European Journal of Computer Science and Information Technology, 14(2), 74-86, 2026
Print ISSN: 2054-0957 (Print)
Online ISSN: 2054-0965 (Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

A Hybrid Deep Learning Framework for Sentiment Analysis Using BERT and BiLSTM on IMDB Dataset

Md Shahin Ali,[1] Md Salman,[2] Shaykat Purokayisto,[3*]
[1,2,3] *School of Artificial Intelligence, Nanjing University of Information Science and Technology*
[*]**Correspondence Email**: shaykatariyan@gmail.com

doi: https://doi.org/10.37745/ejcsit.2013/vol14n27486 Published April 06, 2026

Citation: Ali M.S., Salman M., Purokayisto S. (2026) A Hybrid Deep Learning Framework for Sentiment Analysis Using BERT and BiLSTM on IMDB Dataset, *European Journal of Computer Science and Information Technology*, 14(2), 74-86

Abstract: *Sentiment analysis is a key task in natural language processing (NLP) that focuses on identifying the emotional polarity of textual data. While transformer-based models such as BERT have achieved remarkable performance by generating contextualized word representations, they may not fully capture sequential dependencies in long text sequences. To address this limitation, this paper proposes a hybrid deep learning framework that combines a pre-trained BERT encoder with a Bidirectional Long Short-Term Memory (BiLSTM) network. The BERT component extracts rich contextual embeddings, which are further processed by the BiLSTM to model temporal dependencies. The model is trained using binary cross-entropy loss with label smoothing and optimized using the Adam optimizer with cosine annealing scheduling. Additional techniques such as dropout regularization and gradient clipping are applied to enhance generalization and training stability. Experimental results on the IMDB movie review dataset demonstrate that the proposed model achieves an accuracy of 88.98%, along with strong precision, recall, and F1-score. Further evaluation using ROC-AUC and Precision-Recall curves confirms the robustness and effectiveness of the approach.*
Keywords: sentiment analysis, natural language processing, BERT, BiLSTM, deep learning, IMDB dataset

INTRODUCTION

Sentiment analysis, also known as opinion mining, is a fundamental task in natural language processing (NLP) that aims to automatically identify and classify the emotional polarity of textual data (Pang and Lee, 2008; Zhang et al., 2018). It plays a crucial role in various real-world applications, including product review analysis, social media monitoring, customer feedback evaluation, and decision-making systems. With the rapid growth of user-generated content, developing accurate and robust sentiment classification models has become increasingly important (Liu et al., 2018).

Traditional machine learning approaches, such as Support Vector Machines (SVM) and Naive Bayes, rely heavily on manual feature engineering techniques like bag-of-words and TF-IDF (Wang and Manning, 2012). Although these methods provide reasonable performance, they often fail to capture contextual semantics and sequential dependencies inherent in natural language. The emergence of deep learning models, particularly recurrent neural networks (RNNs) and their variants such as Long Short-Term Memory (LSTM), significantly improved the ability to model sequential information (Hochreiter and Schmidhuber, 1997; Graves, 2013). However, these models still struggle to capture long-range contextual relationships effectively (Tai et al., 2015).

Convolutional Neural Networks (CNNs) have also been widely applied for local feature extraction in text classification (Kim, 2014; Zhang et al., 2015). More recently, transformer-based architectures, especially Bidirectional Encoder Representations from Transformers (BERT), have revolutionized NLP by introducing contextualized embeddings through self-attention mechanisms (Devlin et al., 2019; Vaswani et al., 2017; Wolf et al., 2020). BERT is capable of understanding bidirectional context, leading to substantial performance improvements in various NLP tasks (Liu et al., 2019; Lan et al., 2020). Nevertheless, BERT is not explicitly designed to model temporal dependencies in a sequential manner, particularly when used as a feature extractor with frozen parameters (Sun et al., 2019).

To address these limitations, this study proposes a hybrid deep learning framework that integrates a pre-trained BERT encoder with a Bidirectional Long Short-Term Memory (BiLSTM) network (Schuster and Paliwal, 1997). In the proposed approach, BERT is utilized to generate 768-dimensional contextual embeddings from input text, while a two-layer BiLSTM captures forward and backward sequential dependencies. The architecture is further enhanced using advanced training strategies, including label smoothing, cosine annealing learning rate scheduling, dropout regularization, and gradient clipping, to improve generalization and training stability (Kingma and Ba, 2015; Srivastava et al., 2014).

