International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 31, No. 5, May, 2026, pp. 1667-1676
Journal homepage: http://iieta.org/journals/isi

HCSE-Net: A Hybrid BiLSTM Framework with Structural Feature Fusion for Tweet Sentiment Classification

Roma R. Jain[1,2*], Supriya O. Rajankar[2]

[1] Department of Electronics and Telecommunication, Bharati Vidyapeeth's College of Engineering for Women, Pune 411043, India
[2] Department of Electronics and Telecommunication, Sinhgad College of Engineering, Pune 411041, India

Corresponding Author Email: romajain30@gmail.com

https://doi.org/10.18280/isi.310522

Received: 26 February 2026
Revised: 20 April 2026
Accepted: 5 May 2026
Available online: 31 May 2026

Keywords:
twitter sentiment analysis, Bidirectional Long Short-Term Memory, structural feature fusion, deep learning, social media analytics, text classification

ABSTRACT

Sentiment analysis of social media content has become increasingly important for opinion mining, customer feedback analysis, public sentiment monitoring, and decision-support systems. However, accurately classifying sentiment in Twitter data remains challenging because tweets are often short, noisy, and characterized by informal language, emojis, hashtags, abbreviations, and context-dependent expressions. While existing deep learning (DL) approaches have demonstrated strong representation learning capabilities, many of them primarily focus on semantic information and insufficiently exploit tweet-specific structural cues. This study proposes Hybrid Contextual and Structural Encoding Network (HCSE-Net), a hybrid sentiment classification framework that combines bidirectional contextual sequence learning with auxiliary structural feature encoding. The proposed architecture integrates word embeddings, a Bidirectional Long Short-Term Memory (BiLSTM) network, and a structural feature fusion mechanism that incorporates tweet-level attributes, including emoji frequency, hashtag usage, punctuation intensity, uppercase word ratio, and tweet length. By jointly modeling semantic dependencies and structural sentiment indicators, the framework aims to improve sentiment discrimination in noisy social media environments. The proposed model was evaluated on three publicly available Twitter sentiment datasets containing positive, negative, and neutral sentiment classes. Experimental results demonstrate that HCSE-Net achieved an accuracy of 94.63%, precision of 94.12%, recall of 93.88%, and F1-score of 94.00%. Comparative experiments further show that the proposed framework consistently outperformed Multi-Feature (MF)-Convolutional Neural Networks (CNN)-BiLSTM, CNN- Long Short-Term Memory (LSTM), CNN-BiLSTM, and a tuned Bidirectional Encoder Representations from Transformers (BERT) baseline under identical evaluation settings. These findings indicate that integrating structural tweet features with contextual sequence modeling provides an effective and computationally practical approach for Twitter sentiment classification.

1. INTRODUCTION

The rapid evolution of social media platforms has transformed how individuals communicate and express opinions online. Twitter has gained significant popularity as a source of real-time popular mood because of the nature of tweets, where most of the content is brief, and the user provides a lot of participation. Consequently, the Natural Language Processing (NLP) task aimed at automatically identifying the emotional polarity of tweets. Sentiment analysis makes it possible to reveal significant knowledge about a huge amount of textual data that otherwise would be hard to assess manually [1]. With the increased use of social media in communication and information exchange, automated sentiment analysis has become a valuable component of the current data-driven decision-making system in businesses and research in the past few years, as people have learned how effective and valuable it is to analyse the sentiments on Twitter in order to learn more about the opinion of the audience and maintain their reputation and predict what new social trends will occur. Most tweets are brief, informal, and full of expressive features such as emojis, hashtags, abbreviations, and slang. This is unlike news reports and official paperwork. These attributes give both possibilities and difficulties to the sentiment classification processes. Tweets offer instant and real-time feedback by the users; however, since they are not structured, it is difficult to comprehend the meaning of their tweets [2]. As a result, it has been proposed that the development of useful computer algorithms to detect the sentiment of a tweet has become a major research agenda. Accurate sentiment detection provides valuable insights for business analytics, political forecasting, healthcare monitoring, and crisis management. Firms tend to use customer tweets to check the performance of their products and how to market it. Policymakers analyze sentiment trends to determine how the citizens of the country are feeling about the government projects or social activities. Moreover, in case of crisis or other disasters, the mood analysis of the people can assist the government to locate false data, emotional distress, or genuine concerns of people [3]. These uses indicate that emotion recognition is not just a technique issue, but a useful tool to make informed decisions in a variety of real-life situations.

The reason is that Tweet sentiment analysis is significant; however, it contains numerous issues that distinguish it from other text categorisation tasks. The tweets are usually full of loud language, bad syntax and statements that are dependent on the situation, which may deceive the average algorithms. Implicit attitude, sarcasm, and irony make it more difficult to interpret what an individual means and therefore models need to locate unobtrusive links between words in the context. Additionally, on the internet, language is a constantly evolving phenomenon, so new terms, memes, and acronyms are generated continuously. This renders fixed lexicons or rule systems useless as time goes by Rosenthal et al. [4]. Overcoming these issues requires sophisticated modelling techniques that are able to comprehend local word-level patterns as well as global context linkages.

The vast majority of initial sentiment analysis systems were constructed based on lexicons and using conventional machine learning (ML) techniques. Lexicon-based methods give ratings of the word or sentence with respect to its polarity based on existing emotion dictionaries, giving an interpretable result, but with low flexibility to contextual variations. Some examples of machine learning models include Naive Bayes, Support Vector Machine (SVM) and Decision Trees, which enhance the accuracy of classification by identifying statistical patterns within the data that has been labelled. However, the latter techniques heavily rely on the structural aspects and do not consider sequential relations inherent in language [5]. The need to generate representations of features from raw text using deep learning (DL) algorithms emerged when researchers started to study how to generate representations of features and enhance understanding of the resulting representations as the amount and complexity of social media data grew. Development of deep neural networks, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has contributed greatly to the research in sentiment analysis. CNNs are effective at discovering local semantics, such as phrases or n-grams, but Long Short-Term Memory (LSTM) networks are more effective at modelling the time-connectivity of word sequences. Bidirectional Long Short-Term Memory (BiLSTM) architectures also perform sentiment prediction with more accuracy because they use both forward and backward analysis of text to build a stronger contextual understanding. Recent literature has shown that hybrid architectures with CNN and BiLSTM layers outperform more traditional machine learning algorithms particularly when dealing with short and noisy social media input [6, 7].

