IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15807

BLIP-Driven Multi-modal Deep Learning Framework for Temporal Depression Prediction from Social Media Posts

Dr Bharathi M P[1], Amulya M[2], Bhumika U[3]

Associate Professor, Surana College (Autonomous), Bengaluru, India[1]

Student, Surana College (Autonomous), Bengaluru, India[2]

Student, Surana College (Autonomous), Bengaluru, India[3]

Abstract: Depression remains one of the most widespread mental health conditions worldwide, and social media has become an unusually rich window into how people express emotional and behavioral change. Most existing detection systems however still look at a single channel either the wording of a post or a single accompanying image and largely ignore how a person's posting behavior shifts over time. This paper addresses that gap with a temporal, multi-modal pipeline that reads depression risk not from one post in isolation but from a sequence of a user's posts, combining what is written, what is pictured, the emotional tone of both, and how these change across time. Images are converted into descriptive captions using BLIP, which are then encoded alongside raw visual features and a set of behavioral and temporal signals processed through a GRU. When these three streams text, image, and temporal-behavioral are fused, the resulting model reaches 71% classification accuracy. The results support the view that behavioral and temporal cues, even when individually weak predictors, meaningfully sharpen a multi-modal system's ability to separate depressed from non-depressed posting patterns.

Keywords: Depression Detection; Social Media Analytics; Multi-modal Fusion; Temporal Behaviour Modelling; Image Captioning; Sentiment Analysis; BERT; GRU.

I. INTRODUCTION

Depression affects a substantial share of the global population the WHO estimates roughly 3.8% of people overall, with the figure rising to about 5.7% among adults over 60 [1]. At the same time, social media use has become close to universal, recent figures put active users at 4.62 billion worldwide, or roughly 58.4% of the global population, with the average person spending around seven hours per day online including two hours thirty minutes on social media platforms [2]. Because platforms like Twitter and Weibo function as informal, largely unfiltered outlets for emotion, they have become a practical through imperfect data source for studying mental health at scale. This pattern documented as early work analyzing depressive language in tweets [3].

Most prior systems built on this data source, however commit to a single modality: they analyze either the text of a post or a single image, rarely both, and almost never track how either change as a user keeps posting [4,5]. This is a real limitation, because the signal for depression in practice is rarely confined to one channel it shows up jointly in what someone writes, what they photograph or share, the emotional register of both, and the rhythm of when and how often they post. Prior multi-modal work has shown that combining these complementary signals measurably improves prediction accuracy over any single-channel baseline [6,7], which motivates the approach taken here. A framework that treats image, text, and behavioral signal not as competing inputs but as three views of the same underlying state, tracked over time rather than snapshot by snapshot.

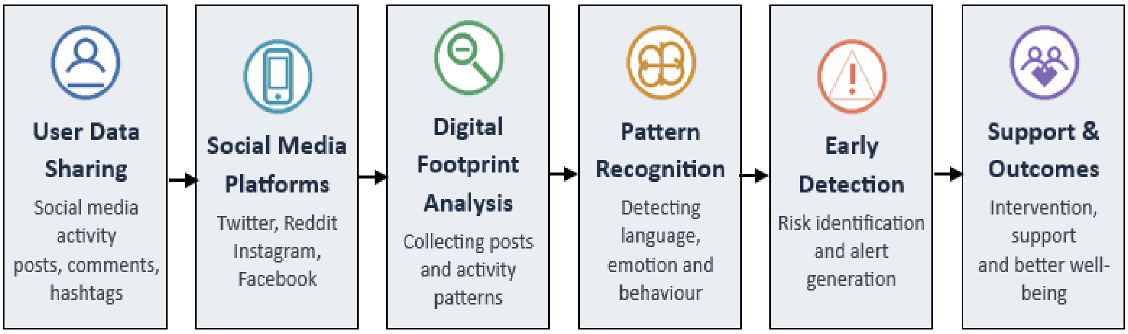

Figure 1: General Workflow of Social Media-Based Depression Detection

ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15807

Figure 1 sketches this idea at a high level: user activity across platforms such as posts, comments, and hashtags is mined for behavioral, linguistic, and emotional patterns, which then feed into an early-detection layer intended to support timely outreach rather than replace clinical judgment. Recent literature broadly agrees that deep learning CNN's, transformer encoders like BERT, and recurrent architectures like GRU is now capable of extracting these signals more reliably than earlier hand-engineered features, even though hand-crafted features do retain some ability to capture temporal and semantic structure in multi-modal posts [8]. This paper builds directly on that direction, combining BLIP, BERT, ResNet50, and GRU into a single temporal-multi-modal architecture built for depression screening.