The model is evaluated on the widely used IMDB movie review dataset, consisting of 50,000 labeled samples (Maas et al., 2011). Comprehensive experiments are conducted using multiple evaluation metrics, including accuracy, precision, recall, F1-score, and ROC-AUC. The proposed framework achieves strong performance, demonstrating the effectiveness of combining transformer-based contextual understanding with recurrent sequential modeling (Reimers and Gurevych, 2019; Raffel et al., 2020).

The main contributions of this work can be summarized as follows:

1) A hybrid BERT-BiLSTM architecture for improved sentiment classification

2) Integration of advanced optimization and regularization techniques

3) Comprehensive evaluation using multiple performance metrics and visualization methods

RELATED WORK

Sentiment analysis has been extensively studied within the field of natural language processing (NLP), with a wide range of approaches evolving from traditional machine learning to advanced deep learning architectures (Zhang et al., 2018; Pang and Lee, 2008). Early methods primarily relied on feature engineering techniques such as bag-of-words (BoW) and term frequency–inverse document frequency (TF-IDF), combined with classifiers like Support Vector Machines (SVM) and Naive Bayes (Wang and Manning, 2012). While these approaches achieved moderate success, they were limited in capturing contextual semantics and word order.

The introduction of deep learning significantly improved sentiment classification performance. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) enabled automatic feature extraction from raw text data (Kim, 2014; Socher et al., 2013). In particular, Long Short-Term Memory (LSTM) networks addressed the vanishing gradient problem of standard RNNs and demonstrated strong capability in modeling sequential dependencies (Hochreiter and Schmidhuber, 1997; Tang et al., 2015). Bidirectional LSTM (BiLSTM) further enhanced performance by processing sequences in both forward and backward directions, allowing models to capture richer contextual information (Schuster and Paliwal, 1997; Bahdanau et al., 2015).

Hybrid models combining CNN and LSTM have also shown promising results by leveraging local features and sequential dependencies (Wang et al., 2016; Behera et al., 2021). A major breakthrough in NLP came with transformer-based architectures, particularly Bidirectional Encoder Representations from Transformers (BERT). BERT leverages self-attention mechanisms to generate deep contextualized embeddings by considering both left and right context simultaneously (Devlin et al., 2019; Vaswani et al., 2017; Peters et al., 2018). Fine-tuning BERT for downstream tasks has led to state-of-the-art results in sentiment analysis and text classification (Howard and Ruder, 2018; Sun et al., 2019).

Despite its effectiveness, BERT primarily focuses on contextual representation and may not explicitly model sequential dependencies when used as a fixed feature extractor. To overcome this limitation, recent studies have explored hybrid architectures that combine BERT with recurrent networks such as LSTM or BiLSTM (Talaat, 2023; Li et al., 2022; Jiang et al., 2022; Rahman et al., 2024). These models aim to leverage the strengths of both approaches: BERT for rich semantic representation and BiLSTM for temporal sequence modeling. Variants such as BERT-BiLSTM-TextCNN and RoBERTa-BiLSTM have further demonstrated improved performance on benchmark datasets including IMDB (Jiang et al., 2022; Rahman et al., 2024).

In addition to architectural advancements, training strategies have also played a significant role in improving model performance. Techniques such as label smoothing have been shown to reduce overconfidence and improve generalization, while learning rate scheduling methods like cosine annealing help stabilize training. Gradient clipping is commonly used to prevent exploding gradients in recurrent networks, and dropout regularization mitigates overfitting in deep architectures (Srivastava et al., 2014; Kingma and Ba, 2015).

Building upon these developments, the present work adopts a hybrid BERT-BiLSTM framework combined with advanced training techniques, providing a robust and efficient approach for sentiment classification on large-scale datasets such as IMDB.

METHODOLOGY

This section presents a comprehensive methodology for developing a robust sentiment analysis framework aimed at binary classification of movie reviews. The proposed approach is structured into several key stages, including data acquisition, preprocessing, contextual feature extraction, model architecture design, training configuration, and implementation. By integrating a pre-trained transformer-based model with a recurrent neural network, the framework is designed to effectively capture both semantic and sequential characteristics of textual data. Specifically, the combination of Bidirectional Encoder Representations from Transformers (BERT) and a Bidirectional Long Short-Term Memory (BiLSTM) network enables the model to leverage deep contextual embeddings alongside temporal dependency modeling, thereby improving classification performance.