Nevertheless, when tweets contain structural information such as hashtags or emoticons which can convey some critical emotional signals, deep learning methods fail. The other emerging trend of sentiment study is the inclusion of additional features and structural data into the neural networks. Basic word embeddings may offer a more detailed view of the underlying emotion, with models typically incorporating meta-linguistic information like punctuations, number of mentions, and usage of emojis. The goal of these hybrid frameworks is to bridge language comprehension and situational awareness, and therefore they are more suitable to real-world Twitter data. With the development of the field, there is the growing interest to produce the systems that are very balanced in terms of performance, interpretability, and computational efficiency [8, 9].

Overall, the tweet sentiment recognition is becoming a subject of study involving machine learning, deep learning, and language analysis. Previous approaches provided the structure upon which automated opinion mining can be established, but the latest deep architectures are reassuring in their ability to deal with the complexity of the language employed on social media. The enduring nature of noise, ambiguity and the constantly changing conditions of the online communication styles highlight the fact that new hybrid models are required that can be effective in obtaining both semantic and structural information. Coupled with these issues, the present-day research work has focused on the creation of innovative architectures that are able to create more contextual knowledge whilst being robust in the context of real-world sentiment analysis in Twitter. Despite significant advancements in deep learning-based sentiment analysis, existing hybrid models primarily focus on semantic embeddings while often ignoring tweet-specific structural sentiment indicators such as emojis, hashtags, punctuation intensity, and informal writing patterns. Although transformer-based architectures provide strong contextual modelling, they introduce substantial computational complexity and deployment overhead for real-time sentiment analytics systems. Therefore, there remains a need for a computationally efficient framework capable of jointly learning contextual and structural sentiment information from noisy Twitter data.

In addition, modern information systems increasingly rely on automated sentiment-aware analytics for decision-support applications, customer feedback mining, crisis management, and public opinion monitoring. The proposed HCSE-Net framework is designed to support such intelligent social media analytics systems through efficient and context-aware sentiment prediction. In this work, the performance and generalizability of a proposed Twitter sentiment analysis system were tested using three publicly available benchmark datasets. The initial one is the Twitter Sentiments Dataset of Hussein [10]. It contains tweets belonging to a number of categories, which have been tagged and their sentiment polarity. The dataset of CS98X Twitter Sentiment Classification was uploaded on Kaggle by Azzopardi [11]. It contains a good combination of tweets that can be employed in supervised sentiment learning. The remote supervision dataset that is famous and was created by Go et al. [12] was also used to offer more diversity and size to the training samples.

To ensure consistency across sources, all datasets underwent preprocessing before testing, including normalisation, tokenisation, and noise reduction techniques. This research categorised sentiment labels into three categories: positive, negative, and neutral. This made it capable of applying the same assessment paradigm to all the datasets. Adding different datasets strengthens the model by subjecting it to differences in style of language, topic distribution and annotation methods. This is used to remove bias in the datasets.

The rest of this paper is intended to outline the research that clearly and logically explains the research that was recommended. Section 1 discusses the key principles of tweet sentiment recognition, including but not restricted to why it matters, what the pitfalls of it are, and what should be done with the work. Section 2 is an intensive scrutiny of the literature that discusses the constraints of the available machine learning methods and deep learning methods, as far as handling informal writing regarding social media is concerned. Section 3 discusses the planned approach, such as architectural design preprocessing techniques, embedding, and the hybrid architecture of deep learning used in the categorisation of sentiment. In Section 4, the experimental setup, the dataset, the evaluation metrics, and an intensive examination of the findings to determine the level at which the HCSE-Net has been achieved are presented. Lastly, Section 5 concludes with the most appropriate findings, identifies the scientific contributions, and suggests potential paths of further study of Twitter sentiment analysis.

2. LITERATURE REVIEW

The literature review will explore modern advances in tweet sentiment analysis, with the emphasis on traditional machine learning, deep learning, and hybrid neural network models. The effectiveness of CNN, LSTM, BiLSTM, and transformer-based systems in enhancing sentiment classification up to now has been demonstrated in previous studies. Still, there are huge issues, particularly with the possibility of interpreting informal language, the Twitter structure, and ambiguity. This evaluation will form part of the research gaps that will be identified to enable a more advanced sentiment analysis system to come into existence. Badholia et al. [13] studied sentiment analysis based on machine learning in a large variety of fields and compared traditional algorithms, including Naive Bayes and SVM. Their studies revealed that despite the speed with which ML methods can be solved, they do not always explain the relationships among informal tweets.

Mao et al. [14] also did a thorough evaluation of sentiment analysis methodologies, and this evaluation revealed the shift between the conventional machine learning and the advanced deep learning frameworks. Their analysis revealed that, despite the fact that the present-day NLP structures can strengthen the contextual awareness, they still encounter difficulties with noisy social media content and sarcasm detection, which means that hybrid frameworks incorporating various feature representations are necessary. The comparison of CNN, LSTM and BiLSTM architectures in categorisation of Twitter moods was carried out by Xu et al. [15]. They discovered that the BiLSTM models are more effective than the traditional methods since they can capture bidirectional context. They also demonstrated that quality preprocessing is significant to overall performance.

Mirdan et al. [16] analysed advanced deep learning algorithms to deal with sentiment analysis within the context of business analytics, with a focus on the ability of AI-based sentiment prediction to develop sales forecasting and customer engagement strategies.

Geethanjali and Valarmathi [17] suggested a deep learning hybrid model, which is a combination of CNN and LSTM, in the study of multimodal emotion during crisis. Their results also revealed the need to include numerous sources of features to enhance the capabilities of classification in vibrant social media circumstances.

El Koufi et al. [18] suggested a hybrid CNN-LSTM model to be used in the analysis of customer sentiment on online marketplaces. Their results supported the effectiveness of deep hybrid models in learning semantic patterns and decision-making in the marketing environment. Another recently developed LSTM model with multi-head attention is proposed by Yi et al. [19], and it demonstrated much higher accuracy in sentiment classification compared to the traditional LSTM networks. Their findings emphasized the role of attentional mechanisms in the summing up of complex emotional reactions in text.

The proposed CNN-LSTM model by Alasmari et al. [20] was used to evaluate sentiments in social media within large datasets. Their approach demonstrated the usefulness of the combination of convolutional feature extraction with sequential learning to locate short-term and long-term patterns in text.

