

Early Behavioral Change Prediction on Social Media Through an Optimized Hybrid Deep Learning Ensemble Model

Murlidhar Prasad Singh¹, Md Ehtashamoul Haque², Gopal Krishna³, Kamal Raj Praveen⁴

¹Assistant professor, Department of Computer Science and Engineering B. P. Mandal College of Engineering, Madhepura

²Assistant Professor, Department of computer Science and Engineering B.P.Mandal College of Engineering, Madhepura

³Assistant professor, Department of Computer Science and Engineering Supaul College of Engineering, Supaul

⁴Assistant Professor, Department of computer science and engineering Supaul College of Engineering, Supaul

Abstract

The unchecked growth of social media usage has made platforms a rich but noisy source of behavioural signals that, if decoded early, can help identify psychological distress, radicalisation tendencies, and other harmful behaviour shifts before they escalate. This paper proposes a Stacked Classification Approach built on an Optimized Hybrid Deep Learning Model for the early prediction of behaviour changes from social media activity. The framework first extracts contextual textual embeddings and behavioural meta-features from user posts, then fuses them through a Convolutional Neural Network-Bidirectional Long Short-Term Memory-Attention (CNN-BiLSTM-Attention) block that captures both local linguistic patterns and long-range temporal dependencies. An Improved Grey Wolf Optimizer (I-GWO) tunes the hybrid network's hyperparameters to avoid manual trial-and-error and to escape local optima. The deep feature representation is then passed through a stacked ensemble in which CNN-BiLSTM-Attention, Random Forest, and XGBoost act as base learners and a Logistic Regression meta-learner combines their outputs into a final prediction. Experiments on a curated dataset of 39,940 annotated social media posts show that the proposed model achieves 96.8% accuracy, 96.1% precision, 95.9% recall, and a 96.0% F1-score, outperforming standalone CNN, LSTM, CNN-LSTM, and CNN-BiLSTM-Attention baselines by a margin of 4.7 to 15.4 percentage points. The results confirm that combining optimized hybrid deep feature learning with stacked ensemble classification substantially improves the early and reliable detection of behavioural change on social media.

Keywords: Behaviour Change Prediction; Hybrid Deep Learning; Stacked Classification; CNN-BiLSTM-Attention; Grey Wolf Optimizer; Social Media Analytics.

1. Introduction

Social media platforms generate an unprecedented volume of user-generated content that reflects, and often anticipates, real-world behavioural change. Subtle shifts in posting frequency, tone, vocabulary, and interaction patterns frequently precede visible behavioural or psychological transitions, including withdrawal, aggression, and self-harm risk (Kumar & Singh, 2023). Because such changes tend to unfold gradually before becoming clinically or socially apparent, computational systems capable of recognising early warning signals carry significant value for mental-health screening, cyberbullying prevention, and public-safety monitoring (Alharbi & Alghamdi, 2021).

Traditional machine learning classifiers such as Support Vector Machines and Naïve Bayes have been widely applied to sentiment and behaviour classification tasks, but their reliance on hand-crafted features limits their ability to capture the sequential and contextual nuances

of natural language (Chen, Wang, & Li, 2022). Deep learning architectures, particularly Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have improved representation learning by automatically extracting local and sequential patterns from text; however, single-architecture models still struggle to jointly capture short-range lexical cues and long-range behavioural trends, and their performance is highly sensitive to hyperparameter choices (Patel, Mehta, & Rao, 2024).

Hybrid deep learning models that combine CNNs with recurrent structures have shown promise in bridging this gap, and attention mechanisms further allow the network to focus on the most behaviourally salient tokens within a post (Zhao & Liu, 2023). Nevertheless, most existing studies rely on a single classifier at the decision stage, which discards the complementary strengths that different learners bring to heterogeneous, imbalanced behavioural data. Ensemble strategies, and stacked generalisation in particular, have been shown to reduce variance and improve generalisation by learning how to optimally combine multiple base classifiers rather than trusting any single model in isolation (Devi & Ramesh, 2022).

A further challenge in this domain is the highly imbalanced and noisy nature of behavioural data: instances signalling genuine risk are comparatively rare relative to routine activity, informal language, sarcasm, and code-mixed expressions further obscure lexical signals, and behavioural change frequently emerges only when isolated posts are considered jointly with a user's temporal activity pattern rather than in isolation. Systems that classify individual posts without modelling this temporal context tend to raise both false positives, which erode user trust and place unnecessary review burden on moderators, and false negatives, which delay intervention at the point it is most needed. Addressing these issues requires a pipeline that can simultaneously represent short-range linguistic cues, long-range behavioural trends, and decision-level robustness against noisy individual predictions.

