicit textual information, and then combine it with explicit textual information to obtain the final required features. Xu et al. resized images to 224×224 pixels and then used VGG19 to extract depressive emotion features contained in the images, which were then fed into a classifier for depression detection [23, 24]. Huang et al. employed transfer learning to pre-train the CNN model, enabling it to automatically capture features from images and represent them using vectors [25].

2.3. Multimodal Fusion of Image and Text Data

Feature fusion is a crucial step in multimodal sentiment recognition. Common feature fusion methods primarily include Feature-level Fusion, Decision-level Fusion, and Model-level Fusion. Feature-level Fusion, also known as early fusion, involves extracting features from multimodal data such as text and images and then combining them before inputting them into the subsequent model for processing. This approach allows the data features of each modality to be fully represented but may face challenges due to the high heterogeneity of the data, leading to higher requirements for data preprocessing. Decision-level Fusion is performed at a later stage. It involves independently extracting and classifying features from multimodal data to obtain local decisions or classifications for each modality. These results are then fused (e.g., through weighted averaging) to obtain the final decision. This method avoids the problem of high data heterogeneity but may overlook the correlations and influences between different modalities. Model-level Fusion is also a late-stage fusion method. It involves inputting the features from each modality into different models for processing, obtaining their respective outputs, and then integrating these outputs. This method makes full use of the characteristics of each modality but similarly weakens the connections between different modalities. The attention mechanism plays an important role in Model-level Fusion. Long et al. utilized the attention mechanism to fuse textual and image features for depression detection. Compared with simple concatenation fusion methods, the attention-based fusion method achieved higher accuracy, recall, and F1 score [26].

3. Analysis of Typical Research Methods

3.1. Comparison of Research Status

Current social media depression detection researches based on text-image methods and evaluation results are shown in table 1.

Table 1. Comparison of current methods.

Ref	Data Source	Methods	Accuracy	Precision	Recall	F1 Score
[27]	Twitter	LSTM+CNN	0.9844	-	0.9844	0.9692
[28]	WU3D	ALBERT+VGG16	0.9382	0.9253	0.8858	0.9051
[29]	Twitter	BERT+RNN	0.938	-	-	0.968
[30]	Twitter	EmoBERTa+CLIP	0.9172	0.878	0.9	0.908
[31]	Twitter	VNN+?	0.931	0.911	0.9	0.905
[32]	WU3D	BERT+BiLSTM+VGG16	0.905	0.9083	0.905	0.9048
[33]	WU3D	BERT+VGG19	0.9313	0.9328	0.9313	0.9312
[34]	Reddit	LSTM+CNN	0.9398	0.942	0.937	0.94
[35]	Facebook	CNN+BiLSTM	0.93	0.91	0.95	0.93
[36]	WU3D	XLNet+ResNet18	-	0.989	0.945	0.9665
[37]	WU3D	ALBERT+VGG16	0.8310	0.8468	0.8390	0.8429

Research has shown that deep learning algorithms outperform machine learning algorithms in depression detection, and the data volume directly affects the accuracy of both deep learning and machine learning methods. The larger the dataset, the more accurate the deep learning algorithm, and vice versa.

3.2. Applications of Multimodal Fusion Technology

Detecting depression based on multi-modal data of social media, some researchers have put forward many relevant practical application ideas. Social media platforms can establish an early emotion detection system to detect users in real time, identify users' emotional changes in a very short time according to the collected user voice, text, image and other multi-modal data, and automatically push psychological treatment plans, pleasing music, and one-on-one counseling services to users with depressive tendencies. Conduct individualized mental health intervention guidance. If the user still has a depressive tendency in the subsequent detection, the platform can guide the user to the hospital for treatment to prevent further deterioration of the condition. At the same time, the platform will also detect and mark suicidal texts, images and other content, if it is determined that the user has suicidal thoughts, it can inform the local mental health service center or police to launch emergency intervention measures to avoid the tragedy.

4. Future Research Directions and Challenges

4.1. Data Acquisition and Privacy Protection on Social Media

The construction of depression datasets based on social media is a major challenge for researchers. While platforms like Twitter, Instagram, and Facebook offer vast amounts of user-generated content, such as images and text posts, for research purposes, accessing this data could infringe on users' privacy rights. Utilizing this data for experimental purposes without platform oversight and user consent could also raise ethical concerns. Furthermore, users with depressive tendencies may be unaware that their personal information is being used as research data, and given their potentially fragile mental state, they may be unwilling to expose their private information. Failure to properly address transparency issues may leave a more serious damage on these individuals. With the advancement of the big data era, it has become inevitable for researchers worldwide to utilize social media to mine user information for research and analysis. However, this practice does affect users' privacy to some extent [38]. How to legally and effectively collect user data from social media without infringing on privacy rights and while respecting ethical standards is a challenge that must be faced in the construction of depression databases.

