ISSN 2579-7298

A Multi-Feature Fusion Framework for Sentiment Analysis Based on Textual and Affective Signals

Hussein Ala'a Alkaabi [a], ali kadhim jasim [b]

[a] Ministry of Education Iraq, General Directorate of Education in Al-Najaf, Al-Najaf 54001, Iraq., Iraq
[b] Department of Computer Engineering, Imam Ja'afar Al-Sadiq University, Maysan - Iraq
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Sentiment Analysis,
Feature Fusion,
Textual Representation,
Emotion Signals,
Deep Learning

ABSTRACT

Sentiment analysis of social media content, particularly on platforms like Twitter, presents significant challenges due to the informal, brief, and context-dependent nature of user-generated text. Traditional lexicon-based and shallow machine learning approaches often fail to capture nuanced sentiment expressions, especially in the presence of slang, abbreviations, sarcasm, and emotionally charged language. To address these limitations, this paper proposes a novel tri-stream feature fusion framework that integrates contextual semantics, sequential dependencies, and affective signals for robust sentiment classification. The framework employs RoBERTa to extract rich contextual embeddings, Bidirectional Long Short-Term Memory (BiLSTM) networks to capture word-order and temporal patterns, and lexicon-based emotion vectors to enhance emotional cue detection. These heterogeneous features are concatenated at the representation level to form a comprehensive feature space, which is subsequently used to predict sentiment polarity via a fully connected neural network classifier. Extensive experiments conducted on the Sentiment140 dataset, comprising 1.6 million labeled tweets, demonstrate that the proposed approach significantly outperforms conventional baselines and recent hybrid models, achieving an accuracy of 92.1%. Additionally, ablation studies and misclassification analyses reveal each feature stream's complementary contributions and highlight challenges in detecting sarcasm and implicit sentiment. Future work will integrate sarcasm-aware components and external knowledge sources to further enhance model interpretability and robustness.

1. Introduction

In recent years, sentiment analysis has emerged as a critical task in natural language processing (NLP), with widespread applications in social media monitoring, customer feedback analysis, political forecasting, and public opinion mining [1]. Twitter has become a prominent source for sentiment mining across various platforms due to its massive volume of user-generated content, brevity of expression, and real-time nature [2]. Despite its importance, sentiment classification on Twitter data poses several challenges. Tweets, like other short texts such as SMS messages, are typically brief, informal, and often include slang, abbreviations, emojis, and domain-specific language [3]. These characteristics render traditional bag-of-words or lexicon-based approaches insufficient, as they usually fail to capture the nuanced, context-dependent nature of sentiment expression. Furthermore, sarcasm, ambiguity, and emotionally charged language compound the complexity of accurate classification [4]. To address these limitations, recent research has shifted towards deep learning and transformer-based models that capture semantic and syntactic features

from text [5]. However, relying solely on a single representation, such as contextual embeddings or sequential features, may not fully exploit the diverse linguistic cues embedded in user posts[6]. This limitation motivates the need for multi-perspective feature modeling. In this paper, we propose a novel tri-stream feature fusion framework for sentiment classification that integrates three complementary sources of information. First, contextual semantics are captured using the RoBERTa transformer to understand the nuanced meaning of text within context [7]. Second, sequential dependencies are extracted via Bidirectional Long Short-Term Memory (BiLSTM) networks, which effectively model the input's word order and temporal relationships [8]. Third, affective signals are derived from lexicon-based emotion vectors, providing explicit emotional cues to enhance sentiment detection [9]. This multi-faceted approach aims to improve classification performance by leveraging the strengths of each feature stream. These heterogeneous representations are concatenated at the feature level to form a rich, unified embedding, which then passes through a classification layer to predict sentiment polarity. We evaluate our model on the Sentiment140 dataset, which comprises 1.6 million labeled tweets [10]. The results demonstrate that our approach outperforms traditional baselines and single-stream models, achieving significant improvements in accuracy and robustness.
The key contributions of this study are as follows:

- We introduce a multi-stream feature extraction architecture that jointly models tweets' contextual, sequential, and emotional aspects.
- We perform extensive evaluation on a large-scale real-world dataset and demonstrate the effectiveness of feature-level fusion.
- We provide an ablation study and error analysis to investigate each component's impact and understand model behavior on challenging inputs.