Li et al. [21] constructed an attention-based CNN-BiLSTM model utilising the time and space correlation of tweets. They revealed that the accuracy of sentence classification of sentiment increased, especially when attention methods were employed to focus on words that carry sentiment.

The authors proposed that an MF-CNN-BiLSTM with topic modelling can be used to analyse conversations on Twitter about world events [22]. The authors show that convolutional layers combined with the BiLSTM networks outperform single deep learning methodologies on the feature extraction and contextual modelling.

Das et al. [23] tested sentiment categorisation with a modified version of BERT-based model in comparison with traditional ML and LSTM models. The authors discovered that models based on transformers enhance the accuracy of classification; however, they require substantial computational capacity and sensitive hyperparameter optimization. Si and Wei [24] suggested a LSTM and BERT method to overcome the bad use of features in text sentiment analysis. The authors stated that better representation learning is necessary in order to achieve better classification on a very large range of data.

Zhang et al. [25] a hybrid BERT-BiLSTM-Attention model was proposed by them. Their paper revealed that contextual embeddings applied in sequential modelling were more interpretable and performed better in multi-class sentiment tasks.

Tiwari et al. [26] were able to review ensemble techniques on sentiment classification and found that bagging and boosting techniques are found to have a high accuracy of prediction compared to individual ML classifiers. However, due to their expensive nature to compute the ensemble models, they are difficult to apply in real-time. The models evaluated by Pookduang et al. [27] include conventional machine learning models, LSTM networks, and transformer-based models, including RoBERTa. Their findings show that transformer models are more precise, but LSTM can remain competitive as it is capable of modelling sequences.

Even while earlier ML and DL models have come a long way, many methods don't do a good job of combining semantic context with extra structural hints in tweets. The limitations seen in the MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, and optimised transformer models underscore the need for a more contextually aware and structurally enhanced architecture. Overall, existing sentiment analysis approaches can be broadly categorized into traditional machine learning methods, deep learning architectures, hybrid CNN-LSTM/BiLSTM systems, and transformer-based frameworks. Traditional machine learning methods suffer from limited contextual understanding, while CNN-based models struggle to capture long-range semantic dependencies. Transformer architectures improve contextual representation but often require substantial computational resources and fine-tuning complexity. Moreover, many existing studies insufficiently integrate tweet-specific structural sentiment cues such as emojis, hashtags, punctuation patterns, and informal linguistic expressions. These limitations motivate the development of the proposed HCSE-Net framework. To solve these problems, suggest the HCSE-Net (Aux-BiLSTM) framework, which uses embedding-based representation, bidirectional sequence learning, and auxiliary structural encoding to get better accuracy, faster convergence, and stronger sentiment classification performance.

3. METHODOLOGY

The proposed tweet sentiment classification framework is a hybrid deep learning architecture that employs embedding representation, bidirectional sequence modelling, and extra structural characteristics to improve sentiment classification performance, as shown in Figure 1.

The proposed HCSE-Net framework incorporates several auxiliary structural features to enhance sentiment understanding in noisy Twitter text. These features include emoji count, hashtag frequency, punctuation patterns (such as exclamation and question marks), uppercase word ratio, tweet length, repeated character occurrences, and the number of user mentions and URLs. Each structural feature was numerically encoded and normalized using min-max scaling before integration into the network. The extracted auxiliary feature vector was then concatenated with the contextual feature representation generated by the BiLSTM layer through a feature fusion mechanism. This hybrid representation enabled the model to capture both semantic dependencies and tweet-specific structural cues, thereby improving sentiment classification accuracy and robustness on informal social media data, as shown in Figure 2.