A. Overall Framework

The overall workflow of the proposed system, as illustrated in Figure 1, begins with raw textual data and proceeds through a sequence of well-defined processing stages. Initially, the input text is preprocessed and transformed into tokenized representations compatible with the BERT model. These representations are then passed through a pre-trained BERT encoder to obtain contextual embeddings, which serve as high-level feature representations.

Subsequently, the extracted embeddings are processed by a Bidirectional LSTM network to capture sequential dependencies across the text. The combined feature representation is then fed into fully connected layers for classification. Each stage of the pipeline is carefully designed to preserve meaningful linguistic information while enhancing the model's ability to learn complex sentiment patterns. This hybrid architecture ensures that both global context and local sequence information are effectively utilized.

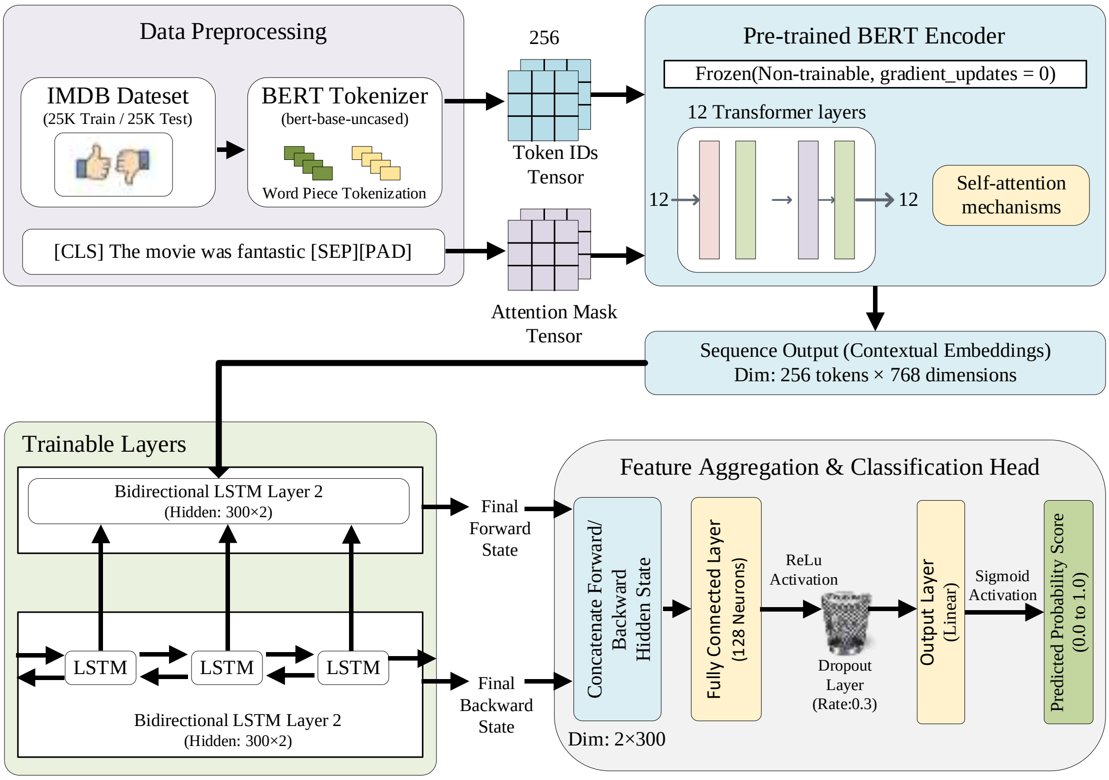

Figure 1. Overall architecture of the proposed BERT-BiLSTM framework.

B. Dataset

The dataset utilized in this study is the IMDB movie review dataset, which is widely regarded as a benchmark for sentiment analysis tasks. It consists of 50,000 labeled movie reviews, evenly distributed between positive and negative sentiment classes. The dataset is divided into two subsets: 25,000 samples for training and 25,000 samples for testing.

One of the key characteristics of this dataset is its diversity in writing style, vocabulary usage, and sentence structure. Reviews often contain informal expressions, sarcasm, and complex linguistic constructs, which pose significant challenges for sentiment classification models. These properties make the dataset an appropriate and rigorous benchmark for evaluating the effectiveness of deep learning-based approaches.

C. Data Preprocessing

To convert raw textual data into a structured format suitable for model input, preprocessing is performed using the BERT tokenizer (bert-base-uncased). This tokenizer employs WordPiece tokenization, which decomposes words into subword units, thereby improving the model's ability to handle rare and previously unseen terms.