The remainder of this paper is organized as follows: Section 2 discusses related work. Section 3 details the proposed methodology. Section 4 outlines the experimental setup, and Section 5 presents the results and analysis. Section 6 concludes the paper with potential directions for future research.

2. Related work

Twitter sentiment analysis has been widely studied due to the abundance of real-time user opinions, which serve as rich sources of real-time public opinion. Various techniques have been explored to improve the accuracy and efficiency of sentiment classification, ranging from traditional machine learning methods to advanced deep learning architectures. Ensemble approaches combining classifiers such as Support Vector Machine (SVM), Naive Bayes, and Long Short-Term Memory (LSTM) have demonstrated improved accuracy on the Sentiment140 dataset, outperforming individual models and supporting applications in business analytics [11]. Comparative studies on classical machine learning algorithms applied to social media data, including Sentiment140 and Twitter-Reddit datasets, have shown that Linear SVC often achieves superior accuracy, reaching up to 81% [12]. Moreover, gradient boosting methods such as LightGBM have been reported to outperform SVM, achieving accuracy levels up to 91%, highlighting their suitability for large-scale, real-time sentiment classification [13]. Deep learning models, especially LSTM, have effectively captured temporal dependencies in tweets, with distant supervision via emoticons enhancing model validation on noisy data [14]. Recently, hybrid models combining transformer-based embeddings (BERT, RoBERTa, DistilBERT) with sequences such as BiLSTM and CNN have gained attention. On Sentiment140, DistilBERT-BiLSTM achieved the highest accuracy of 81%, indicating the value of integrating contextual and sequential features for nuanced sentiment analysis [15]. Furthermore, ConvLSTM architectures with word-level attention have been proposed to focus on salient features, outperforming baseline methods across multiple datasets, including Sentiment140 [16]. A semi-automatic multi-label annotation approach was developed, involving corpus selection, emotional tag assignment, preprocessing, automatic annotation via word matching and weighting, and manual correction for ambiguous cases. Experiments on the Sentiment140 Twitter corpus demonstrated the method's effectiveness, achieving over 70% accuracy and strong consistency with manual annotations. This process facilitated the creation of a multi-label emotion corpus of 6,500 tweets for subsequent model training [17]. Ouni et al [18] propose a novel Soft EMBedding (SoftEMB) method that integrates GloVe and Word2Vec embeddings via a soft-voting algorithm to enhance representation quality. Evaluated across various hybrid sentiment analysis models, including CNN-LSTM, CNN-GRU, CNN-BiLSTM, and CNN-BiGRU, SoftEMB demonstrated notable accuracy improvements. On movie review datasets, the method achieved approximately 88.3% accuracy, while on the Sentiment140 dataset, it achieved accuracies above 82.5%, underscoring its effectiveness across diverse sentiment classification tasks. Several studies have addressed sentiment analysis using various datasets, such as IMDB for movie reviews [19] and Amazon Reviews for product evaluations [20]. While these datasets have proven useful, they are often limited in size or topical diversity. In contrast, the current study utilizes the Sentiment140 dataset, which is significantly larger, comprising over 1.6 million tweets and covering a wide range of real-world topics. This diversity offers a more comprehensive representation of sentiment across different contexts.

3. Method

This section presents the proposed approach for sentiment classification. Figure 1 illustrates the overall architecture of the proposed tri-stream sentiment classification framework, highlighting the three parallel feature extraction streams and the subsequent fusion process.

