International Journal of Scientific Modern Research and Technology
(Volume: 25, Issue: 01, Number: 04) Paper ID: IJSMRT - 26100104
ISSN: 2582-8150

# An Efficient Hybrid Deep Learning Framework for Depression-Related Text Classification Using Social Media Data

*Aklesh Panche[1], Anurag Shrivastava[2], Dr. Puran Gour[3]*
*[1]Research Scholar, Department of Computer Science & Engineering, NIIST, Bhopal (M.P.), India*
*[2]Associate Professor, Department of Computer Science & Engineering, NIIST, Bhopal (M.P.), India*
*[3]Principal NIIST, Bhopal (M.P.), India*

*Abstract: Depression-related content on social media offers an opportunity to study language patterns associated with mental health, but accurate text classification remains challenging. This paper proposes a hybrid deep learning framework combining DistilBERT, Bidirectional Long Short-Term Memory (BiLSTM), and an attention mechanism for depression-related text classification. DistilBERT generates contextual representations, while BiLSTM captures sequential relationships between tokens and attention emphasizes informative features. The model was evaluated on a labelled social media text dataset and compared with TF-IDF-based Logistic Regression and Support Vector Machine baselines. The proposed model achieved 95.82% accuracy, 95.78% precision, 95.89% recall, and a 95.84% F1-score on 10,000 test samples. These results indicate promising classification performance, although the dataset labels are not clinical diagnoses. Testing on diverse datasets is needed.*

*Keywords: Depression Detection, Social Media Text Classification, DistilBERT, BiLSTM, Attention Mechanism, Hybrid Deep Learning.*

*How to cite this article: Aklesh Panche, Anurag Shrivastava and Puran Gour. (2026). An Efficient Hybrid Deep Learning Framework for Depression-Related Text Classification Using Social Media Data, International Journal of Scientific Modern Research and Technology (IJSMRT), ISSN: 2582-8150, Volume-25, Issue-01, Number 04, Oct-2026, pp. 18-21, URL: https://www.ijsmrt.com/wp-content/uploads/2026/11/IJSMRT-26100104.pdf*

*Copyright © 2026 by author (s) and International Journal of Scientific Modern Research and Technology Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0)*

*(http://creativecommons.org/licenses/by/4.0/)*

IJSMRT-26100104

## I. INTRODUCTION

Depression is a common mental health condition that can affect an individual's emotions, thoughts, behaviour, and daily functioning. Early recognition of depression-related symptoms is important because delayed support may increase personal and social difficulties. However, conventional assessment generally depends on clinical interviews, questionnaires, and professional evaluation, which may not always be readily accessible. In recent years, social media platforms have become important spaces where people share personal experiences, emotions, opinions, and daily concerns. The language used in these posts may contain patterns associated with depressive symptoms, creating opportunities for natural language processing (NLP) techniques to support research on depression-related text classification. Such computational approaches are intended to identify patterns in text and should not be considered substitutes for clinical diagnosis.

The growing availability of user-generated content has encouraged researchers to investigate machine learning and deep learning methods for analysing mental-health-related language. Traditional text classification approaches commonly use feature extraction methods such as Term Frequency–Inverse Document Frequency (TF-IDF), followed by classifiers including Logistic Regression and Support Vector Machine (SVM). Although these methods provide useful baselines, they may have difficulty representing contextual meaning, word order, and complex linguistic relationships. Transformer-based language models address some of these limitations by learning contextual representations from large text corpora. Shen and Paik [1] investigated temporal modelling approaches for depression forecasting using pretrained embeddings, highlighting the relevance of contextual and temporal information in social media analysis. Villa-Pérez et al. [2] examined the impact of temporal boundaries and text representations on depression classification, demonstrating the importance of experimental design

when evaluating social media data. Recent studies have also explored sequential and advanced language modelling techniques. Farruque et al. [3] studied deep temporal modelling of clinical depression through social media text, while their later work [4] investigated an LLM-driven semi-supervised approach for modelling depression symptoms. These studies indicate the potential of deep language representations and learning strategies for analysing depression-related content. Bao et al. [5] focused on explainable depression symptom detection, underlining the value of making model predictions more interpretable. In addition, Sadeghi et al. [6] explored multimodal depression detection by combining large language models and facial expressions, illustrating the broader movement towards integrating different information sources.

Motivated by these developments, this paper proposes a hybrid deep learning framework for depression-related text classification using social media data. The framework combines DistilBERT, Bidirectional Long Short-Term Memory (BiLSTM), and an attention mechanism. DistilBERT generates contextual token representations, BiLSTM captures relationships across the token sequence in both forward and backward directions, and attention assigns greater importance to informative features before classification. The proposed model is evaluated against TF-IDF-based Logistic Regression and SVM baselines using a labelled social media text dataset. On the held-out test set of 10,000 samples, the model achieves 95.82% accuracy, 95.78% precision, 95.89% recall, and a 95.84% F1-score. These findings suggest that combining pretrained contextual representations with sequential learning and attention can provide an effective approach to this classification task. The study focuses on text-level classification rather than longitudinal user-level forecasting, and its dataset labels are treated as research labels rather than verified clinical diagnoses. Further validation across datasets remains necessary.