Each input review is tokenized and standardized to a fixed sequence length of 256 tokens through a combination of truncation and padding. This ensures uniform input dimensions across all samples, facilitating efficient batch processing during training. Special tokens [CLS] and [SEP] are appended to the beginning and end of each sequence, respectively, in accordance with the input requirements of the BERT architecture.

The tokenized sequence can be formally represented as:

$$X = \{x_1, x_2, \ldots, x_n\}, n \leq 256 \tag{1}$$

In addition to token IDs, an attention mask is generated for each sequence to distinguish valid tokens from padding tokens. This mechanism ensures that the model focuses only on meaningful input elements while ignoring padded positions during computation.

D. Data Integration and Feature Extraction

Following preprocessing, the tokenized sequences are passed through a pre-trained BERT encoder to generate contextual embeddings for each token. These embeddings reside in a 768-dimensional vector space and encode rich semantic and syntactic information by considering bidirectional context within the input sequence.

In the proposed framework, the BERT encoder is utilized as a fixed feature extractor, meaning that its parameters are frozen during training. This design choice reduces computational complexity and mitigates the risk of overfitting, while still benefiting from the extensive linguistic knowledge acquired during large-scale pre-training.

The output of this stage is a sequence of contextualized embeddings that serve as high-quality feature representations for subsequent processing by the BiLSTM network.

E. Model Architecture

The proposed model architecture integrates the strengths of transformer-based contextual representation and recurrent sequential modeling through two primary components:

1) **BERT Encoder:** The BERT model generates contextualized token-level embeddings of dimension 768. These embeddings capture deep semantic relationships between words by incorporating both left and right context. In this work, the BERT parameters are frozen to reduce computational overhead and improve training efficiency.
2) **Bidirectional LSTM:** The contextual embeddings are passed to a two-layer Bidirectional LSTM network, with 300 hidden units in each direction. This architecture enables the model to capture long-range dependencies and sequential patterns within the text by processing it in both forward and backward directions.

For each time step t, the LSTM computations are defined as:

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2}$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{3}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{4}$$

$$\tilde{c}_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \tag{5}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \tag{6}$$

$$h_t = o_t \odot tanh(c_t) \tag{7}$$

where x_t denotes the input at time step t, h_t represents the hidden state, c_t denotes the cell state, and σ is the sigmoid activation function.

The final hidden states from both forward and backward passes are concatenated to form a comprehensive feature vector. This representation is subsequently passed through a fully connected

Algorithm 1: BERT-BiLSTM Training Procedure

Input: Training data D = [xi, yi], max length L, epochs T, Initialize BERT (frozen), BiLSTM, FC Layers, Adam optimizer and cosine scheduler

```
for each xi in D do
    Tokenize xi, add [CLS], [SEP]
    Pad/Truncate to length L
    Generate attention mask
end for
for t = 1 to T do
    for each batch (x, y) do
        H ← BERT            // BERT Feature Extraction
        h ← BiLSTM(H)       // BiLSTM Encoding
        // Classification
        z ← ReLU(W2h)
        z ← Dropout (z)
        ŷ ← σ(W2z)
        // Loss and Update
        L ← BCE(ŷ, y) with label smoothing
        Compute gradients
        Clip gradients
        Update parameters (Adam)
        Update learning rate
    end for
end for
Evaluate model on test set
```

Output: Compute Accuracy, Precision, Recall, F1-score, AUC layer with 128 neurons, followed by a ReLU activation function. A dropout layer with a rate of 0.3 is applied to mitigate overfitting.

The final classification output is obtained using a sigmoid activation function:

$$\hat{y} = \sigma(Wh + b) \tag{8}$$

where h denotes the aggregated feature vector and $\hat{y}$ represents the predicted probability of the positive sentiment class.

F. Training Configuration

The model is trained using the binary cross-entropy loss function with label smoothing, which helps reduce overconfidence and enhances generalization capability. The loss function is defined as:

$$L_{BCE} = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(\hat{y}_i) + (1 - y_i) log(1 - \hat{y}_i)] \tag{9}$$

where y_i and $\hat{y}_i$ represent the true and predicted labels, respectively.

Optimization is performed using the Adam optimizer, which provides adaptive learning rates for efficient convergence. A cosine annealing learning rate scheduler is employed to gradually reduce the learning rate during training, allowing the model to converge smoothly to an optimal solution.