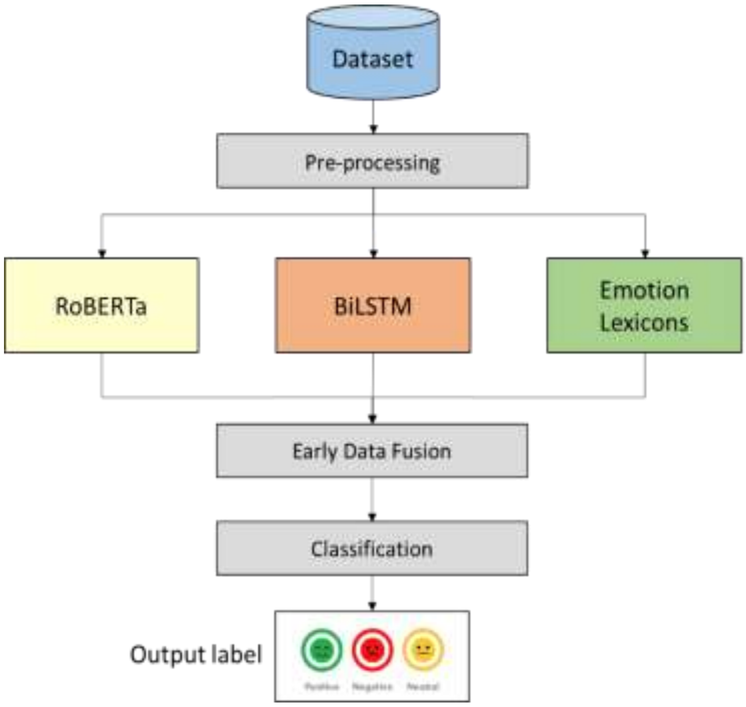

Figure 1. Architecture of the proposed tri-stream sentiment classification framework integrating contextual, sequential, and affective feature streams.

3.1 Data Pre-processing

A comprehensive pre-processing pipeline was applied to the raw tweets to ensure the quality and consistency of the input data. This process includes removing URLs, user mentions, hashtags, and non-alphanumeric characters [21]. Emojis and emoticons were excluded to maintain consistency across feature extraction channels, unless explicitly used in the emotional feature stream. All text was converted to lowercase to minimize sparsity. Common stop words were removed, and tokenization was performed using a subword tokenizer compatible with transformer-based models. This step ensures the textual input is normalized and ready for multi-stream representation [22].

3.2 Feature Extraction

To capture rich and diverse representations of sentiment-bearing texts, we designed a triple-stream feature extraction mechanism, where each stream is responsible for capturing different linguistic and semantic characteristics:

3.2.1 Contextual Representation via RoBERTa

RoBERTa, a robustly optimized BERT variant, captures contextual dependencies within the tweet. Each input sequence is tokenized and passed through the pre-trained Roberta-base model. The [CLS] token embedding from the final hidden layer is extracted to represent the tweet's overall semantic context [23].

3.2.2 Sequential Features via BiLSTM

A Bidirectional LSTM network extracts the sequential structure and word order of tweets. Input sequences are transformed using a pre-trained word embedding layer and then passed to a BiLSTM layer [24]. The concatenated hidden states from the final forward and backward passes are used as the sequential feature vector.

3.2.3 Affective Features via Emotion Lexicons

Emotion-based features are extracted using a lexicon-driven approach to enhance the model's awareness of affective signals. Each tweet is analyzed using the NRC Emotion Lexicon to identify word-level associations with emotions such as joy, anger, sadness, fear, trust, and disgust [25]. These emotion scores are aggregated into a fixed-size vector representing the tweet's affective profile.

3.3 Feature Fusion via Early Data Fusion

We adopt an early data fusion strategy to integrate the extracted features' complementary strengths [26]. In this approach, the contextual embeddings obtained from RoBERTa, the sequential features derived from the BiLSTM network, and the affective representations produced through emotion lexicons are concatenated at the feature level to form a unified representation. This early fusion technique ensures that diverse linguistic and emotional signals are combined before classification, enabling the model to leverage both deep contextual semantics and explicit affective cues simultaneously. By aligning and merging these heterogeneous feature vectors into a single high-dimensional representation, the model benefits from a richer, more discriminative feature space, leading to improved sentiment classification performance. Equation 1 shows the fusion mechanism. Let us consider that Fctx represents Contextual features from RoBERTa, Fseq represents Sequential features from BiLSTM, and Faff represents Affective (emotional) features from lexicon-based vectors.