II. LITERATURE REVIEW

Interest in applying AI spanning classical machine learning, NLP, and modern deep learning to social-media-based depression detection has grown steadily, largely because users tend to disclose feelings and attitudes openly on these platforms, providing usable signal without requiring a clinical interview. On people doing research in this area mostly used text. They made lists of words and phrases and then used old methods to look at them [9,10]. This showed that what people write on media can really tell us something about them but it also showed that these old methods had a big problem they did not look at pictures of how people behaved or when they made their posts.
Then some other researchers tried using methods, like Naive Bayes, Random Forest and SVM. These approaches were reasonably effective at surface-level pattern matching but structurally limited in capturing the more layered, indirect ways depression shows up in language. A recurring methodological gap in this literature is also worth noting: several studies reported accuracy alone, without F1-score, confusion matrices, or bias analysis [11,12] an omission that matters more in a medical-adjacent context than in ordinary NLP benchmark. The shift toward transformer encoders such as BERT, RoBERTa, DistilBERT, DeBERTa improved semantic representation considerably over these classical baselines [13,14], but introduced a different cost where these models are comparatively opaque, and the resulting lack of explainability is a recognized concern when the output feeds into anything resembling a mental-health judgment.
Beyond pure text, some work has folded in sentiment and behavioral signals with encouraging results [15], and related techniques have been extended to crisis detection, though not without raising legitimate ethical questions around consent and safeguards [16]. Other lines of research have pursued multi-modality more directly one notable approach fuses facial-expression features with LLM-derived features for hybrid detection [17], while another draws on contextual, audio-visual emotion analysis [18].
Taken together, the literature converges on two conclusions relevant to this work: multi-modal approaches consistently beat single-modality ones, and temporal information how posting patterns evolve remains comparatively under exploited even in recent multi-modal systems. This is precisely the gap the present framework is built to address, using BLIP, BERT, ResNet50, and GRU jointly to model image content, textual meaning, and temporal-behavioral dynamics within a single pipeline [19].

III. PROPOSED RESEARCH

The proposed model is designed for detection symptoms of depression in a set of posts from the user rather than just looking at one post. It involves analyzing visual content, textual, emotional and behavioral feature of these posts. Most of post that we deal with contains visual content. Therefore, we use the BLIP captioning model to describe this visual content. As we analyze emotions triggered by the use of certain words in the description of the visual content. All postings of each individual are combined together. They are sorted by their creation date starting from the oldest postings to the most recent. As a result, we can track changes in postings of each individual over time. From the posts, we gather information concerning descriptive language style, word count and emotions.
We also obtain temporal information such as creation time of a post. We have three tools that help us understand the posts. The first tool looks at the descriptions of the pictures. Turns them into a special kind of code that computers can understand. The second tool looks at the pictures themselves and gets information from them without reading the descriptions. The third tool looks at how the users posts change over time and uses this information to understand their behavior.
We combine the information, from these three tools. Then use it to guess if a user is depressed based on their posts. The posts and depression are what matter most to us. So, we look at the posts to see if they show sign of depression. We use the post sequence to do this because we want to know if someone's post indicate that they are depressed.
Figure 2 illustrates the proposed methodology is divided into two phases: Data Preparation and Feature Extraction. At the initial phase, the social media dataset is collected and pre-processed in order to clean the dataset for further use in the training of the machine learning models. The pre-processed dataset is divided into three different types such as image data, text data, and behavioral data. The captions for the images are extracted using the BLIP captioning model, while all the irrelevant data in the textual data are removed. Moreover, the behavioral data is pre-processed in order to generate

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15807

temporal features based on the patterns in user behaviour, such as posting frequency. Visual, textual and temporal features are extracted using ResNet50, BERT, and GRU.