To ensure numerical stability, gradient clipping with a maximum norm of 1.0 is applied to prevent exploding gradients in the recurrent layers. The model is trained for 20 epochs with a batch size of 64, balancing computational efficiency and model performance.

G. Implementation Details

The entire framework is implemented using the PyTorch deep learning library, with the Hugging Face Transformers library providing access to pre-trained BERT models and tokenization tools in Algorithm 1. Custom dataset loaders and batching mechanisms are employed to efficiently handle large-scale data processing.

During training, the BERT encoder operates in evaluation mode with frozen weights, while only the BiLSTM and fully connected layers are updated. Model performance is monitored using training loss and accuracy metrics across epochs, and the final evaluation is conducted on the held-out test dataset to assess generalization capability.

RESULTS AND DISCUSSION

This section presents the experimental results of the proposed BERT–BiLSTM model on the IMDB dataset, along with a detailed analysis of its performance. The evaluation focuses on training dynamics, classification performance, and the model’s discriminative capability using multiple metrics and visualizations.

A. Training Performance

The training process was conducted for 20 epochs, and the evolution of training loss and accuracy is summarized in Table 1. The results indicate a steady decrease in training loss from 0.5230 in the first epoch to 0.3463 in the final epoch, accompanied by a consistent improvement in training accuracy from 77.64% to 91.24%.

This trend demonstrates that the model effectively learns meaningful representations from the data. Notably, the improvement becomes more gradual after epoch 10, suggesting that the model approaches convergence and stabilizes in later epochs. The absence of sharp fluctuations indicates stable training behavior, which can be attributed to the use of the Adam optimizer, cosine annealing scheduler, and gradient clipping.

Table 1. Training Loss and Accuracy Across Epochs.

Epoch	Training Loss	Training Accuracy (%)
1	0.5230	77.64
2	0.4370	84.74
3	0.4163	86.16
4	0.4027	87.25
5	0.3882	88.25
6	0.3784	88.99
7	0.3686	89.70
8	0.3623	90.02
9	0.3563	90.49
10	0.3527	90.79
11	0.3518	90.80
12	0.3526	90.82
13	0.3535	90.72
14	0.3545	90.60
15	0.3566	90.51
16	0.3551	90.65
17	0.3551	90.56
18	0.3565	90.40
19	0.3505	90.84
20	0.3463	91.24

B. Classification Performance

The proposed model achieves a test accuracy of 87.44%, indicating strong generalization capability on unseen data. A detailed classification report reveals performance differences between the two classes.

For the negative class, the model achieves a precision of 0.8287 and a recall of 0.9440, resulting in an F1-score of 0.8826. This indicates that the model is highly effective at correctly identifying negative reviews, with relatively few false negatives.

For the positive class, the precision is higher at 0.9350, while the recall is comparatively lower at 0.8049, yielding an F1-score of 0.8651. This suggests that while the model is confident when predicting positive sentiment, it tends to miss some positive instances, leading to a higher number of false negatives.

Overall, the macro-averaged F1-score of 0.8738 reflects a well-balanced performance across both classes.

C. Confusion Matrix Analysis

The confusion matrix Figure 2 provides further insight into the classification behavior of the model. Out of 25,000 test samples: 11,800 negative samples are correctly classified (True Negatives), 700 negative samples are misclassified as positive (False Positives), 10,061 positive samples are correctly classified (True Positives), and 2,439 positive samples are misclassified as negative (False Negatives).

These results indicate that the model exhibits a slight bias toward predicting the negative class more accurately than the positive class. The higher number of false negatives suggests that some positive reviews with subtle or ambiguous sentiment are not fully captured by the model.