3.4 Classification

The fused feature vector is fed into a fully connected feed-forward neural network for sentiment classification. The final classification layer employs a softmax activation function to predict one of the predefined sentiment categories (positive, negative, or neutral). The network is trained using cross-entropy loss and optimized using the Adam optimizer. Dropout regularization is applied to prevent overfitting [27].

4. Evaluation and Experimental Results

This section outlines the experimental setup for evaluating the proposed multi-stream sentiment classification framework. We present details of the dataset, model configuration, evaluation metrics, and a comprehensive analysis of the obtained results.

4.1 Dataset Description

The proposed model was evaluated using the Sentiment140 dataset [10], a widely recognized benchmark corpus for sentiment analysis in the Twitter domain. This dataset comprises 1.6 million labeled tweets collected through the Twitter API. Each tweet is automatically annotated using distant supervision based on the presence of emoticons, allowing for large-scale sentiment labeling without manual annotation. The sentiment labels are encoded as 0 for negative sentiment, 2 for neutral sentiment, and 4 for positive sentiment. Sentiment140 provides a diverse and realistic set of social media texts, making it well-suited for evaluating models to capture sentiment in noisy, user-generated content. To demonstrate the structure and labeling of data in the Sentiment140 dataset, Table 1 presents real tweet examples with their corresponding sentiment annotations.

Table 1. Examples of tweets from the Sentiment140 dataset with corresponding sentiment annotations.

Sentiment	Tweet ID	Tweet Text
Negative	1467810369	I hate it when people text me, I reply, and they don't reply. 😬
Negative	2191879462	I miss my bed so much right now... 😢
Neutral	1621183934	Just got home, now relaxing.
Neutral	2179630032	Going to the store to get some stuff.
Positive	1767284741	I love sunny mornings like this one! ☀□
Positive	2157288231	Awesome day with my best friend! 😄

Our experiments excluded the neutral class to focus on binary sentiment classification (positive vs. negative), which aligns with previous literature. After pre-processing, the dataset was split into 10 folds using 10-fold cross-validation [28].

4.2 Experimental Setup

The model was trained and evaluated using a standardized configuration across all components to ensure fair and reproducible experimentation. The RoBERTa model (roberta-base) from the HuggingFace Transformers library was employed, with a hidden size of 768. The BiLSTM layer comprised 128 units (64 per direction) and used 100-dimensional GloVe embeddings. Emotion-based features were represented as an 8-dimensional vector derived from the NRC Emotion Lexicon. Feature fusion was achieved by concatenating the feature-level outputs of all three streams. The classification module was a fully connected neural network with a softmax activation function for multi-class sentiment prediction. The model was trained with a batch size of 32, optimized using the Adam optimizer and categorical cross-entropy loss. Evaluation metrics included accuracy, precision, recall, and F1-score. All experiments were conducted using PyTorch on a single NVIDIA GeForce RTX 3090 GPU.

4.3 Evaluation Metrics

Standard classification metrics were used to evaluate the proposed model's performance. These metrics offer a comprehensive assessment of predictive accuracy and robustness. They are instrumental in handling class imbalance, which is common in NLP tasks [29].

- **Accuracy**: The ratio of correctly predicted instances to the total number of cases, indicating the overall correctness of the model.

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + FP + TN + FN)} * 100\% \quad (2)$$

- **Precision**: The proportion of true positive predictions among all instances predicted as positive, reflecting the model's ability to avoid false positives.

$$\text{Precision} = \frac{(TP)}{(TP + FP)} \quad (3)$$

- **Recall**: The proportion of actual positive instances correctly identified by the model, measuring its ability to minimize false negatives.

$$\text{Recall} = \frac{(TP)}{(TP + FN)} \quad (4)$$

- **F1-Score**: The harmonic mean of precision and recall provides a balanced measure that accounts for false positives and false negatives.

$$\text{F1 score} = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \quad (5)$$

These metrics provide a balanced view of the classifier's effectiveness, particularly in class-imbalanced settings [30].

5. Results and Analysis

This section presents the experimental findings of evaluating the proposed tri-stream sentiment classification framework. We comprehensively analyze the model's performance compared to traditional baselines and state-of-the-art methods. The evaluation is conducted on the Sentiment140 dataset using standard classification metrics, including accuracy, precision, recall, and F1-score.