4.2. Further Optimization of Multimodal Fusion Technology

Depression detection based on multimodal social media data still faces many challenges. Depression detection that relies on common data types like text and images is relatively simplistic, and its effectiveness often decreases when dealing with posts from potential patients that are too brief, vague, or difficult to understand. Integrating new modalities from social media for fusion is one of the future directions.

The speech patterns of individuals with depression differ from those of normal individuals [39]. Generally, people with depression speak at a slower pace, with a lower pitch, quieter volume, more frequent pauses, a flatter tone, and may even have unclear articulation and prolonged silences. As a result, analyzing users' audio data has become one of the methods for detecting depression. Audio data, such as voice posts uploaded by users on social media or audio frequently listened to on music platforms, can be analyzed as vocal features to assess users' emotional states. Researchers have used the EDAIC-WOZ database to extract one-dimensional features of audio signals using MFCC and GTCC feature extraction techniques, applying them to a Bi-LSTM model to classify depressive states, successfully achieving high accuracy in depression prediction [40]. In the study, the authors utilize MFCC as the representation of audio features in the input layer. Multiple convolutional layers are used to extract features from MFCC, introducing non-linearity through the ReLU activation function, downsampling with Max pooling, and processing the features through a fully connected layer. The authors also integrated features extracted from video, linking the two types of features through late fusion to achieve depression recognition.

Additionally, users' behavioral patterns on social media can also serve as tools to reveal their emotional states. For example, the number, frequency, and timing of a user's posts, as well as behaviors like likes/comments/shares, and even the location from which the user accesses social media, can provide clues. Ji et al. proposed a Behavioral and Timeline Text Multimodal (BTTM) model to handle six behavioral features including followers count and so on. Compared to depression detection that relies solely on text, this model improves the accuracy of the assessment. Deng et al. used the WU3D dataset to capture user's posting frequency and timing of original tweets with images, considering both text and image features, feeding them into a classifier to predict depression.

4.3. Depression Detection in Cross-Cultural and Multilingual Contexts

Present research on social media depression detection is primarily focused on English platforms. There remains multiple challenges in detecting depression in cross-cultural and multilingual contexts. The same words in different cultural backgrounds may contain different meanings, which will increase the difficulty of emotion analysis and feature extraction of the text. Less common languages are often lack of enough annotation data, which makes it difficult to build the database. At the same time, different countries have different laws and regulations on the privacy of social media user data. Therefore, how to collect data legally and reasonably is also a major challenge that needs to be solved.

However, there are also researchers who have collected and utilized databases in less common languages to identify depression on social media. Soudi et al. collected a large number of Arabic tweets and extracted sentences and phrases that represent the meaning of "suicide" in Arabic. After complex data processing, they used a Bi-LSTM + CNN model to classify the Arabic text data, achieving an accuracy of 0.9064. Paramita et al. collected Indonesian language data from Twitter, annotated it and used the annotated data to train the IndoBERT and IndoBERTweet models, which were then used for classification. Uddin et al. collected 5,000 Bengali tweets from Twitter, preprocessed and categorized them into depressive and non-depressive tweets.Then they used a GRU model, with optimized hyperparameters to detect depression. Although this study merely achieved an accuracy of 0.757, it provided a new approach for future depression detection in Bengali and other languages. In another study on depression detection in Bengali social media, the researchers merged and processed the datasets established by Uddin et al. and Khan et al., using an XGBoost classifier to detect depression in Bengali social media posts. Because of the complexity of the Bengali language and the imperfections in the dataset, there was no suitable model for feature selection, which is an area that needs improvement in the future.

5. Conclusion

This paper provides a comprehensive review of the current techniques for detecting depression using multimodal social media data, emphasizing the importance of processing both text and images through a systematic approach. By detailing the five key steps—Data Preprocessing, Feature Extraction, Feature Selection, Model Training, and Model Evaluation—this work offers a clear framework for researchers and practitioners in the field. Additionally, the paper explores three methods for integrating multimodal data, providing insights into how these techniques can be effectively combined to enhance detection accuracy. By comparing and contrasting recent studies, this review highlights the practical applications of these methods in real-world scenarios and identifies the strengths and limitations of existing approaches.