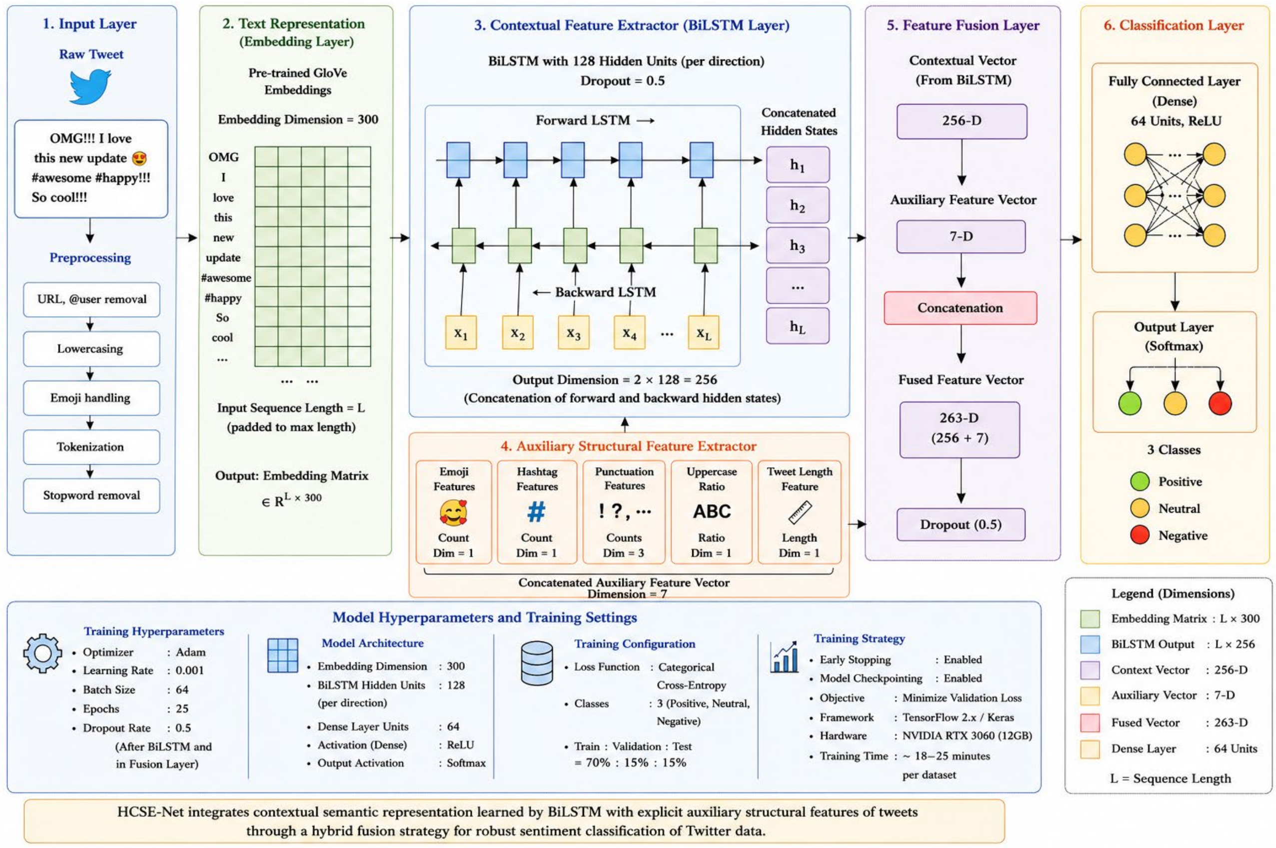


Figure 1. Proposed architecture

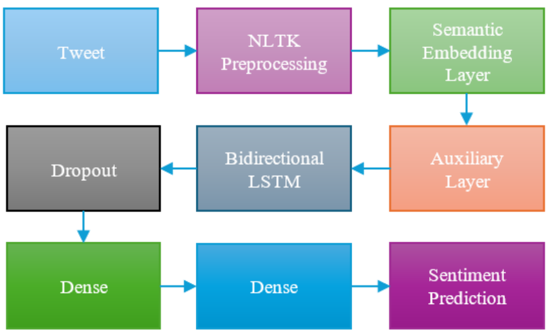


Figure 2. Sequence wise illustration of proposed methodology

Unlike conventional CNN-BiLSTM and attention-BiLSTM architectures that primarily focus on semantic sequence learning, the proposed HCSE-Net framework jointly models contextual semantic dependencies and tweet-specific structural sentiment indicators within a unified feature fusion architecture. This enables improved handling of noisy and highly informal Twitter content while maintaining lower computational complexity than transformer-based architectures.

Let X denote the input tweet sequence and W_e represent the embedding weight matrix. The embedded feature representation E is obtained in Eq. (1):

$$E = X \cdot W_e \qquad (1)$$

where, E is the embedding output used as input to the BiLSTM network.

The BiLSTM processes the sequence in both forward and backward directions to find dependencies in the context. The LSTM unit's internal gating operations control how information flows through the network.

Forget gate as shown in Eq. (2):

$$f_t = \sigma(Wf \cdot [h_{t-1}, E_t] + bf) \quad (2)$$

Input gate as shown in Eq. (3):

$$i_t = \sigma(Wi \cdot [h_{t-1}, E_t] + bi) \quad (3)$$

Candidate memory state as shown in Eq. (4):

$$\hat{C}_t = tanh(Wc \cdot [h_{t-1}, E_t] + bc) \quad (4)$$

Updated cell state as shown in Eq. (5):

$$C_t = f_t \odot C_{t-1} + {}_t \odot \hat{C}_t \quad (5)$$

Hidden output state as shown in Eq. (6):

$$h_t = o_t \odot tanh(C_t) \quad (6)$$

where, σ represents the sigmoid activation function, $\odot$ denotes element-wise multiplication, and h_t corresponds to the hidden representation at time step t. A feature vector A includes not only sequential features but also extra structural attributes taken from tweets, like the number of emojis, the presence of hashtags, and the patterns of punctuation. We combine these extra features with the BiLSTM output H using a fusion operation as shown in Eq. (7):

$$F = [H; A] \quad (7)$$

where, F is the fused feature representation that comes from putting together sequential and structural information.

Then, the fused features go through dense layers to learn more abstract ideas. A softmax classifier is used to figure out the final sentiment prediction as shown in Eq. (8).

$$y = softmax(Wd \cdot F + bd) \quad (8)$$

where, y is the probability distribution of the sentiment classes (positive, negative and neutral). They are done using categorical cross-entropy loss and gradient-based optimization techniques to check the optimum model parameter values. The proposed structure will improve its ability to handle noisy and informal language that characterizes Twitter data through the incorporation of a learning sequence in a contextual manner with structural auxiliary encoding. The probability distribution between classes of sentiments (positive, negative and neutral). The proposed outline strengthens the tolerance of the informal and unintelligent language that is characteristic of Twitter data by combining contextual sequence learning with structural auxiliary encoding.

Let $S = [s_1, s_2, s_3, \dots, s_n]$ represent the auxiliary structural feature vector extracted from each tweet, where s_1 denotes emoji count, s_2 hashtag frequency, s_3 punctuation intensity, s_4 uppercase ratio, s_5 tweet length, and other tweet-specific attributes. The extracted features were normalized using min-max normalization as follows in Eq. (9).

$$s'_i = \frac{s_i - s_{min}}{s_{max} - s_{min}} \quad (9)$$

The normalized auxiliary feature vector was then fused with the contextual representation h_t generated by the BiLSTM network using feature concatenation in Eq. (10).

$$F = [h_t \oplus S'] \quad (10)$$

where, F represents the final hybrid feature vector and $\oplus$ denotes concatenation. This combined representation enabled the proposed HCSE-Net model to simultaneously capture semantic dependencies and structural tweet patterns, improving robustness and sentiment classification accuracy on noisy Twitter data.

The primary novelty of HCSE-Net lies in its lightweight hybrid fusion strategy that combines contextual sequence learning with explicitly engineered tweet-structure sentiment indicators under a unified architecture optimized for noisy social media environments. Unlike transformer-based systems requiring large-scale fine-tuning, HCSE-Net achieves competitive performance with significantly lower computational overhead.

4. EXPERIMENT AND RESULTS

The experimental part measures the suitability of the suggested HCSE-Net (Aux-BiLSTM) model through the thorough performance analysis of various benchmark Twitter sentiment datasets.

4.1 Experimental setup

The experimental setup for the proposed HCSE-Net framework was implemented using the TensorFlow-Keras deep learning library in Python. Every experiment was carried out on a system with an Intel Core i7 processor, 16 GB of RAM, and no GPU acceleration. The average training time for the proposed model HCSE-Net l was approximately 2.5 hours, depending on dataset size and preprocessing overhead. The implementation environment included Python 3.10, TensorFlow 2.x, NumPy, Pandas, Scikit-learn, and NLTK libraries for preprocessing, feature extraction, and performance evaluation.

The proposed HCSE-Net model was implemented empirically tuned hyperparameters to ensure stable convergence and reproducibility. A 300-dimensional embedding layer and a BiLSTM network with 128 hidden units were used for contextual feature extraction, followed by dense layers with ReLU activation. The Adam optimizer was used to train the model using a batch size of 64, a learning rate of 0.001, and 25 training epochs. After the BiLSTM layer, a dropout rate of 0.5 was used to lessen overfitting. Categorical cross-entropy and softmax activation were used for multi-class sentiment classification. A 70:15:15 ratio was used to separate the dataset into training, validation, and testing sets, and early stopping with model checkpointing was employed during training to improve convergence stability and minimize overfitting.