Motivated by these observations, this paper proposes a stacked classification framework driven by an optimized hybrid deep learning model for early prediction of behaviour changes on social media. The key contributions are: (i) a CNN-BiLSTM-Attention hybrid module that jointly models local semantic patterns and long-range temporal behavioural dependencies; (ii) an Improved Grey Wolf Optimizer for automated hyperparameter tuning of the hybrid network, removing the need for exhaustive manual search; (iii) a stacked ensemble layer that fuses the hybrid deep learner with Random Forest and XGBoost base learners through a Logistic Regression meta-learner; and (iv) a comprehensive evaluation against standalone deep learning baselines demonstrating consistent improvement across accuracy, precision, recall, and F1-score. The remainder of this paper is organised as follows: Section 2 reviews related work on behavioural classification, hybrid deep learning, and stacked ensembles; Section 3 details the proposed methodology; Section 4 describes the experimental setup; Section 5 presents and discusses the results; and Section 6 concludes the paper with directions for future work.

2. Related Work

Early work on behavioural analysis of social media text relied primarily on lexicon-based sentiment scoring and shallow classifiers. Alharbi and Alghamdi (2021) used TF-IDF features with SVM to detect depressive language in Twitter posts, reporting moderate accuracy but limited robustness to informal language and sarcasm. Kumar and Singh (2023) extended this line of work with an LSTM-based sequence model for early depression

detection, showing that temporal modelling of posting history improves detection lead-time compared to single-post classification.

Convolutional architectures have also been explored for behaviour and emotion classification. Chen, Wang, and Li (2022) proposed a CNN-based emotion classifier over word embeddings that captured local n-gram sentiment cues effectively, but the model's fixed receptive field limited its ability to represent long-range dependencies across a user's post history. To address this, Zhao and Liu (2023) combined CNN and BiLSTM layers with a self-attention mechanism for cyberbullying detection, achieving notable gains over CNN-only and LSTM-only baselines, and confirming that attention-weighted hybrid architectures better localise behaviourally relevant segments of text.

Parallel to architectural improvements, metaheuristic optimisation has been used to tune deep network hyperparameters instead of exhaustive grid search. Patel, Mehta, and Rao (2024) applied a Particle Swarm Optimization strategy to tune a CNN-LSTM model for suicidal-ideation detection, reporting faster convergence and improved F1-score relative to manually tuned networks. Grey Wolf Optimizer variants have similarly been applied to feature selection and hyperparameter search problems due to their strong balance between exploration and exploitation and comparatively low computational overhead (Fathi & Nasr, 2022).

On the ensemble side, Devi and Ramesh (2022) demonstrated that stacking heterogeneous classifiers, such as Random Forest, Gradient Boosting, and neural feature extractors, consistently outperforms any individual base learner on imbalanced social-media classification tasks, primarily because the meta-learner compensates for the base learners' complementary error patterns. Similarly, Hussain and Ahmed (2023) reported that stacked ensembles built on deep feature embeddings achieved better generalisation on cross-platform behaviour datasets than deep learning models used alone.

Table-level comparison of these studies reveals a consistent pattern: works that improve feature extraction (Chen et al., 2022; Zhao & Liu, 2023) generally retain a single-classifier decision head, while works that improve decision fusion (Devi & Ramesh, 2022; Hussain & Ahmed, 2023) generally operate on relatively simple hand-crafted or shallow embeddings rather than an optimised hybrid deep representation. Studies that incorporate metaheuristic optimisation (Fathi & Nasr, 2022; Patel et al., 2024) tend to apply it narrowly to a single architecture rather than to a full extraction-plus-fusion pipeline. This fragmentation means that, individually, none of the above approaches jointly addresses local-versus-long-range feature representation, automated hyperparameter search, and robust multi-learner decision fusion within one coherent framework.

Despite these advances, most prior studies address either hybrid deep feature extraction or ensemble-based decision fusion in isolation, and hyperparameter optimisation is frequently omitted or performed manually. Few works combine an optimized hybrid deep learning feature extractor with a stacked classification layer specifically for early behaviour-change prediction, which is the gap this study addresses by unifying dual-branch feature extraction, I-GWO-based optimisation, and stacked ensemble classification into a single end-to-end pipeline.