Looking ahead, the field of multimodal depression detection on social media faces several challenges, including the need for more robust data integration techniques, improved model interpretability, and addressing privacy concerns associated with using personal data. Future research should focus on refining these detection models to increase their accuracy and reliability while ensuring ethical considerations are met. As AI and machine learning technologies continue to evolve, there is significant potential for these techniques to play a crucial role in early diagnosis and intervention, ultimately aiding in the timely treatment and support of individuals suffering from depression. This paper serves as a foundational reference for ongoing research and development in this critical area.

References

[1] S. R. Begum, S. Y. Sait, Effective techniques for depression detection on social media: A comprehensive review, 2022 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2022, pp. 1 - 9.

[2] X. L. Ye, W. Zhang, F. F. Zhao, Depression and internet addiction among adolescents: A meta-analysis, Psychiatry Res. 326 (2023) 115311.

[3] World Health Organization, Depression: What is depression? World Health Organization, 2012.

[4] A. De Los Reyes, T. M. Augenstein, M. Wang, et al., The validity of the multi-informant approach to assessing child and adolescent mental health, Psychol. Bull. 141 (2015) 858 - 900.

[5] R. Chiong, G. S. Budhi, S. Dhakal, F. Chiong, A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, Comput. Biol. Med. 135 (2021) 104499.

[6] A. A. Latif, Z. C. Cob, S. M. Drus, R. M. Anwar, H. M. Radzi, understanding depression detection using social media, 2021 6th IEEE International Conference on Recent Advances and Innovations in Engineering (ICRAIE), Kedah, Malaysia, 2021, pp. 1 - 6.

[7] B. S. Satpute, W. P. Rahane, R. Bharati, examining social media posts for identification of anxiety and depression utilizing machine learning techniques, 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2023, pp. 295 - 300.

[8] Yao, A multi-model approach to detection of depression in the Chinese social media entries, 2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT), Nanjing, China, 2024, pp. 2148 - 2151.

[9] K. Abbas, K. Munir, A. Raza, N. A. Samee, M. M. Jamjoom, Z. Ullah, Novel transformer-based contextualized embedding and probabilistic features for depression detection from social media, IEEE Access 12 (2024) 54087 - 54100.

[10] A. Wani, M. A. Elaf Fendi, K. A. Shakil, A. S. Imran, A. A. Abd El-Latif, Depression screening in humans with AI and deep learning techniques, IEEE Trans. Comput. Soc. Syst. 10 (2023) 2074 - 2089.

[11] R. Wang, J. Zhu, S. Wang, T. Wang, J. Huang, X. Zhu, Multi-modal emotion recognition using tensor decomposition fusion and self-supervised multi-tasking, 2024 International Journal of Multimedia Information Retrieval, 13 (4), pp. 39.

[12] M. S. Zulfiker, N. Kabir, A. A. Biswas, T. Nazneen, M. S. Uddin, an in-depth analysis of machine learning approaches to predict depression, Curr. Res. Behav. Sci. 2 (2021) 100044.

[13] X. Xu, G. Zhang, Q. Lu, X. Mao, Multimodal depression recognition that integrates audio and text, 2023 4th International Symposium on Computer Engineering and Intelligent Communications (ISCEIC), Nanjing, China, 2023, pp. 164 - 170.

[14] A. G. Reece, C. M. Danforth, Instagram photos reveal predictive markers of depression, EPJ Data Sci. 6 (2017) 15.

[15] B. Gupta, N. Pokhriyal, K. K. Gola, Mridula, detecting depression in Reddit posts using hybrid deep learning model LSTM-CNN, 2022 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS), Tashkent, Uzbekistan, 2022, pp. 610 - 617.

[16] M. G. Prasad, S. Jamadagni, Detection of depression using data collected from social media, 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023, pp. 1 - 7.

[17] A. N. Subedy, M. A. Khan, N. Kuldiloke, R. Katarya, Depression detection through user’s social network interactions, 2023 IEEE International Conference on Contemporary Computing and Communications (InC4), Bangalore, India, 2023, pp. 1 - 6.

[18] S. Rai, S. Bhattarai, P. Dhar, B. Upadhyaya, K. Sharma, S. Gurung, analyzing social media texts for suicidal risk identification using natural language processing, 2023 9th International Conference on Signal Processing and Communication (ICSC), NOIDA, India, 2023, pp. 227 - 231.