4.2 Dataset description

The proposed HCSE-Net model was tested using three publicly available benchmark Twitter sentiment datasets: the Twitter Sentiments Dataset by Hussein [10], the CS98X Twitter Sentiment Classification dataset by Azzopardi [11], and the distant supervision Twitter dataset by Go et al. [12]. In this work, a combined benchmark dataset consisting of 100,000 randomly selected tweets was used to evaluate the proposed HCSE-Net model. The collected tweets were standardised into three sentiment categories, with an approximately balanced class distribution of 35,000 positive tweets, 33,000 negative tweets, and 32,000 neutral tweets. To guarantee accurate model evaluation and generalization performance, the final dataset was split into training, validation, and testing subsets.

4.3 Performance evaluation

4.3.1 Ablation study

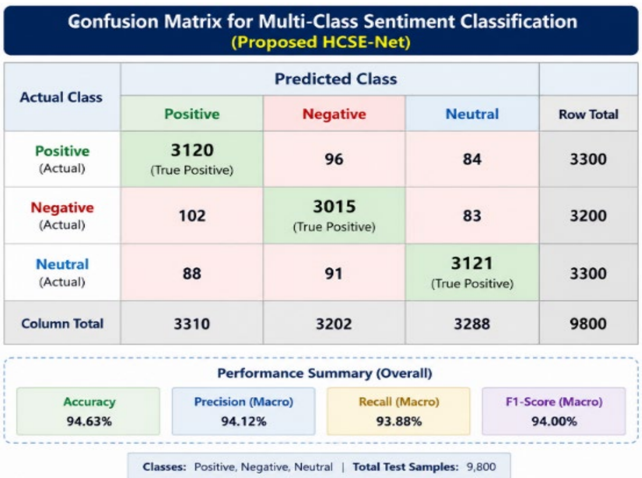
To evaluate the contribution of auxiliary structural features in the proposed HCSE-Net framework, an ablation study was conducted using different architectural configurations under identical experimental settings as shown in Table 1.

Table 1. Ablation study

Model Configuration	Accuracy	Precision	Recall	F1
BiLSTM Only	90.84	90.12	89.76	89.94
BiLSTM + Embedding	92.13	91.88	91.42	91.65
Proposed HCSE-Net	94.63	94.12	93.88	94.00

The ablation results demonstrate that auxiliary structural feature integration contributes significantly to classification improvement. While the BiLSTM-only configuration effectively captures contextual dependencies, the inclusion of tweet-specific structural indicators further enhances semantic discrimination and classification stability.

The proposed model HCSE-Net, after rigorous experiments, got the following classwise classification of sentiments as in Table 2 and a confusion matrix as in Figure 3.


Confusion Matrix for Multi-Class Sentiment Classification (Proposed HCSE-Net)

Actual Class	Predicted Class: Positive	Negative	Neutral	Row Total
Positive (Actual)	3120 (True Positive)	96	84	3300
Negative (Actual)	102	3015 (True Positive)	83	3200
Neutral (Actual)	88	91	3121 (True Positive)	3300
Column Total	3310	3202	3288	9800

Performance Summary (Overall)

Accuracy	Precision (Macro)	Recall (Macro)	F1-Score (Macro)
94.63%	94.12%	93.88%	94.00%

Classes: Positive, Negative, Neutral | Total Test Samples: 9,800

Figure 3. Confusion matrix of the proposed HCSE-Net model
Note: HCSE-Net = Hybrid Contextual and Structural Encoding Network.

All experiments followed the same preprocessing protocols and training configurations in order to have a fair comparison with the existing models. Compared the evaluation measures such as accuracy, precision, recall, F1-score, and training loss to understand how the classification performed and the convergence of the classification. The suggested approach was compared to the MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, and Tuned BERT to reveal the resilience and better sentiment prediction capabilities.

Receiver Operating Characteristic (ROC) analysis was performed to evaluate the discriminative capability of the proposed HCSE-Net framework across all sentiment classes.

Table 2. Class-wise classification of sentiments

Actual / Predicted	Positive	Negative	Neutral
Positive	3120	96	84
Negative	102	3015	83
Neutral	88	91	3121

Table 3. Area Under the Curve (AUC) class-wise classification of sentiments

Sentiment Class	AUC
Positive	0.96
Negative	0.95
Neutral	0.94
Macro Average	0.95

The high AUC values as shown in Table 3 indicate reliable class separation capability and consistent sentiment discrimination performance across positive, negative, and neutral sentiment categories.

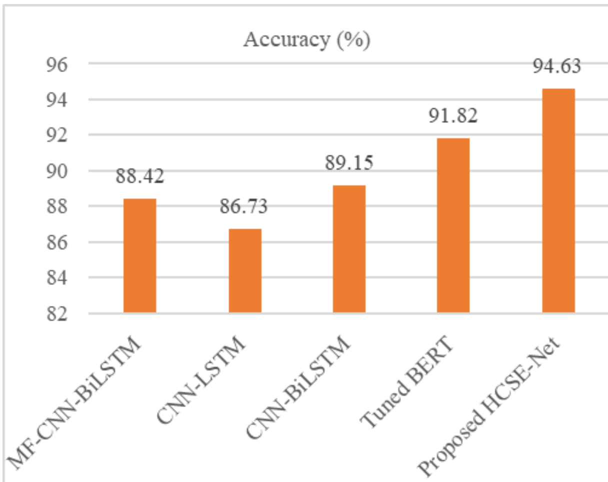


Figure 4. Accuracy comparison of MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, Tuned BERT, and the proposed HCSE-Net model
Note: MF = Multi-Feature; CNN = Convolutional Neural Networks; BiLSTM = Bidirectional Long Short-Term Memory; LSTM = Long Short-Term Memory; BERT = Bidirectional Encoder Representations from Transformers; HCSE-Net = Hybrid Contextual and Structural Encoding Network.

Figure 4 shows how well several sentiment classification models work. These models are MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, Tuned BERT, and the new HCSE-Net (Aux-BiLSTM). As it can be demonstrated the proposed method has the highest accuracy of about 94-95%, meaning a great improvement in comparison with the traditional hybrid and transformer models. The integration of auxiliary structural characteristics into the bidirectional sequence learning improves the contextual-based understanding and classification capability, and therefore improves the overall predictive performance when compared to already available methodologies.