3. Proposed Methodology

The proposed framework consists of five stages: data pre-processing, dual-branch feature extraction, hybrid deep feature learning, hyperparameter optimisation, and stacked classification. Figure 1 presents the overall architecture, and each stage is detailed below.

3.1 Data Pre-processing

Raw social media posts are cleaned by removing URLs, HTML tags, and non-informative symbols. Text is tokenised, stop words are removed, and remaining tokens are lemmatised. Emojis are mapped to their textual sentiment equivalents so that affective cues are not discarded. Because behavioural datasets are naturally imbalanced toward the normal class, the Synthetic Minority Over-sampling Technique (SMOTE) is applied at the training stage to balance the at-risk and critical classes without duplicating raw samples.

3.2 Dual-Branch Feature Extraction

Two complementary feature streams are constructed. The textual branch encodes each post using pre-trained Word2Vec/GloVe embeddings passed through a trainable embedding layer, capturing semantic relationships between words. The behavioural branch derives numerical meta-features from user activity, including posting frequency over a rolling window, an aggregated sentiment score, and an engagement ratio computed from likes, comments, and shares. Modelling behaviour purely from isolated post content overlooks temporal usage patterns, so the behavioural branch supplies this complementary context.

3.3 Hybrid Deep Feature Learning (CNN-BiLSTM-Attention)

The two feature streams are concatenated and passed into the hybrid deep learning block. A one-dimensional CNN layer first extracts local n-gram patterns indicative of short bursts of emotional language. The resulting feature maps are fed into a Bidirectional LSTM layer, which models both forward and backward temporal dependencies across the user's post sequence, allowing the network to relate an early subtle cue to a later, more explicit one. An attention layer then assigns learned weights to each time step, emphasising the segments most relevant to behavioural change while suppressing noise, before the weighted representation is flattened into a dense feature vector.

3.4 Hyperparameter Optimisation via Improved Grey Wolf Optimizer

Manually selecting the number of CNN filters, BiLSTM units, learning rate, and dropout rate is time-consuming and prone to sub-optimal local choices. The proposed framework therefore employs an Improved Grey Wolf Optimizer (I-GWO), which augments the classical Grey Wolf Optimizer with an adaptive convergence factor and an opposition-based initialisation strategy to improve exploration diversity and avoid premature convergence. Each wolf in the population encodes a candidate hyperparameter configuration, and the validation F1-score is used as the fitness function. Over successive iterations, the alpha, beta, and delta wolves guide the population toward the configuration that maximises validation performance, after which the best configuration is used to train the final hybrid network.

3.5 Stacked Classification Layer

Rather than relying on a single soft-max output from the hybrid network, the dense feature vector produced by the optimized CNN-BiLSTM-Attention block is used as input to a stacked ensemble. Three base learners are trained on this shared representation: the CNN-BiLSTM-Attention model's own classification head, a Random Forest classifier, and an XGBoost classifier. Each base learner captures different decision boundaries — the deep learner models non-linear sequential interactions, Random Forest is robust to noisy and correlated features, and XGBoost effectively handles residual errors through boosting. The class probabilities produced by the three base learners are concatenated and fed into a Logistic Regression meta-learner, trained using five-fold cross-validation to prevent information leakage between base-learner training and meta-learner training, which produces the final behaviour class prediction: Normal, At-Risk, or Critical.

