Depression Detection Using Stereo Crossview Global Attention with Fully Convolutional Neural Networks

Kamalamma Puttaraju Impana
Department of Computer Science and Engineering, JSS Academy of Technical Education, Bengaluru, Karnataka, India
impanaraj@gmail.com (corresponding author)

Mullur Puttubuddhi Pushpalatha
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
mppvin@sjce.ac.in

Kotagi Basavarajappa Vikhyath
Department of Computer Science and Engineering, Dr H N National College of Engineering, Bengaluru, Karnataka, India
vikhyath059@gmail.com

Rajanna Raksha
Department of Computer Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka, India
raksha@jssstuniv.in

Received: 13 June 2025 | Revised: 12 August 2025 | Accepted: 25 August 2025

 DOI: https://doi.org/10.48084/etasr.12638

ABSTRACT

Detection of depression involves identifying physiological, behavioral, and textual cues to recognize signs of mental distress. Patterns in text, speech, physiological signals, and facial expressions are employed to assess a person's emotional state. However, traditional methods often fail to interpret data patterns due to limited long-range dependencies, which minimize model performance in depression detection. This research proposes Stereo Cross-view Global Attention with Fully Convolutional Neural Network (SCGA-FCNN) to detect depression effectively. In traditional FCNN, SCGA is incorporated to capture global contextual dependencies among multiple views, enhancing spatial awareness and cross-view interactions, allowing the model to detect subtle emotions more efficiently. Support Vector Machine-Synthetic Minority Over-sampling Technique (SVM-SMOTE) is applied to balance data by generating synthetic minority class samples near the support vectors, which helps to preserve and define more accurate class boundaries. Compared to existing methods such as Embeddings from Language Models (ELMo) with SVM, the proposed SCGA-FCNN achieves a higher accuracy of 89.98%, 85.75%, and 94.17% for NLSAA, PHQ-9, and Dreaddit datasets, respectively.

Keywords-depression; embeddings from language models; fully convolutional neural network; stereo cross-view global attention; support vector machine

I. INTRODUCTION

Depression is a medical condition and one of the main mental diseases that affects millions of people worldwide. Depression affects the learning ability of a person, causing mood fluctuations and minimizing work effectiveness [1, 2]. In geriatric patients, depression is focused on recognizing emotional and cognitive changes related to aging. The World Health Organization (WHO) states that 322 million people currently suffer from depression. Depression leads to social withdrawal and isolation, and social media platforms are increasingly offering affected individuals a way to connect with others who experience similar challenges [3]. Individuals experiencing depression find comfort in communicating emotions and thoughts anonymously and privately through social media platforms [4, 5]. Depression detection provides significant advantages because it allows seeking timely information from physicians, avoiding deterioration of the condition and enhancing outcomes [6]. Numerous studies employ Deep Learning (DL) methods to detect depression due to the rise of transformer-based language models to determine users at depression risk and similar disorders in online environments [7, 8].

In [9], Embeddings from Language Models (ELMo) with Support Vector Machine (SVM) was introduced using sentence or word data to detect depression effectively. In [10], Bag of Words (BoW) with Linear Regression (LR) was used to detect and analyze stress-associated posts in Reddit's academic communities. In [11], a Multi-Explainable Temporal Net (METN) combined a Temporal Convolution Network (TCN)-based transformer model to detect depression. In [12], an ML-based method used a Support Vector Machine (SVM) for social media users to detect depression. In [13], Bidirectional Encoder Representations from Transformers (BERT) and Mental BERT were used to recognize depression and stress on social media. This study proposes a method called Stereo Crossview Global Attention with Fully Convolutional Neural Network (SCGA-FCNN) to detect depression by capturing long-range dependencies through SCGA.

The main contributions of this study are as follows.

- FCNN automatically learns spatial hierarchies by capturing complex patterns in text, ensuring better robustness and generalization in depression cases.
- In traditional FCNN, SCGA is incorporated to capture local and global dependencies that enhance spatial awareness and contextual understanding by focusing on the most appropriate features.
- The Support Vector Machine-Synthetic Minority Oversampling Technique (SVM-SMOTE) is used to balance the classes by generating synthetic data points for the minority class.
- The NSLAA, PHQ-9, and Dreaddit datasets contribute to mental health analysis by providing diverse text-based assessments, enabling robust performance.