5.1 Quantitative Performance Evaluation

Table 2 summarizes the accuracy achieved by various existing sentiment classification models compared to our proposed approach. The baseline models include traditional machine learning classifiers, such as Naive Bayes and SVM, with TF-IDF features, as well as hybrid deep learning architectures, such as DistilBERT-BiLSTM and SoftEMB-based CNN-LSTM variants. As depicted, the proposed method surpasses all baseline configurations, achieving an accuracy of 92.1%, significantly outperforming existing single-stream and dual-stream fusion models.

Table 2. Comparative analysis of existing sentiment classification models and the proposed approach on the Sentiment140 dataset.

Study / Model	Features Used	Fusion Type	Accuracy (%)
Naive Bayes with TF-IDF	Lexical (TF-IDF)	Single	78.5
SVM with TF-IDF	Lexical (TF-IDF)	Single	79.0

DistilBERT + BiLSTM	Contextual + Sequential	Late Fusion	81.0
Attention-based ConvLSTM	Sequential + Attention	Attention-based	82.0
SoftEMB + CNN-BiLSTM	Hybrid Word Embeddings + Sequential	Embedding-level Fusion	82.76
SoftEMB + CNN-LSTM/GRU/BiGRU	Hybrid Word Embeddings + Sequential	Embedding-level Fusion	82.51–83.27
Proposed Method (Tri-Stream Fusion)	Contextual + Sequential + Affective	Early Fusion	92.1

The performance gain of nearly 10% over advanced hybrid models, as shown in Figure 2, highlights the efficacy of integrating contextual, sequential, and affective representations through early fusion. This result substantiates the hypothesis that multi-perspective feature fusion enhances sentiment classification performance by capturing richer linguistic cues.

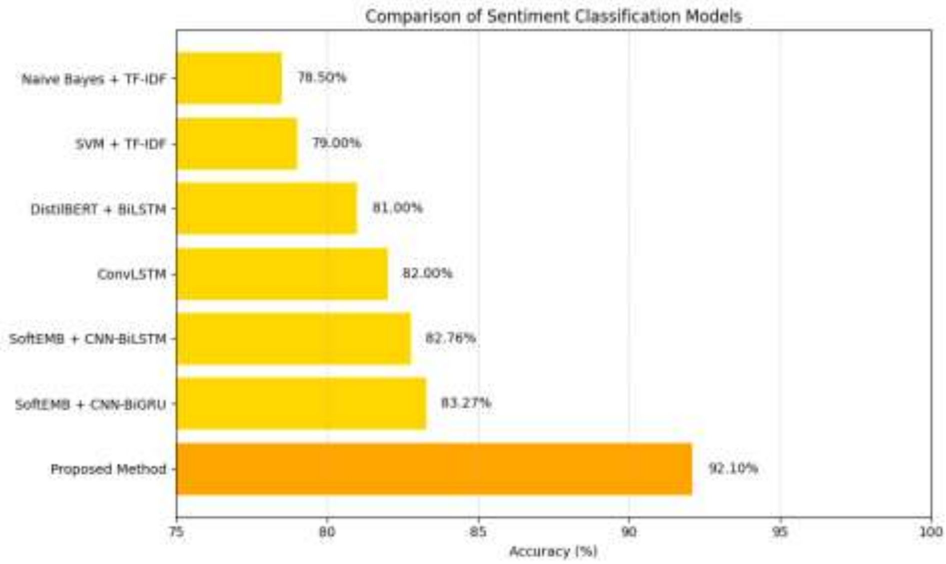

Figure 2. Proposed method performance comparison with the related work models

5.2 Detailed Performance Metrics

Table 3 reports the evaluation metrics, including Precision, Recall, and F1-Score, for the proposed tri-stream model and its component-wise ablations to provide a deeper insight into the model's classification behavior. Specifically, we compare the performance of (i) BiLSTM-only sequential features, (ii) RoBERTa-only contextual embeddings, (iii) dual-stream RoBERTa + BiLSTM fusion, and (iv) the complete tri-stream architecture.