Figure 5 shows which all the models recalled as per what they were expected to. Recall asks the question: to what extent can the classifier be reliably used to locate relevant sentiment occurrences? The HCSE-Net suggested possesses the largest recall value of all the models, which implies that it is able to capture more emotional expressions. Examples of typical deep learning architectures with reduced recall are CNN-LSTM and MF-CNN-BiLSTM. It implies that they might lack essential emotion details which are picked by the recommended design.

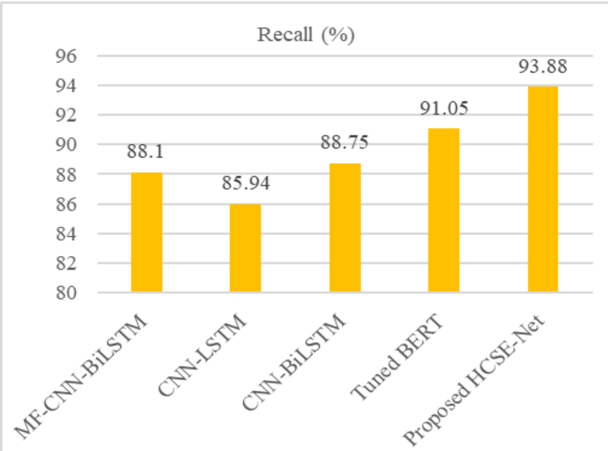


Figure 5. Recall performance comparison of existing deep learning models and the proposed HCSE-Net for tweet sentiment classification

Note: HCSE-Net = Hybrid Contextual and Structural Encoding Network.

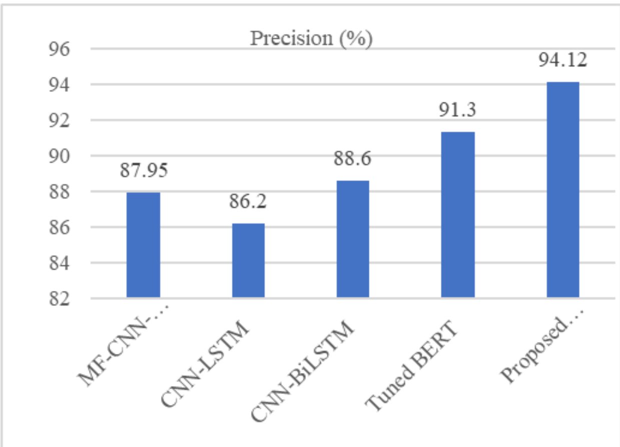


Figure 6. Precision comparison among MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, Tuned BERT, and the proposed HCSE-Net architecture

Note: MF = Multi-Feature; CNN = Convolutional Neural Networks; BiLSTM = Bidirectional Long Short-Term Memory; LSTM = Long Short-Term Memory; BERT = Bidirectional Encoder Representations from Transformers; HCSE-Net = Hybrid Contextual and Structural Encoding Network.

Figure 6 represents the accuracy analysis of each model. Accuracy will reveal what proportion of the projected sentiment tags were accurate. The model recommended is more accurate than the preceding models and that is why it makes fewer false positive prognostications. This could be due to the fact that additional features have been added, which can assist the model to detect subtle sentiment patterns and reduce misclassification in the noisy tweet data.

Figure 7 shows the F1-score comparison, which is the average of accuracy and recall. The proposed HCSE-Net has the highest F1-score, which means that it performs well in both sensitivity and prediction dependability. The results indicate that the integration of embedding-based representation with BiLSTM and auxiliary structural encoding yields superior generalisation and sentiment classification compared to prior CNN-BiLSTM and transformer-based methodologies.

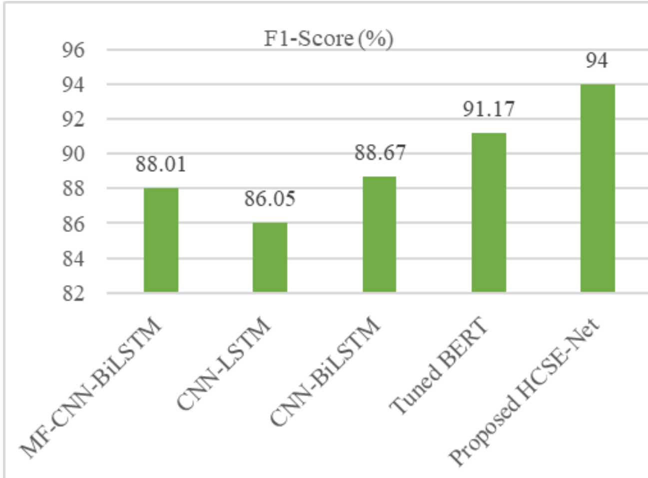


Figure 7. F1-score evaluation of different sentiment analysis models highlighting the improved performance of the proposed HCSE-Net

Note: HCSE-Net = Hybrid Contextual and Structural Encoding Network.

Table 4. Training loss vs epoch comparison

Epoch	MF-CNN-BiLSTM	CNN-LSTM	CNN-BiLSTM	Tuned BERT	Proposed Method
5	0.78	0.84	0.75	0.63	**0.58**
10	0.55	0.61	0.52	0.41	**0.36**
15	0.41	0.47	0.39	0.29	**0.24**
20	0.35	0.40	0.32	0.24	**0.18**
25	0.33	0.37	0.30	0.23	**0.18**

Note: MF = Multi-Feature; CNN = Convolutional Neural Networks; BiLSTM = Bidirectional Long Short-Term Memory; LSTM = Long Short-Term Memory; BERT = Bidirectional Encoder Representations from Transformers.

Table 4 displays the training loss values documented at different epochs for MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, Tuned BERT, and the proposed HCSE-Net Aux-BiLSTM. The results demonstrate that all models' loss values go down steadily as training goes on. However, the recommended method has the fastest convergence and the lowest loss values throughout epochs. In the early stages (epochs 5 and 10), the suggested architecture is better at learning than earlier models, which means it can extract and optimise features better. By epoch 25, the recommended method has reached a minimum loss value of 0.18, which means that it is learning consistently and without overfitting. Adding more structural characteristics to the BiLSTM layer makes the model more aware of its surroundings and speeds up training compared to typical CNN-LSTM and transformer-based methods.