At the second stage, the features are combined together to form fused multi-modal features. A classification model is then trained on these features to learn the relationships between image, text, and behavioral data and predict whether a given social media user has depressed class or non-depressed class. It is expected that the depression detection model will perform better when multi-modal features are used compared to uni-modal features.

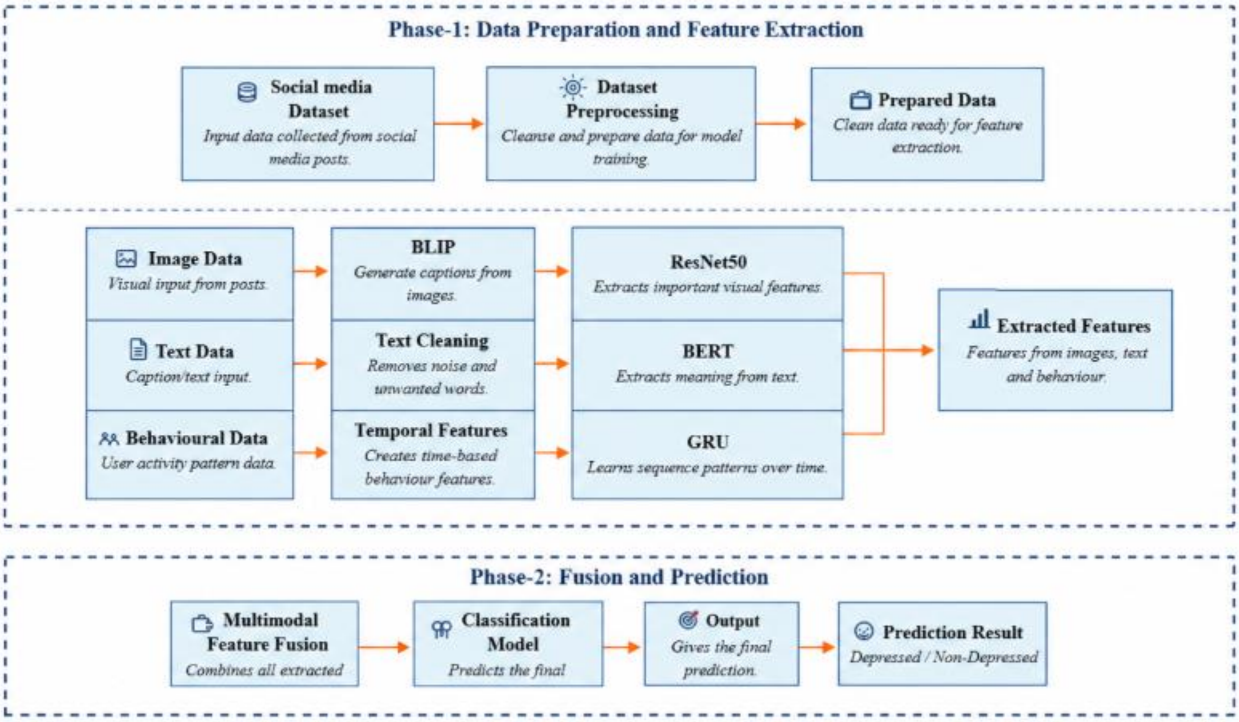

Figure 2: Workflow of the Proposed Multi-modal Depression Detection Framework

A. Dataset Description

The dataset is a Twitter-derived image-post collection sourced from Kaggle [20], containing images paired with associated text and depression-related labels. After filtering out invalid or incomplete records, 4,000 usable samples remained, split evenly between 2,000 depressed (label 1) and 2,000 non-depressed (label 0) cases.