Table 3. Performance metrics (Accuracy, Precision, Recall, F1-Score) of single-stream, dual-stream, and tri-stream models.

Model	Accuracy (%)	Precision	Recall	F1-Score
BiLSTM Only	83.4	0.86	0.84	0.85
RoBERTa Only	87.2	0.89	0.86	0.87
RoBERTa + BiLSTM	90.0	0.93	0.88	0.91
Proposed 3-Stream Model	92.1	0.95	0.93	0.94

The tri-stream model consistently outperforms its single and dual-stream counterparts across all metrics. Notably, including the affective stream, despite its low dimensionality, substantially boosts recall and F1-score. This indicates that emotion-based features are particularly effective in disambiguating sentiments in tweets with subtle or implicit emotional undertones.

5.3 Confusion Matrix Analysis

We generated a confusion matrix heatmap on the test set, as shown in Figure 4, to further understand the model's classification tendencies [31]. The matrix reveals that the proposed model demonstrates high discriminative capability, with minimal false positives and false negatives. However, most misclassifications occur in tweets characterized by sarcasm, idiomatic expressions, or cultural references.

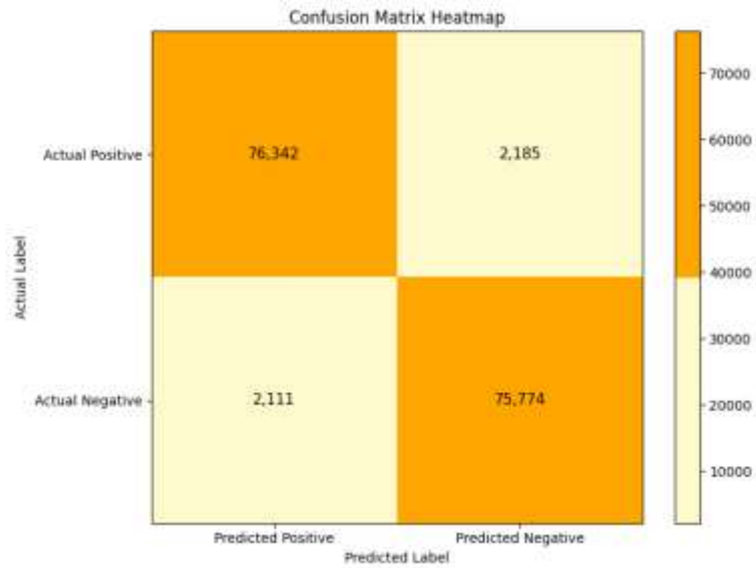

Figure 3. Confusion matrix heatmap illustrating the classification performance of the proposed model on the Sentiment140 test set.

5.4 Misclassification Analysis

Despite the model's strong performance, sarcasm and implicit sentiment remain challenging. Table 4 shows examples of sarcastic tweets misclassified as positive due to misleading lexical cues [32].

Table 4. Examples of Sarcasm-Induced Misclassifications

Tweet Example	Predicted Sentiment	Actual Sentiment	Remarks
"Yeah, I just love when my internet goes down during a meeting…"	Positive	Negative	Sarcasm detected as positive
"Best movie ever! Slept through the entire thing."	Positive	Negative	Sarcastic praise misinterpreted

These cases highlight the model's reliance on literal word meanings, which limits its ability to detect sarcasm or irony. Addressing this issue requires integrating sarcasm-aware classifiers or external commonsense knowledge sources to better capture implicit sentiment patterns in future work.

5. Discussion

The experimental outcomes affirm that fusing heterogeneous contextual, sequential, and affective features within a unified representation space significantly enhances sentiment classification robustness. Notably, the affective stream effectively resolved sentiment ambiguities in tweets lacking explicit emotional indicators. However, challenges remain in accurately detecting sentiments expressed through sarcasm or domain-specific idiomatic phrases. Incorporating sarcasm-aware components and external commonsense knowledge bases could further elevate the model's interpretability and classification precision in future iterations.

