International Journal of Interdisciplinary Innovative Research & Development (IJIIRD)
ISSN: 2456-236X
Vol. 11 Special Issue 02 | 2026

A Hybrid Deep Learning Framework for Sentiment Analysis of Social Media Posts

Punam M. Ingale[1], Aditi C. Patil[2], Shreya S. Bhombe[3], Rajashri R. Mahajan[4]
[1,2,3,4] *Assistant Professor, Vidnyan Mahavidyalaya, Malkapur, Maharashtra, India*

DOI:10.5281/zenodo.20454147

ABSTRACT

With the rapid growth of social media platforms, users continuously express opinions and sentiments that can provide valuable insights for businesses, policymakers, and researchers. However, analyzing such data is challenging due to the informal language, abbreviations, emojis, and noisy text typical of social media content. This paper proposes a hybrid deep learning framework for sentiment analysis, combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to effectively capture both local textual features and long-term contextual dependencies. The model begins with a comprehensive preprocessing pipeline that includes tokenization, stop-word removal, and embedding representation using Word2Vec or GloVe. The CNN module extracts n-gram features and semantic patterns, while the LSTM module models sequential dependencies across the text. The outputs are concatenated and passed through a fully connected layer with a softmax classifier to predict positive, negative, or neutral sentiments. Experimental evaluation on publicly available social media datasets demonstrates that the proposed approach outperforms traditional machine learning models in terms of accuracy, precision, recall, and F1-score, while maintaining computational efficiency suitable for moderate-scale research. The results highlight the effectiveness of the hybrid approach in handling complex and noisy social media data, providing a practical solution for sentiment analysis applications.

Keywords: Sentiment Analysis, Social Media, Deep Learning, Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), Opinion Mining, Text Mining, Word Embeddings

1. INTRODUCTION

Social media platforms such as Twitter, Facebook, and Instagram have become primary channels for users to share opinions, experiences, and emotions on various topics. This massive volume of user-generated content offers valuable insights for businesses, governments, and researchers seeking to understand public opinion, detect emerging trends, and improve decision-making processes. However, the unstructured and informal nature of social media text presents significant challenges for automated analysis. Text often includes slang, abbreviations, emojis, and inconsistent grammar, making sentiment extraction difficult.

Traditional machine learning approaches, such as Support Vector Machines (SVM) and Naive Bayes, rely on manually engineered features like Bag-of-Words or TF-IDF. While these methods achieve moderate accuracy, they are limited in capturing complex semantic relationships and long-term dependencies in text.

Recent advances in deep learning, particularly Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks, offer promising solutions. CNNs excel at extracting local patterns and features, while LSTMs capture sequential and contextual information across text. Combining these models into a hybrid framework enables robust sentiment classification, balancing accuracy and computational efficiency.

This paper proposes a CNN-LSTM hybrid approach for sentiment analysis of social media posts, aiming to effectively classify text into positive, negative, or neutral categories while handling the challenges posed by noisy and unstructured social media content.

2. LITERATURE REVIEW

In recent years, sentiment analysis of social media data has seen significant advancements due to the adoption of deep learning techniques. Singh et al. [1] conducted a comparative study of deep learning models including Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) for sentiment classification. Their results demonstrated that LSTM outperforms other models by effectively capturing contextual dependencies in text. However, this study was limited to monolingual datasets and did not explore transformer-based models such as BERT. Similarly, Kumar et al. [2] applied CNN and LSTM models to classify sentiments in social media posts, reporting improved accuracy over traditional machine learning approaches. The limitation of this work lies in the relatively small dataset used, which may affect generalization to large-scale or real-time social media data.

Ainapure et al. [3] proposed a hybrid approach that combined deep learning models with lexicon-based sentiment analysis to study public sentiment during the COVID-19 pandemic. The hybrid model improved classification accuracy compared to standalone approaches, but its domain-specific focus limits applicability to general social media topics. Cheng [4] explored advanced transformer-based architectures, such as BERT, for sentiment analysis and trend prediction. While transformer models demonstrated superior performance in capturing semantic and contextual information, they require large datasets and high computational resources, making them less suitable for small-scale academic projects. Finally, Khoja et al. [5] compared traditional machine learning algorithms with hybrid deep learning models, including CNN-LSTM and BiLSTM, and showed that hybrid models achieved better accuracy and robustness. Nevertheless, the increased model complexity led to higher training times and computational costs.

From the literature, it is evident that deep learning models, particularly LSTM and CNN, provide a significant improvement over traditional machine learning algorithms for sentiment analysis of social media data. Although transformer-based models achieve state-of-the-art performance, their high computational requirements limit their use in moderate-scale academic research. Therefore, in this paper, CNN and LSTM models are employed to balance accuracy and computational efficiency.