TABLE 1: DATASET FIELDS USED IN THE PROPOSED FRAMEWORK

Data Field	Description
full_text	Original tweet text associated with the social media post.
label_text	Text label indicating non-depression or depression.
label_image	Image label indicating non-depression or depression.
kaggle_output_path	Path of the image file used for visual feature extraction.
created_at	Tweet creation timestamp used for temporal ordering.
favorite_count	Number of likes associated with the tweet.
Generated_Caption	Caption generated from the image using the BLIP model.
Sentiment	Sentiment category or score obtained from the generated caption.
User_ID	Pseudo user identifier used for building temporal post sequences.
Depression_Label	Binary class label: 0 = non-depressed, 1 = depressed.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15807

This dataset is a match for the task because people usually express their feelings through pictures just like they do through words. Each record is like a package that has a picture the meaning of the caption and some extra information, about what the user was doing when they posted it. The user IDs are not names they are just used to keep each user posts in order so we can see what they were doing over time. The user IDs do not reveal who the users really are.

B. Data Preprocessing

Raw images are first resized and normalized for compatibility with the downstream deep learning models, then passed through BLIP to generate a caption describing each image's visual content. Once captions are cleaned, sentiment is computed from the caption text, and this sentiment score is then folded into a broader set of behavioral features sentiment score, a confidence score, a negative-sentiment flag, caption length, word count, post order, the time gap since the prior post, and a rolling sentiment mean. These particular features were chosen because a single post rarely carries enough signal on its own to flag a depressive episode reliably it is typically the accumulation of multiple posts with a negative sentiment trend that carries diagnostic weight.

TABLE 2: SAMPLE OF THE GENERATED MULTIMODAL DATASET

	Image_Name	Generated_Caption	Emotional_Caption	Sentiment	User_ID	Timestamp	Depression_Label	Depression_Status
0	depresi_0000.jpg	the screen of a cell phone with a message on it	the screen of a cell phone with a message on it ...	NEGATIVE	DU050	2024-02-24 10:09:00	1	Depressed
1	depresi_0001.jpg	a woman is seen in the doorway of a building	a woman is seen in the doorway of a building c...	POSITIVE	DU007	2024-04-30 10:35:00	1	Depressed
2	nondepresi_0000.jpg	a black background with a white text that read...	a black background with a white text that read...	POSITIVE	NDU020	2024-03-03 02:14:00	0	Not-Depressed
3	nondepresi_0001.jpg	a street with a tree in the middle of it	a street with a tree in the middle of it,refle...	POSITIVE	NDU005	2024-03-16 14:55:00	0	Not-Depressed
4	depresi_0002.jpg	a cartoon drawing of a person crying in the rain	a cartoon drawing of a person crying in the ra...	NEGATIVE	DU006	2024-05-31 18:59:00	1	Depressed
5	depresi_0003.jpg	a cartoon picture of a person holding a phone	a cartoon picture of a person holding a phone,...	NEGATIVE	DU012	2024-02-12 19:50:00	1	Depressed

C. Temporal Sequence Construction

To build fixed-length input sequences for the GRU branch, each pseudonymous user's posts are sorted chronologically, and exactly five posts per user are used and users with fewer than five available posts are zero-padded to that length. This gives the GRU consistent, fixed-length sequences from which to learn sentiment and behavioral dynamics over time.

D. Multi-modal Fusion and Classification

Three input streams are sent to the classifier at the time: the temporal-behavioral feature vectors which is the behavioral branch the ResNet50 visual embeddings which is the picture branch and the caption embeddings which is the text branch. Image characteristics are moved to a space before being added to the temporal branch and padding masks stop the zero-padded inputs from changing the training. After the GRU layers capture the order of the text and behavioral streams the three branches are joined in a fusion layer. A final sigmoid layer gives the chance, between 0 and 1 that the post sequence's, from a sad user after this combined vector goes through dense layers that use dropout batch normalization and L2 penalties to stop overfitting.

IV. IMPLEMENTATION

The system is built using Python. It uses a lot of libraries like Pandas and NumPy. It also uses PIL, datetime, Matplotlib, TQDM, TensorFlow, Keras, Sentence Transformers and Scikit-learn. BLIP is used to generate captions. BERT and Sentence Transformers are used to encode text. ResNet50 is used to extract features is used to model time. All the experiments were run in a Jupyter Notebook on a machine with 8 GB RAM and an Intel i5 processor. This was done on purpose to make sure the system works outside of a lab.
The data is divided into groups based on the user so all the posts from one user are in one group, either train or validation or test. This is done to avoid a problem where the model learns to recognize the user of the depressive signal. If posts from the user are in both the training and test sets the model can learn the user and not the signal and this makes the

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15807

performance look better than it is. The data is split into 2,758 training samples and 631 validation samples and 611 test samples. There is no overlap of users, across the groups as shown in Table 3. The system uses Python and the libraries to make sure this works.