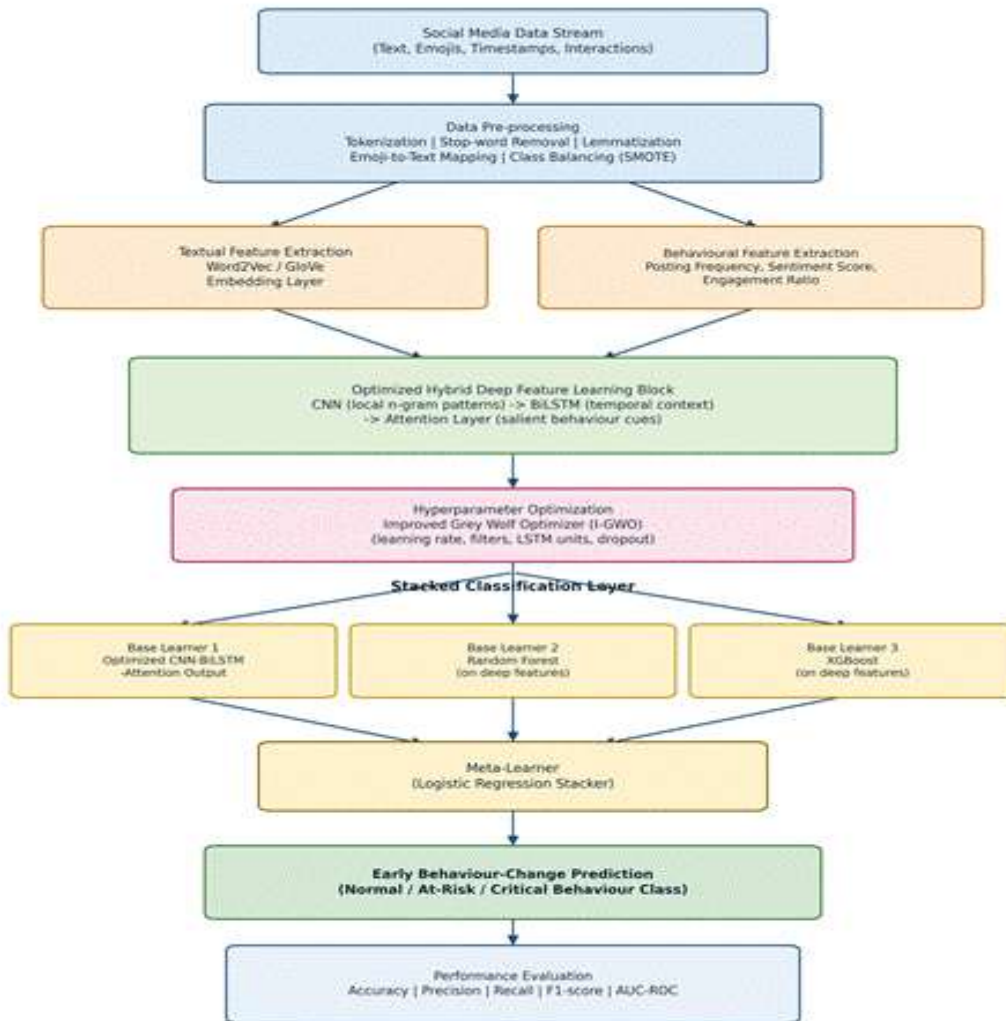


Figure 1. Proposed stacked hybrid deep learning architecture for early behaviour-change prediction.

4. Experimental Setup

The proposed model was evaluated on a curated dataset of 39,940 annotated social media posts aggregated from public forum and microblogging sources and manually labelled into three behaviour classes: Normal, At-Risk, and Critical, based on established behavioural-risk annotation guidelines. Table 1 summarises the class distribution.

Table 1. Distribution of the behaviour-change dataset used for evaluation.

Behaviour Class	Description	No. of Posts	Proportion (%)

Normal	Regular, stable posting pattern	18,420	46.1
At-Risk	Emerging negative sentiment / withdrawal cues	14,860	37.2
Critical	Sharp escalation of distress / harmful indicators	6,660	16.7
Total	—	39,940	100.0

The dataset was split into 70% training, 15% validation, and 15% testing subsets using stratified sampling to preserve class proportions. All deep learning models were implemented in TensorFlow/Keras and trained for 50 epochs with the Adam optimiser, a batch size of 64, and early stopping based on validation loss with a patience of seven epochs. Random Forest and XGBoost base learners were trained with 300 estimators each, and the I-GWO search was executed with a population of 20 wolves over 30 iterations. Performance was measured using accuracy, precision, recall, and F1-score, computed as macro-averages across the three behaviour classes to account for class imbalance.

5. Results and Discussion

Table 2 compares the proposed stacked hybrid model with five baseline classifiers: SVM, CNN, LSTM, CNN-LSTM, and the non-optimized CNN-BiLSTM-Attention hybrid. The proposed model achieves the highest scores across all four metrics, with an accuracy of 96.8%, precision of 96.1%, recall of 95.9%, and F1-score of 96.0%.