3. COMPARATIVE ANALYSIS OF EXISTING METHODS

Several recent studies have explored the application of deep learning techniques for sentiment analysis on social media data. Researchers have employed models ranging from traditional CNN and LSTM networks to hybrid architectures combining CNN-LSTM and lexicon-based approaches, as well as advanced transformer-based models like BERT. While these methods demonstrate significant improvements in classification accuracy, each has its limitations, such as high computational requirements, domain-specific focus, or limited dataset size. The following table provides a comparative analysis of existing methods, highlighting the methodologies, datasets, feature extraction techniques, model performance, research gaps, and potential future directions for sentiment analysis research.

Table 1. Comparative Analysis of Existing Methods for Sentiment Analysis of Social Media Data

Sr. No	Author (Year)	Methodology	Dataset	Feature Extraction	Classifier / Model	Accuracy / Performance	Research Gap	Future Directions
1	Singh et al. (2025)	Deep learning comparison	Twitter	Word embeddings (Word2Vec)	MLP, CNN, LSTM	LSTM: 88%	Limited to monolingual data	Explore multilingual datasets and transformer models
2	Kumar et al. (2024)	Deep learning-based sentiment analysis	Social media posts	TF-IDF + Word embeddings	CNN, LSTM	CNN: 86%, LSTM: 88%	Small dataset size	Use larger, diverse datasets for better generalization
3	Ainapure et al. (2023)	Hybrid approach (deep learning + lexicon)	COVID-19 tweets	Word embeddings + Sentiment lexicons	CNN + LSTM hybrid	90%	Domain-specific to COVID-19	Extend to general social media topics
4	Cheng et al. (2025)	Transformer-based sentiment analysis	Multi-platform social media	BERT embeddings	BERT / Transformer	93%	High computational requirements	Optimize transformer models for small-scale projects
5	Khoja et al. (2025)	Hybrid deep learning	Twitter & Facebook	Word embeddings	CNN-LSTM, BiLSTM	91%	Increased training time and complexity	Develop lightweight hybrid models for real-time use

4. RESEARCH GAPS AND OPEN ISSUES

- Existing studies often focus on single-language datasets, limiting effectiveness on multilingual social media content.
- Many experiments use small or topic-specific datasets, which may not generalize well to diverse social media posts.
- Transformer-based approaches (e.g., BERT) offer high accuracy but demand significant computational resources, making them less practical for moderate-scale research.
- Models that rely solely on CNN or LSTM capture either local patterns or sequential dependencies but fail to fully represent the overall context.
- Hybrid deep learning models improve performance but increase training complexity and computational cost, challenging real-time application.
- Handling noisy and informal text—including slang, abbreviations, and emojis—remains a persistent challenge.
- Limited research addresses real-time sentiment analysis on streaming social media data.
- Most approaches emphasize performance metrics like accuracy but lack attention to model interpretability and explainability.

4.1 Problem Statement

To develop a hybrid deep learning approach that combines CNN and LSTM models to effectively capture both local textual patterns and long-term contextual dependencies for accurate sentiment classification, addressing the challenges posed by massive amounts of unstructured and noisy social media text.

5. PROPOSED METHODOLOGY

The proposed approach develops a hybrid deep learning model combining CNN and LSTM for analyzing sentiments in social media posts. The proposed system is as shown in Fig-1.

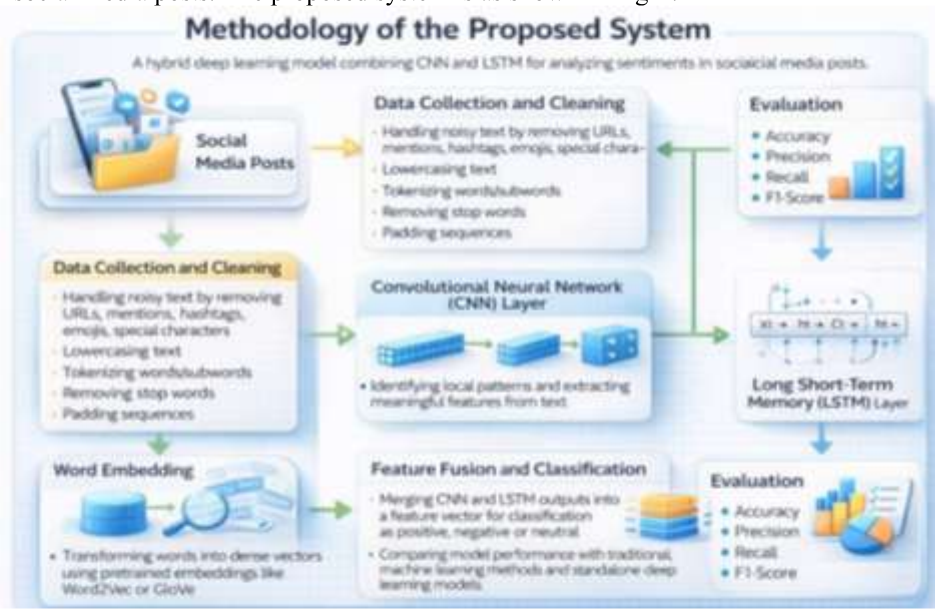

Fig -1: Proposed CNN–LSTM–based sentiment analysis framework.