## II. LITERATURE REVIEW

Clinical depression detection through social media has emerged as a promising approach for enabling early mental health assessment and timely intervention. Recently, deep learning and temporal modeling techniques have attracted considerable attention due to their ability to capture both linguistic patterns and the temporal evolution of users' online behavior, leading to more accurate and reliable depression detection.

Author [1] proposed a deep learning-based temporal modeling framework for depression forecasting using social media posts. The study combined textual representations with personalized temporal features to capture changes in users' emotional states over time. Experimental results showed that integrating temporal dynamics with Transformer-based embeddings significantly improved prediction performance, demonstrating the effectiveness of temporal modeling for accurate depression forecasting.

Author [2] investigated the impact of temporal segmentation of social media timelines for early depression detection. The study analyzed user posts before and after self-reported depression diagnoses using multiple textual representations across English and Spanish datasets. Experimental results demonstrated that temporal segmentation improves classification performance and enhances the generalizability of depression detection models across different languages and datasets.

Author [3] presents a temporal framework for user-level clinical depression detection using social media posts. The proposed approach employs a Depression Symptoms Detection (DSD) classifier to identify clinically relevant symptoms and temporal behavioral patterns, including posting activity and periods of silence. Experimental results demonstrate that semantic representations combined with clinically derived features improve depression prediction, particularly when sufficient user timeline data and similar training–testing distributions are available.

Author [4] proposed a semi-supervised learning framework to enhance Depression Symptoms Detection (DSD) from social media text by combining a fine-tuned pre-trained language model with a zero-shot learning approach. The framework iteratively expanded a clinician-annotated dataset using depression-related tweets collected from self-disclosed depressed users. Experimental results showed that the enlarged dataset significantly improved the performance of both depression symptom detection and depression post detection models, providing a more robust foundation for clinical depression analysis.

Author [5] introduced transformer-based models for clinical depression detection that not only identify users at risk but also provide interpretable symptom-based explanations. The study explored separate and joint learning strategies for depression classification and symptom explanation, while also evaluating conversational Large Language Models (LLMs). Experimental findings demonstrated that the proposed approaches achieved strong detection performance along with clinically meaningful and explainable predictions, enhancing their usefulness for mental health professionals.

Author [6] proposed an automated framework for predicting depression severity by leveraging Large Language Models (LLMs) and multimodal analysis. The study extracted depression-related features from

interview transcripts and combined them with facial information to estimate PHQ-8 depression scores. Experimental results demonstrated that text-based features, particularly when enhanced with speech quality assessment, achieved the highest prediction accuracy, highlighting the effectiveness of LLMs for automated depression severity assessment.

## III. PROPOSED METHODOLOGY

The proposed model combines DistilBERT, Bidirectional Long Short-Term Memory (BiLSTM), and an attention mechanism to perform depression-related text classification. Initially, social media text is preprocessed by removing unnecessary URLs, HTML tags, and extra spaces. The cleaned text is then converted into token IDs and attention masks using the DistilBERT tokenizer. DistilBERT, a lightweight pretrained Transformer model, generates contextual representations of the input tokens. These representations are passed to the BiLSTM layer, which processes the token sequence in both forward and backward directions. This enables the model to capture sequential relationships and contextual information from both sides of each token.

Next, the attention mechanism assigns different weights to the BiLSTM output features according to their importance. The weighted features are combined into a single representation using attention pooling. A dropout layer is applied to reduce overfitting, followed by a fully connected layer that generates classification scores. Finally, the softmax function predicts one of two output classes: Class 0 or Class 1. The model is trained using labelled social media text samples and Cross-Entropy Loss with the AdamW optimizer. Validation loss is monitored to select the best-performing model. This hybrid architecture aims to improve contextual understanding and classification performance by integrating pretrained language representations, sequential feature learning, and attention-based feature selection.




Figure 1: Proposed Architecture Framework

*Dataset Used:* This work uses the Depression Detection dataset, thePixel42/depression-detection, available on Hugging Face. The dataset contains approximately 200,000 Reddit-based social media posts with two attributes: text and label. It is designed for binary classification, where labels 0 and 1 represent the two target categories. The dataset was divided into 140,000 training samples and 60,000 testing samples. For computational efficiency, 20,000 training samples and 10,000 testing samples were selected for experimentation. Additionally, 15% of the training samples were used for validation.

## IV. RESULT ANALYSIS

The experimental results demonstrate that the proposed DistilBERT-BiLSTM-Attention model outperforms the traditional machine learning baselines for depression-related text classification. As presented in the comparison, TF-IDF with Logistic Regression achieved 91.20% accuracy, 92.78% precision, 89.42% recall, and a 91.07% F1-score, with an ROC-AUC of 0.9693. TF-IDF with Support Vector Machine obtained 91.10% accuracy, 92.00% precision, 90.10% recall, and a 91.04% F1-score. Its ROC-AUC value was reported as NaN, so this metric could not be used for comparison. In contrast, the proposed hybrid model achieved the highest performance, recording 95.82% accuracy, 95.78% precision, 95.89% recall, and 95.84% F1-score, together with an ROC-AUC of 0.9917.