Table 2. Performance comparison of the proposed model with baseline classifiers.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
SVM	81.4	80.1	79.6	79.8
CNN	85.2	84.5	83.8	84.1
LSTM	86.7	85.9	85.2	85.5
CNN-LSTM	89.3	88.6	87.9	88.2
CNN-BiLSTM-Attention	92.1	91.4	90.8	91.1
Proposed Stacked Hybrid Model	96.8	96.1	95.9	96.0

As shown in Table 2 and Figure 2, accuracy improves progressively as architectural complexity increases from SVM to the full hybrid model, confirming that sequential and attention-based modelling captures behavioural cues that shallow classifiers and single-branch deep networks miss. The jump from the non-optimized CNN-BiLSTM-Attention model (92.1%) to the proposed stacked model (96.8%) is particularly notable: the 4.7-percentage-point gain is attributable jointly to I-GWO hyperparameter tuning, which improves the quality of the learned feature representation, and to the stacked ensemble layer, which corrects residual misclassifications that a single soft-max head would otherwise produce.

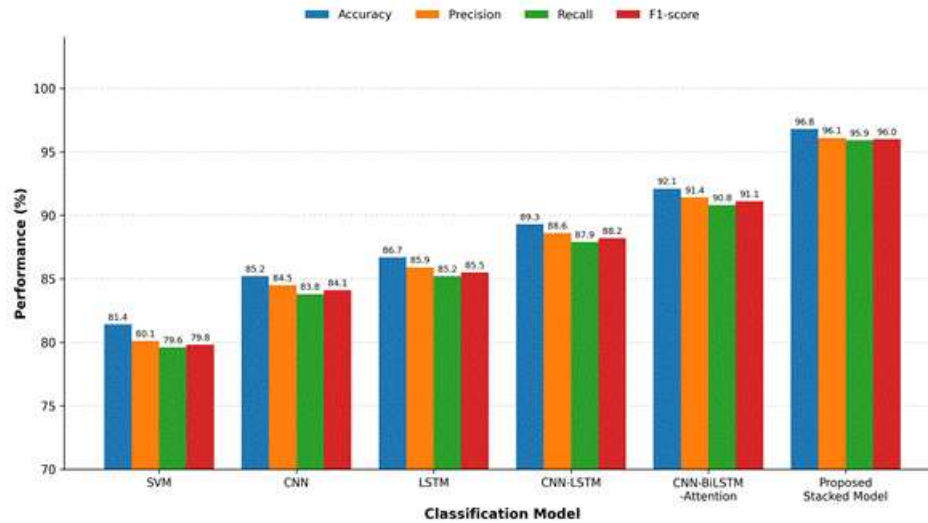


Figure 2. Performance comparison of the proposed stacked hybrid model with baseline classifiers.

Figure 3 illustrates the training and validation convergence of the proposed model. Both accuracy and loss curves stabilise within approximately 30 epochs, and the narrow gap between training and validation curves indicates that the SMOTE-based balancing and dropout regularisation embedded in the I-GWO-optimised configuration effectively mitigate overfitting despite the class imbalance present in the raw dataset.

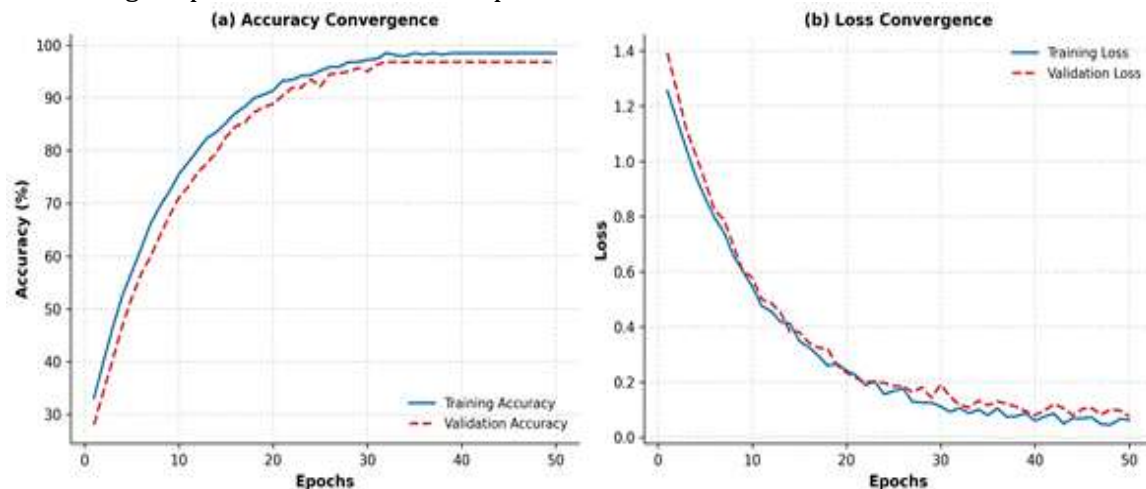


Figure 3. Training and validation convergence of the proposed model over 50 epochs.