II. PROPOSED METHODOLOGY

This study presents SCGA-FCNN to detect depression effectively. Initially, data is obtained from the PHQ-9, NSLAA, and Dreaddit datasets. Then, SVM-SMOTE is applied to balance the data, and the proposed SCGA-FCNN is used for depression detection. Figure 1 shows a block diagram of the proposed SCGA-FCNN.

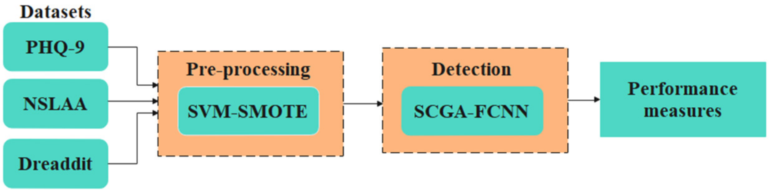


Fig. 1. Block diagram of proposed SCGA-FCNN

A. *Datasets*

The PHQ-9 [14], NSLAA [15], and Dreaddit [16] datasets are used to evaluate model performance. PHQ-9 provides 14-day depression-related ambulatory assessment scores, mod ratings, and depressed symptoms. NLSAA involves data from an 8-year geriatric depression survey among 65+ persons. This data includes 1042 observations and 1263 factors for diagnosing depression in older people. 3553 of 190,000 Dreddit posts are labeled. After preprocessing, the dataset was split into 80% for training and 20% for testing.

B. *Preprocessing*

SVM-SMOTE was used to balance the classes by generating synthetic minority class samples with an SVM decision boundary, which enhances model generalization. This method increases classification performance, particularly when the data have highly skewed class distributions. However, synthetic and original data are highly correlated, which causes SMOTE [17] to generate certain invalid sample data, resulting in poor quality of generated samples. Compared to derived methods and SMOTE, variable weights help to enhance the distribution and weaken the minority samples. The estimated vector w is obtained from SVM's hyperplane as the weight of the variable using:

$$f(x) = w^T x + b \tag{1}$$

where $w = (w_1, w_2, \ldots, w_p)^T$ represents estimated vectors, $f(x) > 0$ indicates that the sample x is categorized as the minority class, and $f(x) < 0$ shows that x is categorized as the majority class. Following preprocessing, the data is fed into the detection process using SCGA-FCNN.

C. *Detection*

After preprocessing, SCGA-FCNN is used to detect depression by capturing both spatial and contextual information across multiple views. FCNN [18] contains convolution, pooling, and deconvolution layers, which are associated with activation functions such as sigmoid.

1) *Stereo Crossview Global Attention (SCGA)*

In traditional FCNN, SCGA is incorporated to enhance feature representation in depression detection by capturing both local and global dependencies, which enhance contextual understanding. Initially, SCGA extracts two complementary types of features: local features F_L that indicate fine-grained information and global features F_R that capture broader and cross-view information. These features are integrated to form a unified representation that preserves both perspectives. The mathematical formulation of SB-LSH is:

$$\lambda_i = \{x_j | argmax(Mx_i) = argmax(Mx_j)\} \quad (2)$$

where $\lambda_i$represents a set of feature vectors x_j that belong to the same hash bucket x_i , x_j denotes the feature vector under comparison, M illustrates an orthogonal projection matrix utilized to calculate hash projections, and $argmax(Mx_j)$ determines the index of the maximum value in the projection used for bucket assignment. The entire process of SCGA is expressed as:

$$SF_L = SCGA(concate(F_L, F_R))L + F_L$$
$$SF_R = SCGA(concate(F_L, F_R))R + F_R \quad (3)$$

Figure 2 shows the architecture of the proposed SCGA-FCNN that enhances spatial dependencies by focusing on significant regions.