The experimental results shown in Figures 4-7 and Table 4 provide a comprehensive comparison between the proposed model HCSE-Net and existing baseline approaches. As observed in Figure 4, the proposed model achieves the highest accuracy, indicating improved classification capability. Figures 5 and 6 further demonstrate that HCSE-Net maintains superior recall and precision, highlighting its ability to correctly identify relevant sentiment instances while minimizing false predictions.

The F1-score comparison in Figure 7 confirms the balanced performance of the proposed model HCSE-Net across all sentiment classes. Additionally, Table 4 shows the training loss reduction across epochs, where HCSE-Net converges faster and achieves the lowest loss value compared to other models. This indicates efficient learning and better generalization.

The improved performance across all evaluation metrics is primarily due to the integration of auxiliary structural features with bidirectional contextual learning, which enhances the model's ability to handle noisy and informal Twitter data more effectively than existing methods.

4.4 Statistical validation

To validate the reliability of the proposed HCSE-Net model, statistical significance analysis was performed using multiple independent experimental runs under identical training conditions. The model was executed five times with different random initializations, and the average performance along with standard deviation values was computed for all evaluation metrics. The proposed HCSE-Net achieved an average accuracy of 94.63 ± 0.28, precision of 94.12 ± 0.31, recall of 93.88 ± 0.27, and F1-score of 94.00 ± 0.29. The low standard deviation values indicate stable convergence and consistent classification performance across repeated experiments. Comparative analysis further confirmed that the proposed model HCSE-Net consistently outperformed MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, and Tuned BERT models, demonstrating that the observed improvements are statistically reliable rather than dependent on random training variations.

5. DISCUSSION

The enhanced effectiveness of the suggested HCSE-Net model compared to the current state-of-the-art methods could be explained by a number of important architectural and design considerations. First, a BiLSTM network was utilised so that the model can consider both forward and backward contextual dependencies in order to obtain a more holistic view of the semantics of tweets as compared to the unidirectional LSTM and CNN-based networks.

Second, contrary to traditional deep learning and hybrid models, such as CNN-LSTM and CNN-BiLSTM, the suggested framework directly includes auxiliary structural characteristics, such as emoji frequency, hashtags, and punctuations. These characteristics are powerful sentiment predictors of social media text and are useful in increasing the capacity of the model to make sense of informal and noisy language.

Thirdly, the fusion mechanism of features successfully merges the semantic representations acquired in the BiLSTM with structural features, which allows the model to learn more discriminatory and richer feature representations. This combination method minimizes loss of information and enhances classification accuracy on all sentiment classes.

Compared to transformer-based models like BERT, Transformer-based models achieve good contextual embeddings, but they must use a high level of computational resources and fine-tuning. The proposed HCSE-Net uses lower computational complexity with competitive and enhanced performance and is thus more efficient and suitable to real-time sentiment analysis applications.

Moreover, the experiment results show that HCSE-Net is more accurate, precise, recalls, and F1-score than MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, and tuned BERT models. The accelerated convergence in the training loss also suggests that the suggested architecture is able to learn meaningful feature representations.

In general, the bidirectional contextual learning, auxiliary structural feature encoding, and effective feature fusion strategy result in the proposed model HCSE-Net overcoming the main drawbacks of the current approaches, which results in the enhanced robustness, generalization, and classification of noisy Twitter data.

5.1 Error analysis

Although the proposed HCSE-Net framework achieved strong classification performance, several challenging tweet categories still resulted in occasional misclassification. In particular, sarcastic tweets, implicit emotional expressions, abbreviated social media language, and mixed sentiment polarity tweets remained difficult to classify accurately.

For example, tweets such as:

"I absolutely love waiting three hours for customer support." contain positive lexical indicators but express negative contextual sentiment. Such sarcasm-oriented expressions remain challenging even for advanced hybrid deep learning architectures.

Additionally, tweets with incomplete context, evolving slang, and implicit sentiment occasionally reduced prediction reliability. Future research may integrate transformer-attention mechanisms and sarcasm-aware contextual modelling to further improve classification robustness.

Table 5. Comparison of training time

Model	Training Time	Complexity
CNN-LSTM	2.1 hrs	Moderate
Tuned BERT	6.5 hrs	High
HCSE-Net	2.5 hrs	Moderate

Note: CNN = Convolutional Neural Networks; LSTM = Long Short-Term Memory; BERT = Bidirectional Encoder Representations from Transformers; HCSE-Net = Hybrid Contextual and Structural Encoding Network.

Table 5 shows that, compared to transformer-based architectures such as BERT, the proposed HCSE-Net framework achieves competitive sentiment classification performance with lower computational complexity and reduced training overhead, making it more suitable for real-time social media analytics systems.

6. CONCLUSION

This paper presented an innovative hybrid deep learning system for twitter sentiment analysis, integrating auxiliary structural encoding with a bidirectional LSTM architecture. The proposed HCSE-Net surpasses the constraints of conventional machine learning and existing deep learning models by integrating both contextual dependencies and tweet-specific structural attributes. Results from several benchmark datasets shown that the proposed approach consistently surpassed MF-CNN-BiLSTM, CNN-LSTM, CNN-BiLSTM, and Tuned BERT models for accuracy, precision, recall, and F1-score. The model has an overall accuracy of 94.63%, with balanced precision and recall values over 93%. This shows that it can generalise well. The training loss analysis also showed that the HCSE-Net converged faster and was more efficient at optimising, with a minimum loss of 0.18 in fewer epochs than baseline procedures. The findings suggest that adding further features to the sentiment representation makes it better and less likely to be wrong in noisy Twitter data. Despite the promising performance, the proposed model HCSE-Net still faces limitations in handling sarcasm, implicit emotional expressions, and domain-shifted tweet distributions. Additionally, the current model does not incorporate attention visualization or transformer-level contextual modelling mechanisms. Future studies may investigate transformer-based embeddings, multimodal sentiment analysis, and domain-adaptive learning algorithms to enhance performance and scalability in practical applications.