6. Conclusion and Future Work

This paper presented a novel tri-stream feature fusion framework for sentiment analysis, which integrates contextual semantics, sequential dependencies, and affective signals to enhance sentiment classification performance on short, informal texts such as tweets. By leveraging RoBERTa for contextual understanding, BiLSTM networks for modeling word-order dependencies, and lexicon-based affective vectors to capture explicit emotional cues, the proposed approach effectively combines multiple linguistic perspectives into a unified feature representation. Experimental evaluations conducted on the Sentiment140 dataset demonstrated that our multi-stream fusion model outperforms traditional machine learning classifiers and recent deep learning baselines, achieving an accuracy of 92.1%. The comprehensive ablation study further confirmed the complementary contributions of each feature stream to the model's overall performance. To mitigate these challenges, future research will focus on integrating sarcasm-aware components, such as specialized transformer models fine-tuned on sarcasm-annotated corpora. Additionally, incorporating external commonsense knowledge graphs and discourse-level information could provide deeper context, enabling the model to interpret implicit sentiment cues better. Another potential avenue is to extend the framework to handle multimodal sentiment analysis by incorporating visual and behavioral signals (e.g., emojis, images, user interaction patterns) to enrich the feature space. Furthermore, exploring adaptive attention mechanisms that dynamically weigh the importance of contextual, sequential, and affective features based on the input's linguistic characteristics may further enhance model robustness and interpretability.

References

[1] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731-5780.

[2] AL-Jumaili, A. S. A., & Tayyeh, H. K. (2020). A hybrid method of linguistic and statistical features for Arabic sentiment analysis. Baghdad Science Journal, 17(1), 26.

[3] Al-Kaabi, H., Darroudi, A. D., & Jasim, A. K. (2024). Survey of SMS spam detection techniques: A taxonomy. *AlKadhim Journal for Computer Science*, *2*(4), 23-34.

[4] Salman Al-Tameemi, I. K., Feizi-Derakhshi, M. R., Pashazadeh, S., & Asadpour, M. (2023). An efficient sentiment classification method with the help of neighbors and a hybrid of RNN models. *Complexity*, *2023*(1), 1896556.

[5] Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, *66*(12), 7305-7361.

[6] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, *8*(4), e1253.

[7] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

[8] Xu, G., Meng, Y., Qiu, X., Yu, Z., & Wu, X. (2019). Sentiment analysis of comment texts based on BiLSTM. *Ieee Access*, *7*, 51522-51532.

[9] Qi, Y., & Shabrina, Z. (2023). Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. *Social network analysis and mining*, *13*(1), 31.

[10] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N project report, Stanford*, *1*(12), 2009.

[11] A. Jazib, W. Tariq and M. Mahmood, "Sentiment Analysis using Ensemble Classifier for Entrepreneurs based on Twitter Analytics," *2022 19th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, Islamabad, Pakistan, 2022, pp. 207-212,

[12] P. Upadhyay, S. Saifi, R. Rani, A. Sharma and P. Bansal, "Machine Learning-Based Sentiment Analysis for the Social Media Platforms," *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, Mathura, India, 2023, pp. 1-5,

[13] C. Anitha, "Real-Time Social Media Sentiment Analysis: A Comparative Study of LightGBM and SVM," *2025 International Conference on Electronics and Renewable Systems (ICEARS)*, Tuticorin, India, 2025, pp. 1947-1950,

[14] Bouassida, Y., & Mezali, H. (2025). Enhancing Twitter sentiment analysis using hybrid transformer and sequence models. *Japan J. Res*, *6*(1), 089.

[15] Wang, S., Sun, F., & Liu, P. (2024, November). A ConvLSTM model with word-level attention for sentiment analysis of review data. In *Fourth International Conference on Advanced Algorithms and Neural Networks (AANN 2024)* (Vol. 13416, pp. 723-729). SPIE.