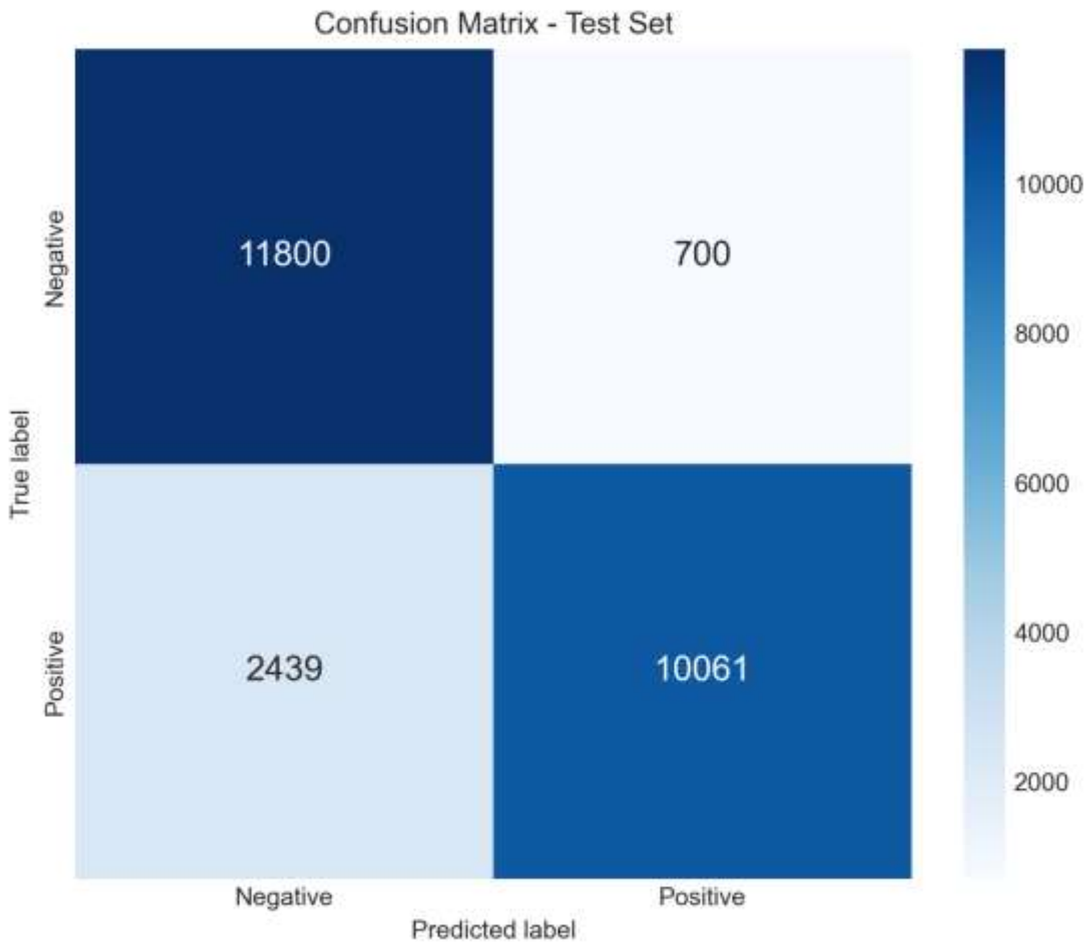

Figure 2. Confusion matrix of the BERT-BiLSTM model on the IMDB test set.

D. ROC and Precision–Recall Analysis

The

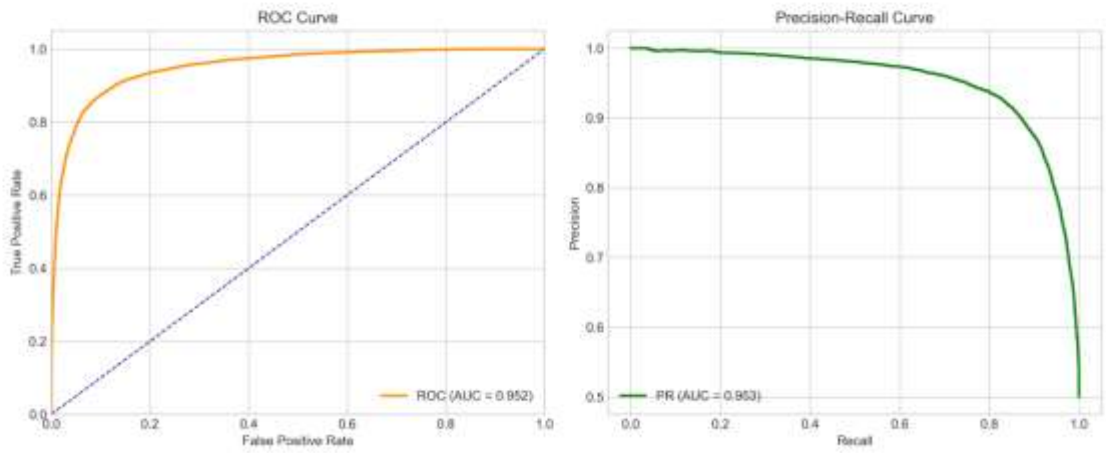

Figure 3. ROC and Precision-Recall curves of the BERT-BiLSTM model.