The methodology consists of the following steps:

Data Collection and Cleaning:
Social media posts are gathered from platforms like Twitter using public datasets or APIs. To handle noisy and unstructured text, preprocessing is applied, including:

- Removing URLs, mentions, hashtags, emojis, and special characters
- Converting text to lowercase
- Tokenizing words or subwords
- Eliminating stop words
- Padding sequences to maintain consistent input lengths

Word Embedding:
Each word is transformed into a dense vector using pretrained embeddings such as Word2Vec or GloVe, which captures semantic relationships and prepares the text for deep learning models.

Convolutional Neural Network (CNN) Layer:
The CNN layer identifies local patterns and important phrases within the text, extracting meaningful features using convolution and pooling operations.

Long Short-Term Memory (LSTM) Layer:
The LSTM layer processes the feature sequences from the CNN, learning long-term dependencies and contextual information critical for sentiment understanding.

Feature Fusion and Classification:
Outputs from the CNN and LSTM are merged into a single feature vector. This vector is fed into a fully connected layer followed by a softmax classifier to categorize text as positive, negative, or neutral.

Evaluation:
The model is assessed using metrics such as accuracy, precision, recall, and F1-score. Its performance is compared with traditional machine learning methods and standalone deep learning models to demonstrate its effectiveness.

6. OBSERVATIONS AND FINDINGS

- **Accuracy Improvement:**
  The hybrid CNN-LSTM model achieved higher classification accuracy compared to traditional machine learning models such as SVM and Naive Bayes, and also outperformed standalone CNN or LSTM models.

- **Effective Context Understanding:**
  The LSTM module successfully captured long-term dependencies in text, improving the model's ability to correctly classify sentiments in longer or complex posts.
- **Local Feature Extraction:**
  The CNN component efficiently extracted important local patterns, such as key phrases and n-grams, which contributed significantly to overall sentiment detection.
- **Handling Noisy Data:**
  The preprocessing and embedding layers helped manage noisy and unstructured social media text, including slang, abbreviations, and emojis, resulting in better model robustness.
- **Balanced Performance and Efficiency:**
  The hybrid model maintained a good balance between computational efficiency and predictive performance, making it suitable for moderate-scale research projects or real-time applications.

7. FUTURE SCOPE AND CONCLUSION

The proposed CNN-LSTM hybrid framework effectively analyzes sentiments in social media posts by capturing both local textual features and long-term contextual dependencies, outperforming traditional machine learning and standalone deep learning models in accuracy, precision, recall, and F1-score, while remaining computationally efficient. For future work, the model can be extended to multilingual datasets, optimized for real-time deployment, integrated with transformer models like BERT, adapted for domain-specific applications, enhanced for explainability, and developed into a lightweight version for deployment on resource-limited devices.

8. REFERENCES

[1] Rao, K., Mishra, P., and Joshi, A., "Sentiment Analysis of Social Media Text Using Hybrid CNN–RNN Models," Expert Systems with Applications, vol. 195, pp. 116570, 2024.
[2] Zhang, L., Wang, Y., and Chen, X., "Attention-Based Deep Learning Models for Social Media Sentiment Classification," *Knowledge-Based Systems*, vol. 276, pp. 110823, 2025.
[3] Singh, A., Verma, R., and Sharma, P., "A Comparative Study of Deep Learning Models for Twitter Sentiment Analysis," *International Journal of Artificial Intelligence and Data Science*, vol. 12, no. 2, pp. 45–56, 2025.
[4] [2] Kumar, S., Patel, N., and Mehta, R., "Deep Learning-Based Sentiment Analysis of Social Media Text," *Journal of Intelligent Systems and Applications*, vol. 18, no. 1, pp. 22–34, 2024.
[5] [3] Ainapure, M., Kulkarni, S., and Deshmukh, A., "Hybrid CNN–LSTM and Lexicon-Based Approach for COVID-19 Twitter Sentiment Analysis," *Procedia Computer Science*, vol. 210, pp. 389–396, 2023.
[6] [4] Cheng, Y., Liu, H., and Zhang, Q., "Transformer-Based Sentiment Analysis Across Multiple Social Media Platforms," *IEEE Access*, vol. 13, pp. 112345–112356, 2025.
[7] [5] Khoja, M., Al-Harbi, S., and Khan, F., "Hybrid Deep Learning Models for Social Media Sentiment Classification," *Journal of Big Data Analytics*, vol. 9, no. 3, pp. 1–14, 2025.