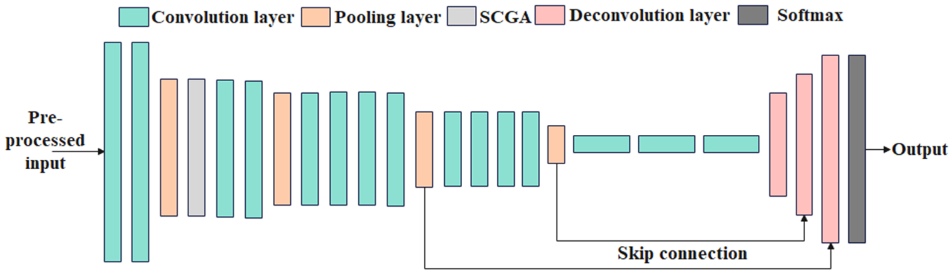


Fig. 2. Architecture of the proposed SCGA-FCNN.

III. EXPERIMENTAL RESULTS

The NLSAA and PHQ-9 datasets were used for performance analysis, whereas the Dreaddit dataset was utilized to compare the SCGA-FCNN with existing methods. The NLSAA dataset is used to detect whether a geriatric patient is depressed or not.

A. *Performance Analysis*

Table I demonstrates the performance analysis of different DL methods with the SCGA attention mechanism for depression detection. Existing methods, such as SCGA-Recurrent Neural Network (RNN), SCGA-Long Short-Term Memory (LSTM), and SCGA-Deep Neural Network (DNN), are compared with the proposed SCGA-FCNN.

Compared to these methods, SCGA-FCNN obtains a high accuracy of 89.98%, 85.75%, and 94.17% due to its ability to capture both global and local contextual information. BERT and RoBERTa struggled with contextual ambiguity, whereas ViT and CNN-BiLSTM have limited spatial feature representations, leading to suboptimal performance. Table II shows the performance analysis of different DL models without an attention mechanism for detecting depression. The FCNN obtains an accuracy of 78.80%, 77.60%, and 92.50% due to its inability to focus on the most appropriate features in intricate data. Figure 3 shows a graphical representation of different activation functions based on depression detection. The existing activation functions, such as Rectified Linear Unit (ReLU), tanh, and softmax, are compared with the sigmoid function.

TABLE I. PERFORMANCE ANALYSIS OF DIFFERENT DL METHODS WITH SCGA ATTENTION MECHANISM FOR DEPRESSION DETECTION

Datasets	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
NLSAA	SCGA-RNN	90.12	89.50	88.70	89.10
PHQ-9		83.45	83.00	83.10	83.05
Dreaddit		95.12	94.90	95.30	95.10
NLSAA	SCGA-LSTM	91.25	90.80	90.10	90.45
PHQ-9		84.30	83.85	83.95	83.90
Dreaddit		95.85	95.75	96.00	95.87
NLSAA	SCGA-BERT	88.65	87.90	87.50	87.70
PHQ-9		84.10	83.80	84.00	83.90
Dreaddit		92.75	92.40	92.60	92.50
NLSAA	SCGA-ViT	86.40	85.30	85.10	85.20
PHQ-9		82.75	82.10	82.50	82.30
Dreaddit		91.85	91.60	91.50	91.55
NLSAA	SCGA-RoBERTa	89.10	88.20	87.90	88.00
PHQ-9		85.20	84.90	84.95	85.00
Dreaddit		93.20	93.00	92.90	92.95
NLSAA	SCGA-CNN-BiLSTM	87.45	86.70	86.10	86.40
PHQ-9		83.40	83.10	83.00	83.05
Dreaddit		91.95	91.50	91.40	91.45
NLSAA	SCGA-FCNN	89.98	89.00	87.00	88.00
PHQ-9		85.75	85.28	85.35	85.32
Dreaddit		94.17	94.12	94.4	94.26