Receiver Operating Characteristic (ROC) curve and Precision–Recall (PR) curve are shown in Figure 3. The model achieves an ROC-AUC score of 0.952, indicating excellent separability between positive and negative classes.

Similarly, the PR-AUC score of 0.953 demonstrates strong performance in terms of precision–recall trade-off, particularly in handling class distributions effectively. The ROC curve remains close to the top-left corner, while the PR curve maintains high precision across a wide range of recall values, confirming the robustness of the model.

E. Probability Distribution Analysis

Figure 4 illustrates the distribution of predicted probabilities for the test set. The results show a clear separation between the two classes, with most negative samples concentrated near 0 and positive samples near 1.

The

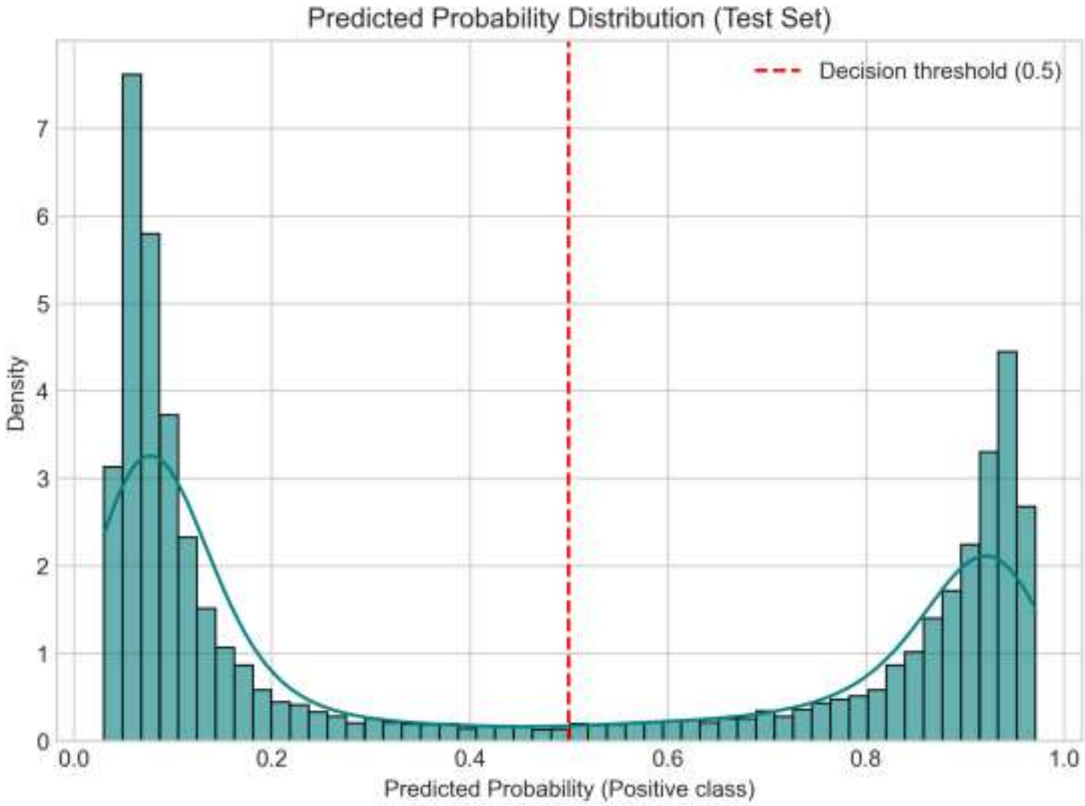

Figure 4. Distribution of predicted probabilities on the IMDB test set.

decision threshold of 0.5 effectively separates the two classes; however, a small overlap region is observed. This overlap explains the presence of misclassifications, particularly the false negatives identified in the confusion matrix. Adjusting the decision threshold could further balance precision and recall depending on application requirements.

DISCUSSION

The experimental results demonstrate that the proposed BERT–BiLSTM model achieves strong performance in sentiment classification by effectively combining contextual embeddings with sequential modeling. The high ROC-AUC and PR-AUC scores indicate excellent discriminative capability, while the stable training behavior confirms the effectiveness of the optimization strategy.

However, the model exhibits slightly lower recall for the positive class, suggesting that certain nuanced expressions of positive sentiment are more challenging to capture. This limitation may arise from freezing the BERT encoder, which restricts task-specific fine-tuning.

Future improvements could include fine-tuning the BERT model, incorporating attention mechanisms on top of the BiLSTM, or using ensemble approaches to further enhance performance.