TABLE 3: GROUP-WISE EXPERIMENTAL SPLIT USED FOR MODEL TRAINING AND EVALUATION

Subset	Samples	Class Distribution	Users	User Overlap
Training	2758	0: 1427, 1: 1331	69	0
Validation	631	0: 330, 1: 301	16	0
Testing	611	0: 243, 1: 368	15	0

The training process uses the Adam optimizer. Additionally, the model also uses binary cross-entropy and label smoothing. The training process makes use of class weights to handle any class imbalance remaining. In order to avoid overfitting during the training process, we used early stopping, ReduceLROnPlateau, dropout, batch normalization and L2 regularization.
The classification threshold was not kept at 0.5 by default. Rather, the classification threshold was tuned using the validation set for obtaining the F1-score. The classification threshold of 0.32 was used for the test set without any modification.

TABLE 4: TRAINING CONFIGURATION OF THE PROPOSED MULTIMODAL TEMPORAL MODEL

Parameter	Value
Sequence length	5 posts
Text feature dimension	384
Image feature dimension	2048
Behaviour feature dimension	8
Optimizer	Adam
Learning rate	8e-5
Loss function	Binary cross-entropy with label smoothing
Batch size	16
Maximum epochs	25
Regularization	Dropout, batch normalization, L2 regularization
Callbacks	Early stopping, ReduceLROnPlateau, Model Checkpoint
Decision threshold	0.32 (selected using validation F1-score)

The output of the caption generator is processed by the BERT network to derive semantics and context from the generated captions to make the model aware of the meanings of the text. ResNet50, pre-trained on ImageNet, helps in deriving features from the image, which might point towards the existence of depressive behavior. The GRU model identifies temporal sequences by analyzing the changes in emotions and posting behavior of the user. The features derived from BERT, ResNet50, and GRU models are merged in the fusion layer to form one single feature vector. This combined feature vector is fed to a dense layer with a sigmoid activation function to classify the input as either Depressed or Not Depressed.

V. RESULTS AND DISCUSSION

Model performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, binary cross-entropy loss, the confusion matrix, and training/validation curves a broad set chosen deliberately to capture both overall correctness and, more specifically, how reliably the model catches genuinely depressed cases arguably the more consequential error direction in this setting.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⫶ Peer-reviewed & Refereed journal ⫶ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15807

TABLE 5: FINAL TEST PERFORMANCE OF THE PROPOSED MODEL

Metric	Value
Accuracy	0.71
Precision	0.72
Recall	0.84
F1-score	0.78
ROC-AUC	0.75
Best validation threshold	0.32

On the held-out test set, the model reached 71% accuracy, 72% precision, 84% recall, a 78% F1-score, and an ROC-AUC of 0.75. The gap between recall and precision, recall running notably higher indicates the model errors toward flagging borderline cases as depressed rather than missing them, which is the more defensible failure mode for a screening tool. A false positive prompts unnecessary but low-cost follow-up, while a false negative risks missing someone who needed support.

TABLE 6: MODEL EVALUATION FORMULAE

Metric	Formula	Purpose
Accuracy	(TP+TN) / (TP+TN+FP+FN)	Measures overall correct classification
Precision	TP / (TP/FP)	Measures correctness among predicted depressed samples
Recall	TP / (TP+FN)	Measures how many actual depressed samples are identified
F1-score	2 x (Precision x Recall) / (Precision + Recall)	Balanced precision and recall
ROC-AUC	$\int_0^1 TPR(FPR)\, d(FPR)$	Measures discrimination across decision threshold
Best validation threshold	$-\frac{1}{N}\sum_{i=1}^{N}[\, y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)\,]$	Optimizes binary classification during training