TABLE II. PERFORMANCE ANALYSIS OF DIFFERENT DL WITHOUT ATTENTION MECHANISM FOR DETECTING DEPRESSION

Datasets	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
NLSAA		76.50	74.80	74.70	74.75
PHQ-9	RNN	75.30	75.10	75.00	75.05
Dreaddit		90.20	90.10	90.00	90.05
NLSAA		77.00	75.40	75.30	75.35
PHQ-9	LSTM	76.10	75.80	75.70	75.75
Dreaddit		91.00	90.80	90.70	90.75
NLSAA		75.18	71.25	70.15	70.69
PHQ-9	BERT	70.36	72.36	72.08	72.21
Dreaddit		77.26	80.26	75.09	77.58
NLSAA		72.98	75.29	71.69	73.44
PHQ-9	ViT	72.69	75.18	75.39	75.28
Dreaddit		79.36	81.49	82.93	82.20
NLSAA		75.90	70.48	73.08	71.75
PHQ-9	RoBERTa	75.18	76.29	75.09	75.68
Dreaddit		85.19	86.29	82.64	84.42
NLSAA		76.87	71.06	75.69	73.30
PHQ-9	CNN-BiLSTM	75.14	75.09	73.68	74.37
Dreaddit		89.36	87.06	85.17	86.10
NLSAA		78.80	76.50	76.50	76.80
PHQ-9	FCNN	77.60	77.40	77.40	77.60
Dreaddit		92.50	92.60	92.60	96.50

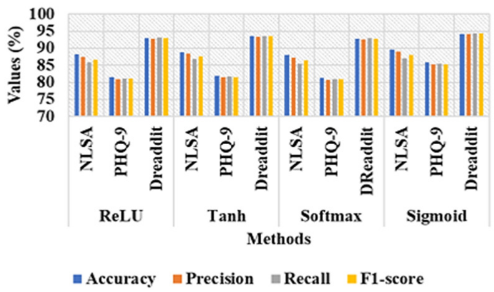


Fig. 3. Graphical representation of different activation functions based on depression detection.

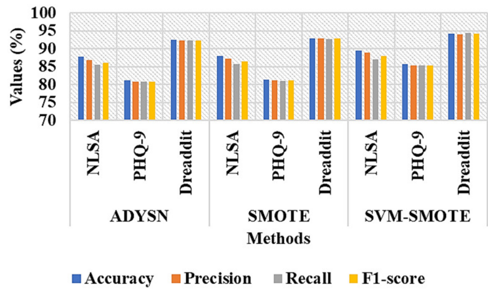


Fig. 4. Graphical representation of different imbalance methods for the preprocessing stage.

Figure 4 demonstrates the performance evaluation of different balancing methods for the preprocessing stage. The Adaptive Synthetic Sampling (ADASYN) and SMOTE are compared with SVM-SMOTE, where the latter achieves a high accuracy of 89.55%, 85.75%, and 94.17% for NLSAA, PHQ-9, and Dreaddit datasets by effectively managing class imbalances through the generation of synthetic data. Hence, SVM-SMOTE makes a more balanced misclassification in minority class instances and enhances accuracy. Figure 4 shows the graphical representation of different balancing methods.

Figure 5 plots epochs vs loss for the proposed SCGA-FCNN in (a) NLSAA, (b) PHQ-9, and (c) Dreaddit datasets. In NLSAA and PHQ-9, validation and training loss minimize steadily, showing that the model is learning effectively. For NLSAA, validation loss stabilizes at a low level with certain fluctuations in training loss, indicating better learning. In PHQ-9, both loss curves follow each other, indicating better generalization, whereas in Dreaddit, convergence is enhanced with less overfitting. These results indicate effective training and performance over three different datasets.