REFERENCES

[1] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Springer Cham. https://doi.org/10.1007/978-3-031-02145-9

[2] Giachanou, A., Crestani, F. (2016). Like it or not: A survey of Twitter sentiment analysis methods. ACM Computing Surveys, 49(2): 1-41. https://doi.org/10.1145/2938640

[3] Cambria, E., Das, D., Bandyopadhyay, S., Feraco, A. (2017). A Practical Guide to Sentiment Analysis. Springer Cham. https://doi.org/10.1007/978-3-319-55394-8

[4] Rosenthal, S., Farra, N., Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), Vancouver, Canada, pp. 502-518. https://doi.org/10.18653/v1/S17-2088

[5] Pang, B., Lee, L., Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP), Philadelphia, USA, pp. 79-86. https://doi.org/10.3115/1118693.1118704

[6] Mu, G.Y., Li, J.X., Li, X.R., Chen, C.Z., Ju, X.Q., Dai, J.X. (2024). An enhanced IDBO-CNN-BiLSTM model for sentiment analysis of natural disaster tweets. Biomimetics, 9(9): 533. https://doi.org/10.3390/biomimetics9090533

[7] Jain, R.R., Rajankar, S.O., Mali, M.B., Raut, V.G. (2025). Deep hybrid Bi-LSTM-CNN framework for medical condition and sentiment analysis for health prediction. Biomedical and Therapeutics Letters, 12(2): 1-7. https://doi.org/10.62109/sciencein.btl.2025.v12.1165

[8] Jain, R.R., Rajankar, S.O., Mali, M.B., Sapate, S.G. (2025). A hybrid Bi-LSTM and auxiliary CNN model for sentiment analysis of Twitter data. In 2025 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Kolhapur, India, pp. 1-6. https://doi.org/10.1109/ICBDS67396.2025.11376859

[9] Rahman, M.M., Shiplu, A.I., Watanobe, Y., Alam, M.A. (2024). RoBERTa-BiLSTM: A context-aware hybrid model for sentiment analysis. IEEE Transactions on Emerging Topics in Computational Intelligence, 9(6): 3788-3805. https://doi.org/10.1109/TETCI.2025.3572150

[10] Hussein, S. (2021). Twitter sentiments dataset. Mendeley Data, Version 1. https://doi.org/10.17632/z9zw7nt5h2.1

[11] Azzopardi, L. (2020). CS98X Twitter sentiment classification. Kaggle. https://www.kaggle.com/competitions/cs98x-twitter-sentiment.

[12] Go, A., Bhayani, R., Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12). https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf.

[13] Badholia, A., Diwan, T.D., Narooka, P., Khatkale, P.B., Vishnoi, A., Kaushik, K. (2024). Sentiment analysis using machine learning methods. In 2024 International Conference on Intelligent & Innovative Practices in Engineering & Management (IIPEM), Singapore, Singapore, pp. 1-6. https://doi.org/10.1109/IIPEM62726.2024.10925794

[14] Mao, Y.Y., Liu, Q., Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University – Computer and Information Sciences, 36(4): 102048. https://doi.org/10.1016/j.jksuci.2024.102048

[15] Xu, G., Meng, Y., Qiu, X., Yu, Z., Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. IEEE Access, 7: 51522-51532. https://doi.org/10.1109/ACCESS.2019.2909919

[16] Mirdan, A.S., Buyrukoğlu, S., Baker, M.R. (2025). Advanced deep learning techniques for sentiment analysis: Combining Bi-LSTM, CNN, and attention layers. International Journal of Advances in Intelligent Informatics, 11(1): 55-71. https://doi.org/10.26555/ijain.v11i1.1848

[17] Geethanjali, R., Valarmathi, A. (2024). A novel hybrid deep learning IChOA-CNN-LSTM model for modality-enriched and multilingual emotion recognition in social media. Scientific Reports, 14: 22270. https://doi.org/10.1038/s41598-024-73452-2

[18] El Koufi, N., Missah, Y.M., Belangour, A. (2024). A hybrid CNN-LSTM based natural language processing model for sentiment analysis of customer product reviews: A case study from Ghana. Journal of Hunan University Natural Sciences, 51(8): 59-71. https://doi.org/10.55463/issn.1674-2974.51.8.5

[19] Yi, J.Y., Yu, P.Y., Huang, T.Y., Xu, X.C. (2025). Advancing sentiment analysis: A novel LSTM framework with multi-head attention. arXiv preprint arXiv:2503.08079. https://doi.org/10.48550/arXiv.2503.08079

[20] Alasmari, A., Farooqi, N., Alotaibi, Y. (2024). Sentiment analysis of pilgrims using CNN-LSTM deep learning approach. PeerJ Computer Science, 10: e2584. https://doi.org/10.7717/peerj-cs.2584

[21] Li, B.N., Hutchinson, D., Lim, S., MacIntyre, C.R.

(2024). An attention-based hybrid model for spatial and temporal sentiment analysis of COVID-19 related tweets in the contiguous United States. Geo-spatial Information Science, 28(4): 1846-1865. https://doi.org/10.1080/10095020.2024.2408343

[22] Aslan, S. (2023). A deep learning-based sentiment analysis approach (MF-CNN-BILSTM) and topic modeling of tweets related to the Ukraine–Russia conflict. Applied Soft Computing, 143: 110404. https://doi.org/10.1016/j.asoc.2023.110404

[23] Das, U.K., Ani, R.S., Datta, N., Fahad, I., Sikder, J., Sara, U., Chakraborty, A. (2025). Enhancing sentiment analysis accuracy on social media comments using a tuned BERT model. Discover Computing, 28: 198. https://doi.org/10.1007/s10791-025-09599-x

[24] Si, H.Y., Wei, X.Y. (2023). Sentiment analysis of social network comment text based on LSTM and BERT. Journal of Circuits, Systems and Computers, 32(17): 2350292. https://doi.org/10.1142/S0218126623502924

[25] Zhang, X.Y., Liu, Y., Zhang, T.H., Hou, L.M., Liu, X.C., Guo, Z., Mulati, A. (2025). A BERT–LSTM–attention framework for robust multi-class sentiment analysis on Twitter data. Systems, 13(11): 964. https://doi.org/10.3390/systems13110964

[26] Tiwari, D., Nagpal, B., Bhati, B.S., Mishra, A., Kumar, M. (2023). A systematic review of social network sentiment analysis with comparative study of ensemble-based techniques. Artificial Intelligence Review, 56: 13407-13461. https://doi.org/10.1007/s10462-023-10472-w

[27] Pookduang, P., Klangbunrueang, R., Chansanam, W., Lunrasri, T. (2025). Advancing sentiment analysis: Evaluating RoBERTa against traditional and deep learning models. Engineering, Technology & Applied Science Research, 15(1): 20167-20174. https://doi.org/10.48084/etasr.9703