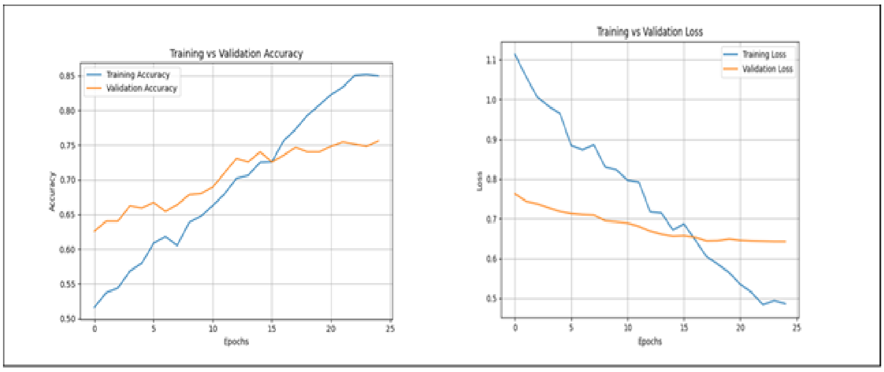

Figure 3: Training and Validation Performance Analysis

Figure 3 tracks training and validation curves across epochs: training accuracy climbs steadily throughout, while validation accuracy levels of around 75%, and both training and validation loss trend downward before flattening, a pattern consistent with the regularization scheme controlling overfitting without stalling learning.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15807

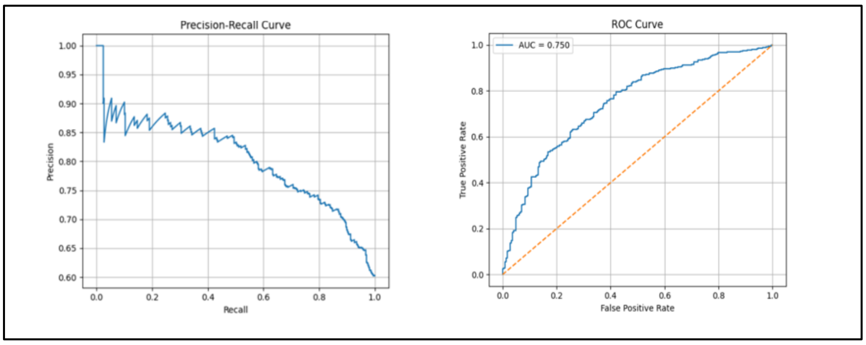

Figure 4: Precision-Recall and ROC Curve Analysis

Figure 4 demonstrates the Precision- Recall and ROC curves of the proposed approach. According to the Precision-Recall curve, the proposed classifier shows a high level of precision for lower levels of recall, which is followed by a gradual decrease in precision when the recall increases, meaning the existence of the precision-recall trade-off relationship. The ROC curve is placed above the diagonal line, which means a sufficient discrimination capacity of the model for separating the depressed from the non-depressed cases. The AUC score of 0.75 additionally proves this statement.