[19] Shivangi, Sangeeta, R. Rautela, Predicting depression with social media images, 2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC), Tandojam, Pakistan, 2024, pp. 1 - 8.

[20] X. Zhu, C. Guo, H. Feng, Y. Huang, Y. Feng, X. Wang, and R. Wang, "A Review of Key Technologies for Emotion Analysis Using Multimodal Information," Cogn. Comput., vol. 1, no. 1, pp. 1 - 27, 2024.

[21] K. Singh, U. Gupta, P. Renukadevi, Psychological analysis using social media tweets, 2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 2024, pp. 917 - 924.

[22] P. Hu, et al., "Making the implicit explicit: Depression detection in web across posted texts and images," in 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Istanbul, Turkiye, 2023, pp. 4807 - 4811.

[23] L. Xu, X. Shu, and J. Shu, "Research on depression tendency detection based on image and text fusion," in 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2022, pp. 326 - 331.

[24] S. Maxim, N. Ignatiev, and I. Smirnov, "Predicting depression with social media images," in Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods (ICPRAM), vol. 2, 2020, pp. 128 - 138.

[25] N. Ignatiev, I. V. Smirnov, and M. Stankevich, "Predicting depression with text, image, and profile data from social media," in Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2022, pp. 753 - 760.

[26] X. Zhu, Y. Huang, X. Wang, and R. Wang, "Emotion recognition based on brain-like multimodal hierarchical perception," Multimed. Tools Appl., vol. 83, no. 18, pp. 56039 - 56057, 2024.

[27] S. Liu, J. Shu, Y. Liao, Depression tendency detection for microblog users based on SVM, 2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2021, pp. 802 - 806.

[28] A. Anshul, G. S. Pranav, M. Z. U. Rehman, N. Kumar, A multimodal framework for depression detection during COVID-19 via harvesting social media, IEEE Trans. Comput. Soc. Syst. 11 (2024) 2872 - 2888.

[29] J. Yan, X. Shu, J. Shu, Depressive emotion tendency detection for users on social platform based on fusion of graph and text, 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2022, pp. 337 - 341.

[30] X. Zhu, H. Xu, Z. Zhao, X. Wang, X. Wei, Y. Zhang, and J. Zuo, "An environmental intrusion detection technology based on WiFi," Wirel. Pers. Commun., vol. 119, no. 2, pp. 1425 - 1436, 2021.

[31] R. B. Soudi, M. S. Zaghloul, O. M. Badawy, Framework for suicide detection from Arabic tweets using deep learning, 2022 32nd International Conference on Computer Theory and Applications (ICCTA), Alexandria, Egypt, 2022, pp. 96 - 103.

[32] P. Singh, G. Singh, S. Bharti, the predictive model of mental illness using decision tree and random forest classification in machine learning, 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2022, pp. 01 - 05.

[33] S. Zheqi, I. Park, predicting depression on Twitter with word embedding by pretrained language model, 2023 12th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 2023, pp. 247 - 252.

[34] G. V. Paramita, E. W. Andangsari, A. P. Kemala, M. M. Ali, A predictive model for identifying depression from social media using natural language processing techniques, 2021.

[35] A. H. Uddin, D. Bapery, A. S. M. Arif, Depression analysis of Bangla social media data using gated recurrent neural network, 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1 - 6.

[36] M. B. Mohammed, A. S. M. Abir, L. Salsabil, M. Shahriar, A. Fahmin, Depression analysis from social media data in Bangla language: An ensemble approach, 2021 Emerging Technology in Computing, Communication and Electronics (ETCCE), Dhaka, Bangladesh, 2021, pp. 1 - 6.

[37] M. R. H. Khan, U. S. Afroz, A. K. M. Masum, S. Abujar, S. A. Hossain, Sentiment analysis from Bengali depression dataset using machine learning, 2020 11th International Conference on Computing Communication and Networking Technologies (ICCCNT), 2020, pp. 1-5.

[38] V. Rajderkar, A. Bhat, Multilingual depression detection in online social media across eight Indian languages, 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1 - 6.

[39] Z. Wang, B. Deng, X. Shu, J. Shu, Multimodal depression detection model fusing emotion knowledge graph, 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2023, pp. 21 - 26.

[40] X. Long, Y. Zhang, X. Shu, J. Shu, Image-text fusion model for depression tendency detection based on attention, 2023 6th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2023, pp. 730 - 734.