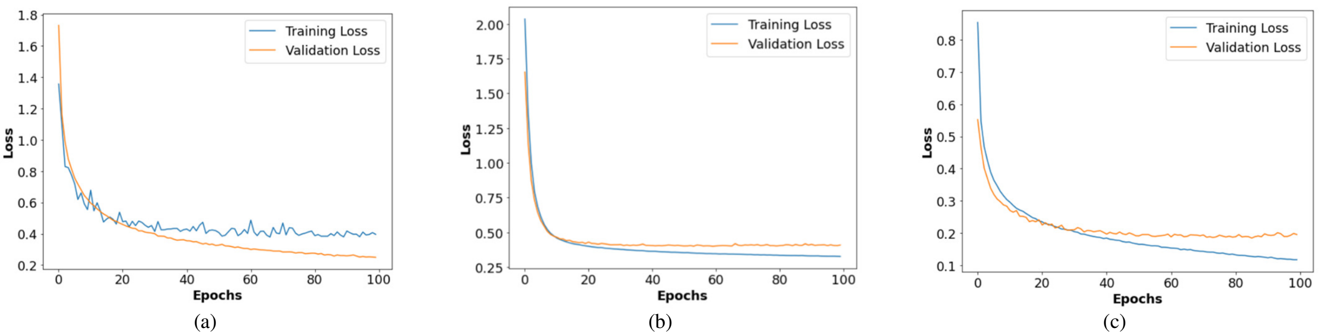
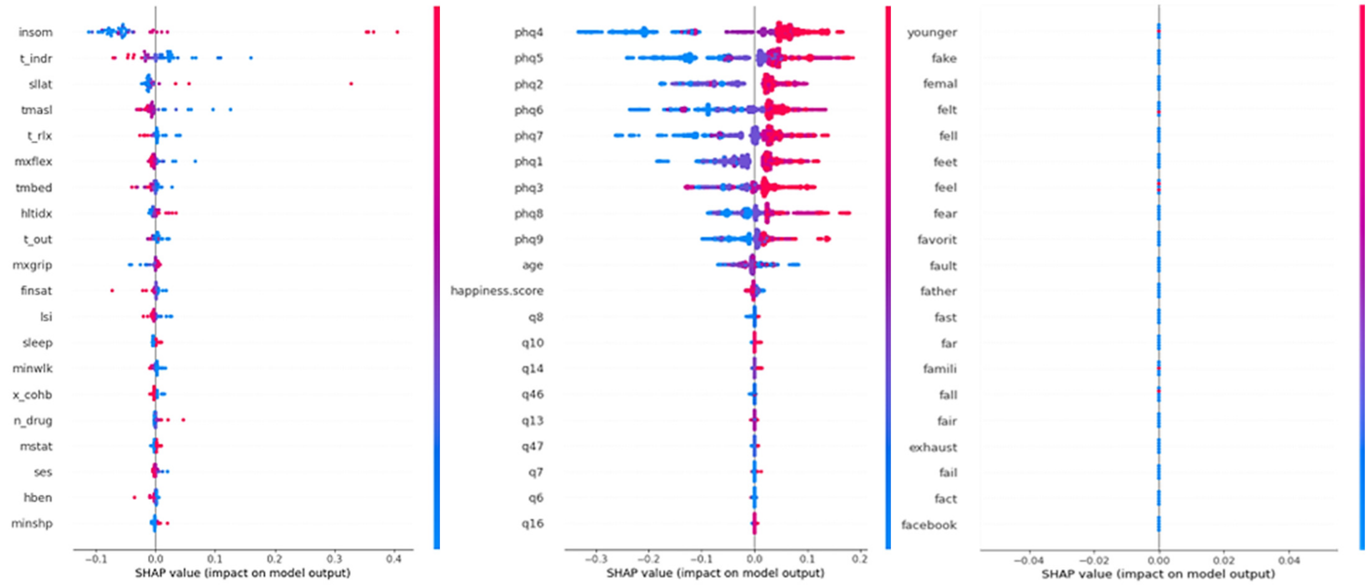


Fig. 5. Performance analysis of epochs vs loss for proposed SCGA-FCNN, (a)NLSAA, (b)PHQ-9, (c) Dreaddit datasets.

Figure 6 shows a visual graph of SHapley Additive exPlanations (SHAP), providing an interpretable understanding of feature importance across the three datasets: (a) NLSAA, (b) PHQ-9, and (c) Dreaddit. Features such as insom, sllat, and t_indr have the highest impact on model output in the NSLAA dataset, indicating that sleep-associated factors and insomnia are significant indicators of depressive symptoms.

In PHQ-9, phq4, ph5, and phq2 dominate in importance, demonstrating the model's sensitivity for depression detection, whereas the Dreaddit dataset shows less variation in SHAP values, indicating that text-based features contribute less to the model's decision-making. These visualizations determine the relevance of selected features for depression detection. As a result, SHAP helps determine the most influential attributes for an accurate detection of depression.