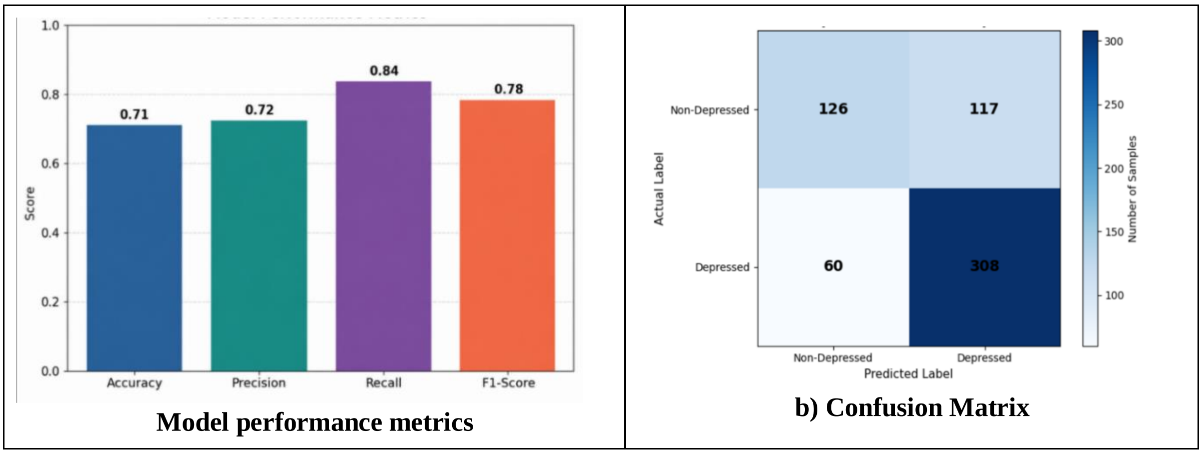

Figure 5: Model Evaluation Results

Figure 5 gives the evaluation metrics and confusion matrix of the proposed framework. Recall at 0.84 was attained by the model, which suggests the high capability of the model in identifying the depressed samples. An F1-score at 0.78 is an indication of a well-balanced model in terms of precision and recall. From the confusion matrix, it can be seen that the model correctly identifies 308 depressed and 126 non-depressed samples but commits 60 false negatives and 117 false positives. False positives may arise from emotionally expressive samples.

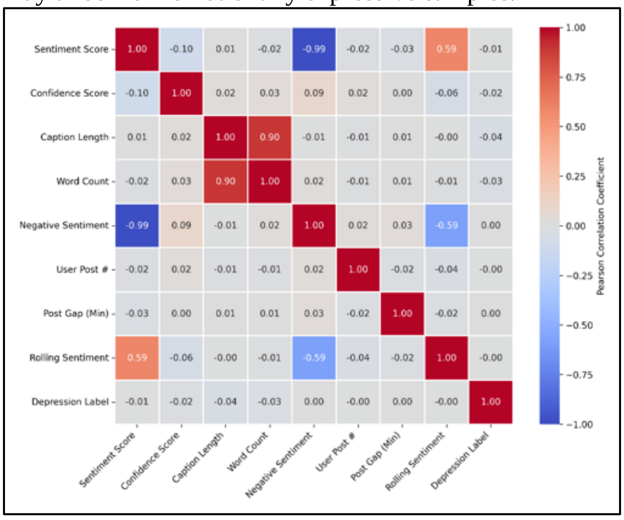

Figure 6: Behavioral Feature Correlation Analysis

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪿ Peer-reviewed & Refereed journal ⪿ Vol. 15, Issue 8, August 2026

DOI: 10.17148/IJARCCE.2026.15807

Figure 6 illustrates behavioral feature correlation heat map indicates that caption length and word count are strongly correlated, while sentiment score and negative sentiment show a strong inverse relationship. Most behavioral features show weak direct correlation with the depression label. This supports the use of multi-modal fusion, because no single behavioral feature alone is sufficient for reliable classification.

TABLE 7: MODEL COMPARISON USING INDIVIDUAL AND FUSED MODALITIES

Model	Features Used	Accuracy	Precision	Recall	F1-score	ROC-AUC
BERT Model	Text features only	0.66	0.70	0.78	0.74	0.68
ResNet50 Model	Image features only	0.68	0.75	0.71	0.73	0.73
GRU Model	Behavioral features only	0.59	0.60	0.67	0.74	0.53
Proposed Multimodal Temporal GRU Model	Image + Text + Behavioral Features	0.71	0.72	0.84	0.78	0.75

The proposed multi-modal framework achieved better performance than the individual text-only, image-only, and behaviour-only models by combining supporting details from all three modalities. While the text and image models captured semantic and visual features, the behavioral features contributed additional temporal context that improved the overall prediction. Although the proposed framework produced encouraging results, its performance is limited by the use of generated captions and evaluation on a single dataset. Future work will focus on validating the model on larger and more diverse datasets while incorporating explainability techniques to improve interpretability. Since the framework analyses sensitive social media data, it should be used only as a screening support tool and not as a substitute for clinical diagnosis.