Compared with Logistic Regression, the proposed model improves accuracy by 4.62 percentage points and F1-score by approximately 4.77 points. The improvement over SVM is 4.72 percentage points in accuracy and approximately 4.80 points in F1-score. These gains suggest that contextual representations generated by DistilBERT, sequential information captured by BiLSTM, and attention-based feature weighting provide complementary benefits. The consistently high precision and recall also indicate a more balanced classification performance than the TF-IDF baselines. Overall, the findings support the effectiveness of the proposed architecture on the evaluated test dataset. However, these results reflect dataset-specific classification performance and should not be interpreted as clinical diagnosis. Further validation on independent and diverse datasets and different user populations is required to assess generalizability.

Table 1: Performance Comparison of Baseline Model with Proposed Model

| Model | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| TF-IDF + Logistic Regression | 0.912 | 0.9278 | 0.9278 | 0.9106 |
| TF-IDF + SVM | 0.911 | 0.92 | 0.92 | 0.9103 |
| Proposed DistilBERT-BiLSTM-Attention | 0.9582 | 0.9578 | 0.9589 | 0.9583 |

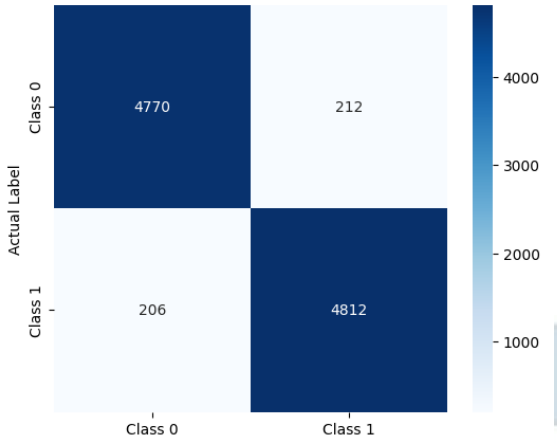



Figure 2: Proposed Model Confusion Matrix

## V. CONCLUSION

This study presents a hybrid DistilBERT-BiLSTM-Attention framework for depression-related text classification using social media data. The proposed model combines contextual language representations, sequential feature learning, and attention-based feature selection to improve classification performance. Experimental results demonstrate that the proposed approach achieves 95.82% accuracy, 95.78% precision, 95.89% recall, and a 95.84% F1-score, outperforming TF-IDF-based Logistic Regression and SVM models. These findings highlight the potential of hybrid deep learning techniques for identifying depression-related language patterns. However, further evaluation using diverse datasets is necessary to improve generalization and ensure reliable performance. The model is intended for research purposes, not clinical diagnosis.

## References

[1] Zheqi Shen and Incheon Paik "Temporal Modeling of Social Media for Depression Forecasting: Deep Learning Approaches with Pretrained Embeddings" Appl. Sci. 2025, 15, 11274 https://doi.org/10.3390/app152011274

[2] Miryam Elizabeth Villa-Pérez et al. "A comparative and statistical analysis of depression classification in social media: assessing the impact of temporal boundaries and text representations" Social Network Analysis and Mining (2026) 16:44 https://doi.org/10.1007/s13278-026-01589-1

[3] Nawshad Farruque et al. "Deep temporal modelling of clinical depression through social media text" Natural Language Processing Journal 6 (2024) 100052, https://doi.org/10.1016/j.nlp.2023.100052

[4] Nawshad Farruque et al. "Depression symptoms modelling from social media text an LLM driven semi-supervised learning approach" Language Resources and Evaluation (2024) 58:1013–1041, https://doi.org/10.1007/s10579-024-09720-4

[5] Eliseo Bao et al. "Explainable depression symptom detection in social media" Health Information Science and Systems (2024) 12:47, https://doi.org/10.1007/s13755-024-00303-9

[6] Misha Sadeghi et al. "Harnessing multimodal approaches for depression detection using large language models and facial expressions" https://doi.org/10.1038/s44184-024-00112-8, Mental Health Research 2024

[7] World Health Organization. Depression. https://www.who.int/newsroom/fact-sheets/detail/depression. Accessed on 3 June 2024.

[8] Woody, C., Ferrari, A., Siskind, D., Whiteford, H. & Harris, M. A systematic review and meta-regression of the prevalence and incidence of perinatal depression. J. Affect. Disord. 219, 86–92 (2017).

[9] Evans-Lacko, S. et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the who world mental health (wmh) surveys, Psychol. Med. 48, 1560–1571 (2018)

[10] Kroenke, K. et al. The phq-8 as a measure of current depression in the general population. J. Affect. Disord. 114, 163–173 (2009).

[11] American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Publishing, Washington, DC, 2013), 5th edn.

[12] thePixel42, "Depression Detection Dataset," Hugging Face Datasets. Available: https://huggingface.co/datasets/thePixel42/depression-detection