(a) (b) (c)

Fig. 6. Visual analysis of SHAP across three datasets (a) NLSAA, (b) PHQ-9, (c) Dreaddit

Table III represents an ablation study that demonstrates the individual contribution of key components. Compared to FCNN, FCNN+SVM-SMOTE, and SCGA-FCNN, SCGA-FCNN+SVM-SMOTE obtains high accuracy due to addressing both class imbalance and feature representation effectively. The incorporation of SCGA-FCNN offers rich feature representation, whereas SVM-SMOTE ensures balanced learning through generating synthetic samples near decision boundaries.

TABLE III. ABLATION STUDY DEMONSTRATING INDIVIDUAL CONTRIBUTION OF KEY COMPONENTS

Datasets	Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
NLSAA	FCNN	78.80	76.50	76.50	76.80
PHQ-9		77.60	77.40	77.40	77.60
Dreaddit		92.50	92.60	92.60	92.50
NLSAA	FCNN +SVM-SMOTE	83.70	82.85	82.60	82.72
PHQ-9		81.20	80.90	80.70	80.80
Dreaddit		93.00	92.85	92.90	92.87
NLSAA	SCGA-FCNN	86.45	85.90	85.10	85.50
PHQ-9		83.15	82.85	82.90	83.00
Dreaddit		93.65	93.45	93.40	93.42
NLSAA	SCGA-FCNN+SVM-SMOTE	89.98	89.00	87.00	88.00
PHQ-9		85.75	85.28	85.35	85.32
Dreaddit		94.17	94.12	94.40	94.26

B. Discussion

The limitations of existing methods such as ELMo [9], which generates dynamic word embeddings, and SVM, which relies on fixed feature spaces, lead to a loss of deep contextual understanding. BoW [10] ignores word order and semantics, which leads to the misinterpretation of nuanced stress-related expressions. METN [11] struggles with capturing long-range dependencies in highly irregular depression patterns because of TCN's receptive field. BERT and Mental BERT [13] struggle with contextual ambiguity in detecting stress and depression due to their misinterpretation of sarcasm or subtle cues in text. The proposed SCGA-FCNN overcomes these method limitations by capturing both global and local dependencies. SCGA enhances spatial and cross-view interactions, making the model learn a rich representation. Moreover, integrating SCGA with FCNN enhances the model's discriminative power, which improves detection accuracy while minimizing FN and FP. Therefore, this combination results in a more interpretable and reliable depression detection system.

IV. CONCLUSION

This study presented SCGA-FCNN to detect depression effectively. FCNN captures contextual features in text by employing convolutional layers that enable efficient performance. It processes inputs of varying lengths without the need for fixed-size vectors, which enhances model performance. SCGA captures both contextual and spatial information across multi-view data, which increases the model's ability to focus on appropriate features, leading to more robust as well as accurate detection. SVM-SMOTE solves imbalanced data by generating synthetic samples for the minority class and increases the model's ability to learn from underrepresented data, leading to superior performance and generalization. Hence, the proposed SCGA-FCNN achieves a higher accuracy of 89.98%, 85.75%, and 94.17% for NLSAA (focused on geriatric patients), PHQ-9, and Dreaddit datasets, which is better than existing methods such as ELMo.

However, real-time clinical data and emoji data are not considered, as the attention mechanism has only limited data. In the future, emoji data will be integrated to better capture emotional context. In addition, real-time data will be collected from social media platforms to support the deployment of the model in real-world clinical monitoring scenarios.

REFERENCES

[1] B. H. Bhavani, M. Sreenatha, and N. C. Kundur, "Diagnosis and Classification of Depressive Disorders using ML and DL Models," *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21383–21389, Apr. 2025, https://doi.org/10.48084/etasr.10017.

[2] M. K. Myee, R. D. C. Rebekah, T. Deepa, G. D. Zion, and K. Lokesh, "Detection of Depression in Social Media Posts using Emotional Intensity Analysis," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16207–16211, Oct. 2024, https://doi.org/10.48084/etasr.7461.

[3] J. P. Thekkekara, S. Yongchareon, and V. Liesaputra, "An attention-based CNN-BiLSTM model for depression detection on social media text," *Expert Systems with Applications*, vol. 249, Sept. 2024, Art. no. 123834, https://doi.org/10.1016/j.eswa.2024.123834.