VI. CONCLUSION

In this paper, we have introduced a temporal multi-modal framework for depression detection based on social media data, through the use of vision, language, affective, and behavioral features for the first time. It was proved that the integration of BLIP, BERT, ResNet50, sentiment analysis, and GRU enables obtaining better results than each model in isolation. The accuracy of 71% was achieved for the considered dataset. It was found that a temporal component should be included when performing depression detection from social media based on multimodal data. The future work will focus on validating the framework using large real-world datasets and advanced deep learning models.

REFERENCES

[1] Predicting Depression Severity Using Multimodal Functions, Cornell University AI and more, 2017.
[2] Digital 2022: Global Overview Report, DataReportal, 2022. Available: https://datareportal.com/reports/digital-2022-global-overview-report
[3] M. Park, C. Cha, and M. Cha, "Depressive Moods of Users Portrayed in Twitter," in Proc. 18th ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD '12), New York, NY, USA, 2012.
[4] T. Baltrušaitis, C. Ahuja, and L. P. Morency, "Multimodal Machine Learning: A Survey and Taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2019.
[5] Y. Zhang et al., "A Depression Detection Method Based on Multi-Modal Feature Fusion Using Cross-Attention," arXiv preprint, arXiv:2407.12825, 2024.
[6] L. He et al., "ContextVecNet: A Context-Driven Multimodal Learning Framework for Depression Detection," IEEE J. Biomed. Health Inform., vol. 29, no. 9, pp. 5446–5457, 2025.
[7] Y. Wang et al., "A Sentiment Pre-Trained Text-Guided Multimodal Cross-Attention Transformer for Improved Depression Detection," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), 2024.
[8] A. Rajawat, S. Bedi, S. B. Goyal, P. Bhaladhare, A. Aggarwal, and R. Singhal, "Fusion Fuzzy Logic and Deep Learning for Depression Detection Using Facial Expressions," Procedia Comput. Sci., vol. 218, pp. 2795–2805, 2023.
[9] R. Wang et al., "Contextual Interaction-Based Multimodal Emotion Analysis with Enhanced Semantic Information," IEEE Trans. Comput. Social Syst., vol. 1, pp. 1–11, 2025.
[10] M. Kerasiotis, L. Ilias, and D. Askounis, "Depression Detection in Social Media Posts Using Transformer-Based Models and Auxiliary Features," Soc. Netw. Anal. Min., vol. 14, p. 196, 2024.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 8, August 2026
DOI: 10.17148/IJARCCE.2026.15807

[11] A. Yadav, A. Tiwari, A. Kumar, and A. K. Verma, "Detecting Mental Health Issues Through Social Media Using NLP and Machine Learning," IJRASET, 2025.

[12] B. Hadzic et al., "Enhancing Early Depression Detection with AI: A Comparative Use of NLP Models," SICE J. Control Meas. Syst. Integration, 2024.

[13] S. Hegde and M. Kumar B., "Depression Detection Using Machine Learning," IJERT, 2023.

[14] G. Coppersmith, R. Leary, P. Crutchley, and A. Fine, "Natural Language Processing of Social Media as Screening for Suicide Risk," Biomed. Inform. Insights, vol. 10, 2018.

[15] S. Anbukkarasi, S. Jeevapriya, A. Kaaviyaa, T. Lawvanyapriya, and S. Monisha, "Deep Learning Based Depression Detection from Social Media Text," Appl. Comput. Eng., 2023.

[16] B. G. Bokolo and Q. Liu, "Deep Learning-Based Depression Detection from Social Media: Comparative Evaluation of ML and Transformer Techniques," Electronics, vol. 12, no. 21, p. 4396, 2023.

[17] N. Goel, R. Anand, A. Amoli, K. Chuki, P. Sharma, and S. Harnal, "Automated Depression Detection System: Integrating Sentiment Analysis and Behavioral Data," IJERT, 2024.

[18] M. A. Mansoor and K. H. Ansari, "Early Detection of Mental Health Crises Through Artificial-Intelligence-Powered Social Media Analysis," J. Pers. Med., vol. 14, no. 9, p. 958, 2024.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.

[20] Twitter Depression Detection Dataset, Kaggle, 2026. Available: https://www.kaggle.com/datasets/wisnusatrio/twitter-depression-v2