[16] Ramirez-Alcocer, U.M., Tello-Leal, E., Hernandez-Resendiz, J.D., Romero, G. (2024). A Hybrid CNN-LSTM Approach for Sentiment Analysis. In: Kumar, S., Balachandran, K., Kim, J.H., Bansal, J.C. (eds) Fourth Congress on Intelligent Systems. CIS 2023. Lecture Notes in Networks and Systems, vol 869. Springer, Singapore. https://doi.org/10.1007/978-981-99-9040-5_31

[17] Liu, X., Zhou, G., Kong, M., Yin, Z., Li, X., Yin, L., & Zheng, W. (2023). Developing Multi-Labelled Corpus of Twitter Short Texts: A Semi-Automatic Method. Systems, 11(8), 390. https://doi.org/10.3390/systems11080390

[18] Ouni, C., Benmohamed, E., & Ltifi, H. (2024). Deep learning-based Soft word embedding approach for sentiment analysis. *Procedia Computer Science*, *246*, 1355-1364.

[19] Maas, A., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., & Potts, C. (2011, June). Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 142-150).

[20] McAuley, J., Pandey, R., & Leskovec, J. (2015, August). Inferring networks of substitutable and complementary products. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794).

[21] Alkaabi, H., Jasim, A. K., & Darroudi, A. (2025). From Static to Contextual: A Survey of Embedding Advances in NLP. *PERFECT: Journal of Smart Algorithms*, *2*(2), 64-73.

[22] Nafea, A. A., Muayad, M. S., Majeed, R. R., Ali, A., Bashaddadh, O. M., Khalaf, M. A., ... & Steiti, A. (2024). A brief review on preprocessing text in Arabic language dataset: Techniques and challenges. *Babylonian Journal of Artificial Intelligence*, *2024*, 46-53.

[23] Al-Kabbi, H. A., Feizi-Derakhshi, M. R., & Pashazadeh, S. (2024). A Hierarchical Two-Level Feature Fusion Approach for SMS Spam Filtering. *Intelligent Automation & Soft Computing*, *39*(4).

[24] Wu, K., Peng, X., Chen, Z., Su, H., Quan, H., & Liu, H. (2023). A novel short-term household load forecasting method combined BiLSTM with trend feature extraction. *Energy Reports*, *9*, 1013-1022.

[25] Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, *44*(4), 491-511.

[26] Derakhshi, M. R. F., Zafarani-Moattar, E., Al-Kabi, H. A. A., & Almarashy, A. H. J. (2024). Pclf: parallel cnn-lstm fusion model for sms spam filtering. In *BIO Web of Conferences* (Vol. 97, p. 00136). EDP Sciences.

[27] Alameady, M. H. H., Mosa, M. O., Aljarrah, A. A., & Razzaq, H. S. (2022). Deep convolutional neural network classified the pneumonia and coronavirus diseases (covid-19) by softmax nonlinearity function. *International Journal of Nonlinear Analysis and Applications*, *13*(1), 2245-2251.

[28] Olawale-Shosanya, S. O., Olusanya, O. O., Joseph, A. O., Idowu, K. O., Eriwa, O. B., Adebare, A. O., & Usman, M. A. (2024). A Meta-Ensemble Predictive Model For The Risk Of Lung Cancer. *Al-Bahir*, *5*(1), 4.

[29] Dang, V. M. H., & Verma, R. M. (2024, April). Data quality in nlp: Metrics and a comprehensive taxonomy. In *International Symposium on Intelligent Data Analysis* (pp. 217-229). Cham: Springer Nature Switzerland.

[30] Jasim, A. K., Al-Rikabi, M. R., Al-Ibraheem, F. A., Al-Kaabi, H. A., & Kamber, A. (2025, September). BERT-Enhanced Dual-Attention RNN for Short Text Spam Detection. In *International Conference on Cybersecurity and Artificial Intelligence Strategies* (pp. 58-70). Cham: Springer Nature Switzerland.

[31] Al-Kabbi, H. A., Feizi-Derakhshi, M. R., & Pashazadeh, S. (2023). Multi-type feature extraction and early fusion framework for sms spam detection. *IEEE Access*, *11*, 123756-123765.

[32] Falter, M., Godderis, D., Scherrenberg, M., Kizilkilic, S. E., Xu, L., Mertens, M., ... & Dendale, P. (2024). Using natural language processing for automated classification of disease and to identify misclassified ICD codes in cardiac disease. *European Heart Journal-Digital Health*, *5*(3), 229-234.