[4] A. Gopalakrishnan *et al.*, "Prenatal depression level prediction using ensemble based deep learning model," *International Journal of Cognitive Computing in Engineering*, vol. 6, pp. 267–279, Dec. 2025, https://doi.org/10.1016/j.ijcce.2024.12.002.

[5] Z. N. Vasha, B. Sharma, I. J. Esha, J. A. Nahian, and J. A. Polin, "Depression detection in social media comments data using machine learning algorithms," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 2, pp. 987–996, Apr. 2023, https://doi.org/10.11591/eei.v12i2.4182.

[6] E. Bao, A. Pérez, and J. Parapar, "Explainable depression symptom detection in social media," *Health Information Science and Systems*, vol. 12, no. 1, Sept. 2024, Art. no. 47, https://doi.org/10.1007/s13755-024-00303-9.

[7] C. Xin and L. Q. Zakaria, "Integrating Bert With CNN and BiLSTM for Explainable Detection of Depression in Social Media Contents," *IEEE Access*, vol. 12, pp. 161203–161212, 2024, https://doi.org/10.1109/ACCESS.2024.3488081.

[8] E. Kerz, S. Zanwar, Y. Qiao, and D. Wiechmann, "Toward explainable AI (XAI) for mental health detection based on language behavior," *Frontiers in Psychiatry*, vol. 14, Dec. 2023, Art. no. 1219479, https://doi.org/10.3389/fpsyt.2023.1219479.

[9] S. Inamdar, R. Chapekar, S. Gite, and B. Pradhan, "Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing," *Human-Centric Intelligent Systems*, vol. 3, no. 2, pp. 80–91, June 2023, https://doi.org/10.1007/s44230-023-00020-8.

[10] N. Oryngozha, P. Shamoi, and A. Igali, "Detection and Analysis of Stress-Related Posts in Reddit's Acamedic Communities," *IEEE Access*, vol. 12, pp. 14932–14948, 2024, https://doi.org/10.1109/ACCESS.2024.3357662.

[11] A. Zafar, D. Aftab, R. Qureshi, Y. Wang, and H. Yan, "Multi-Explainable TemporalNet: An Interpretable Multimodal Approach using Temporal Convolutional Network for User-level Depression Detection," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2024, pp. 2258–2265, https://doi.org/10.1109/CVPRW63382.2024.00231.

[12] M. R. Febriansyah, Nicholas, R. Yunanda, and D. Suhartono, "Stress detection system for social media users," *Procedia Computer Science*, vol. 216, pp. 672–681, Jan. 2023, https://doi.org/10.1016/j.procs.2022.12.183.

[13] L. Ilias, S. Mouzakitis, and D. Askounis, "Calibration of Transformer-Based Models for Identifying Stress and Depression in Social Media," *IEEE Transactions on Computational Social Systems*, vol. 11, no. 2, pp. 1979–1990, Apr. 2024, https://doi.org/10.1109/TCSS.2023.3283009.

[14] "PHQ-9 Depression Assessment." Available: https://www.kaggle.com/datasets/thedevastator/phq-9-depression-assessment.

[15] Y. Yan and Q. Huang "Dataset of depression and anxiety among the elderly derived from The Nottingham Longitudinal Study of Activity and Ageing (NLSAA) project." https://doi.org/10.57760/sciencedb.06263.

[16] E. Turcan and K. McKeown, "Dreaddit: A Reddit Dataset for Stress Analysis in Social Media," in *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*, Hong Kong, China, Nov. 2019, pp. 97–107, https://doi.org/10.18653/v1/D19-6213.

[17] A. R. Salehi and M. Khedmati, "A cluster-based SMOTE both-sampling (CSBBoost) ensemble algorithm for classifying imbalanced data," *Scientific Reports*, vol. 14, no. 1, Mar. 2024, Art. no. 5152, https://doi.org/10.1038/s41598-024-55598-1.

[18] K. Liu, Y. Ye, S. Li, and H. Tang, "Accurate de novo peptide sequencing using fully convolutional neural networks," *Nature Communications*, vol. 14, no. 1, Dec. 2023, Art. no. 7974, https://doi.org/10.1038/s41467-023-43010-x.