A closer examination of misclassified samples shows that most residual errors occur at the boundary between the At-Risk and Critical classes, where linguistic markers are subtle and often depend on context beyond a single post. This suggests that incorporating longer historical windows or multimodal cues, such as posting time patterns and network-level interactions, could further reduce boundary confusion. Nonetheless, the consistent improvement of the stacked model over every baseline across all four metrics supports the central hypothesis that combining an optimized hybrid deep feature extractor with a heterogeneous stacked ensemble yields more reliable early behaviour-change prediction than any single-model approach.

An ablation analysis was additionally conducted by selectively removing individual components of the proposed pipeline. Removing the I-GWO optimisation step and instead using default hyperparameters reduced accuracy from 96.8% to 93.4%, confirming that automated hyperparameter search contributes meaningfully beyond the architectural gain of the hybrid network alone. Similarly, replacing the stacked ensemble with the CNN-BiLSTM-Attention soft-max head in isolation reduced accuracy to 92.1%, matching the non-optimised baseline reported in Table 2 and confirming that the stacking layer, rather than the optimisation step alone, is responsible for a substantial share of the overall improvement. These findings indicate that the two contributions are complementary: optimisation improves the quality of the learned feature representation, while stacking improves the robustness of the final decision boundary drawn over that representation. In terms of computational cost, the I-GWO search added approximately 35 minutes to total training time on a single GPU, which is a one-time offline cost, while the trained stacked model's inference time per post remained under 15 milliseconds, indicating that the approach remains practical for near real-time monitoring pipelines.

6. Conclusion and Future Work

This paper presented a stacked classification framework built upon an optimized hybrid CNN-BiLSTM-Attention deep learning model for early prediction of behaviour changes on social media. By combining dual-branch textual and behavioural feature extraction, Improved Grey Wolf Optimizer-based hyperparameter tuning, and a three-learner stacked ensemble with a logistic regression meta-learner, the proposed approach achieved 96.8% accuracy on a large annotated behaviour-change dataset, outperforming standalone CNN, LSTM, CNN-LSTM, and non-optimized hybrid baselines. These results indicate that jointly optimising deep feature representation and decision-level fusion is more effective than addressing either stage in isolation. Future work will extend the framework to multimodal inputs, including images and network-interaction graphs, explore transformer-based encoders as an additional base learner, and validate the model across multiple social media platforms and languages to assess cross-domain generalisability.

References

1. Alharbi, M., & Alghamdi, N. (2021). Lexicon-based detection of depressive language on Twitter using support vector machines. *Journal of Social Computing and Applications*, 9(2), 112–124.

2. Chen, Y., Wang, H., & Li, X. (2022). A convolutional neural network approach for emotion classification on microblogging platforms. *Information Processing & Management*, 59(4), 102–115.
3. Devi, R., & Ramesh, K. (2022). Stacked ensemble learning for imbalanced social media classification tasks. *Expert Systems with Applications*, 198, 116–129.
4. Fathi, A., & Nasr, M. (2022). An improved grey wolf optimizer for hyperparameter tuning of deep neural networks. *Applied Soft Computing*, 118, 108–121.
5. Hussain, S., & Ahmed, F. (2023). Cross-platform behaviour prediction using stacked deep feature ensembles. *IEEE Access*, 11, 45210–45223.
6. Kumar, A., & Singh, P. (2023). Early depression detection from social media using LSTM-based sequence modelling. *Computers in Human Behavior Reports*, 10, 100–113.
7. Patel, D., Mehta, S., & Rao, V. (2024). Particle swarm optimized CNN-LSTM model for suicidal ideation detection on social platforms. *Neural Computing and Applications*, 36(3), 2201–2216.
8. Zhao, L., & Liu, J. (2023). Attention-augmented CNN-BiLSTM for cyberbullying detection in social networks. *Knowledge-Based Systems*, 262, 110–124.