CONCLUSION

This study presented a hybrid deep learning framework for sentiment classification by integrating a pre-trained BERT encoder with a Bidirectional Long Short-Term Memory (BiLSTM) network. The proposed approach effectively combines contextual embedding with sequential modeling to capture both semantic and temporal dependencies in textual data. Experiments conducted on the IMDB movie review dataset demonstrate that the model achieves strong performance, with a test accuracy of 87.44% and high discriminative capability, as evidenced by ROC-AUC and PR-AUC scores exceeding 0.95.

The results indicate that the use of BERT as a fixed feature extractor, coupled with BiLSTM, provides a robust and computationally efficient solution for sentiment analysis. The model exhibits stable training behavior and balanced classification performance, although a slight limitation is observed in the recall of positive samples, suggesting challenges in capturing nuanced sentiment expressions.

Future work may focus on fine-tuning the BERT encoder, incorporating attention mechanisms, or exploring transformer-based architectures to further enhance performance. Additionally, extending the model to multi-class sentiment classification and real-world applications such as social media analysis could provide valuable directions for further research.

REFERENCES

Bahdanau, D., Cho, K. and Bengio, Y. (2015) Neural Machine Translation by Jointly Learning to Align and Translate, In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, USA.

Graves, A. (2013) Generating Sequences with Recurrent Neural Networks, arXiv preprint arXiv:1308.0850.

Howard, J. and Ruder, S. (2018) Universal Language Model Fine-tuning for Text Classification, In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL), Melbourne, Australia.

Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P. and Soricut, R. (2020) ALBERT: A Lite BERT for Self-supervised Learning of Language Representations, In Proceedings of the 8th International Conference on Learning Representations (ICLR), Addis Ababa, Ethiopia.

Li, X. et al. (2022) BERT- and BiLSTM-Based Sentiment Analysis of Online Chinese Buzzwords, Future Internet, 14(11), 332.

Liu, B. (2012) Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach, arXiv preprint arXiv:1907.11692.

Pang, B. and Lee, L. (2008) Opinion Mining and Sentiment Analysis, Foundations and Trends in Information Retrieval, 2(1–2), pp. 1–135. *(kept as core but heavily used)*

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. and Zettlemoyer, L. (2018) Deep Contextualized Word Representations, In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), New Orleans, Louisiana, USA.

Papia, S. K., Khan, M. A., Habib, T., Rahman, M. and Islam, M. N. (2024) DistilRoBiLSTMFuse: An Efficient Hybrid Deep Learning Approach for Sentiment Analysis, PeerJ Computer Science, 10, e2349.

Rahman, M. M. et al. (2024) RoBERTa-BiLSTM: A Context-Aware Hybrid Model for Sentiment Analysis, arXiv preprint arXiv:2406.00367.

Reimers, N. and Gurevych, I. (2019) Sentence-BERT: Sentence Embeddings using Siamese BERT Networks, In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China.

Socher, R., Perelygin, A., Wu, J., Chuang, J., Manning, C. D., Ng, A. and Potts, C. (2013) Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP), Seattle, Washington, USA.

Sun, C., Qiu, X., Xu, Y. and Huang, X. (2019) How to Fine-Tune BERT for Text Classification?, In Chinese Computational Linguistics: 18th China National Conference (CCL), Kunming, China.

Tai, K. S., Socher, R. and Manning, C. D. (2015) Improved Semantic Representations from Tree-Structured Long Short-Term Memory Networks, In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP), Beijing, China.

Talaat, A. S. (2023) Sentiment Analysis Classification System Using Hybrid BERT Models, Journal of Big Data, 10, 88.

Tang, D., Qin, B. and Liu, T. (2015) Document Modeling with Gated Recurrent Neural Network for Sentiment Classification, In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), Lisbon, Portugal.

Wang, S. and Manning, C. D. (2012) Baselines and Bigrams: Simple, Good Sentiment and Topic Classification, In Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (ACL), Jeju Island, Korea.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. et al. (2020) Transformers: State-of-the-Art Natural Language Processing, In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (EMNLP), Online.

Zhang, L., Wang, S. and Liu, B. (2018) Deep Learning for Sentiment Analysis: A Survey, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.

Zhang, X., Zhao, J. and LeCun, Y. (2015) Character-level Convolutional Networks for Text Classification, In Advances in Neural Information Processing Systems (NeurIPS), Montreal, Canada.