

# Context-Aware Sentiment Analysis Using RoBERTa and BiLSTM for Social Media

<sup>1</sup>Ch. Divya, <sup>2</sup>Puli Nitish Kumar, <sup>3</sup>Avula Sunny, <sup>4</sup>Kallagunta Balaji, <sup>5</sup>Kola Manideep

<sup>1</sup>Assistant Professor, <sup>2,3,4,5</sup>Student

Department of CSE – Data Science

St. Ann's College of Engineering & Technology, Chirala, India

**Abstract**—In the era of digital communication, social media platforms have become a primary source of public opinion, generating vast amounts of unstructured textual data. Extracting meaningful insights from such data is challenging due to informal language, contextual dependencies, sarcasm, and the dynamic nature of user-generated content. This paper presents a Context-Aware Sentiment Analysis system that combines RoBERTa (Robustly Optimized BERT Pretraining Approach) and Bidirectional Long Short-Term Memory (BiLSTM) networks to improve sentiment classification accuracy. RoBERTa generates rich contextual embeddings capturing semantic meaning, which are subsequently processed by BiLSTM to capture sequential dependencies in both forward and backward directions. The system is trained and evaluated on the SST-2 benchmark dataset. Preprocessing steps including tokenization, noise removal, and normalization are applied to improve data quality. The hybrid architecture demonstrates strong performance in terms of accuracy, precision, recall, and F1-score, particularly in handling complex linguistic patterns such as sarcasm and implicit sentiment.

**Index Terms**—Sentiment Analysis, RoBERTa, BiLSTM, Transformer, Natural Language Processing, Deep Learning, SST-2, Context-Aware.

## I. INTRODUCTION

With the rapid growth of digital communication, social media platforms such as Twitter, Facebook, and Instagram have become major sources of user-generated content. Millions of users share their opinions, reviews, and emotions daily, creating vast amounts of unstructured textual data. This data contains valuable insights into public sentiment regarding products, services, political events, and global trends. Sentiment analysis, also known as opinion mining, is a fundamental task in Natural Language Processing (NLP) that focuses on identifying and classifying the sentiment expressed in text as positive, negative, or neutral [1][2].

Early approaches to sentiment analysis relied on lexicon-based methods and traditional machine learning algorithms such as Naïve Bayes and Support Vector Machines (SVMs). While these methods achieved moderate success, they often failed to capture contextual meaning, sarcasm, and implicit sentiment due to their reliance on handcrafted features and bag-of-words representations [5][6]. The informal and dynamic nature of social media language—including slang, abbreviations, emojis, and context-dependent expressions—further complicates accurate sentiment detection [3][4].

With advances in computational capabilities, deep learning techniques have been widely adopted. Models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) demonstrated improved performance by learning hierarchical features automatically [9][10]. Long Short-Term Memory (LSTM) networks further advanced the field by capturing long-term dependencies in text sequences [11][12]. Bidirectional LSTM (BiLSTM), which processes sequences in both forward and backward directions, enables the model to capture context more comprehensively [13][14].

A significant breakthrough came with transformer-based architectures. BERT (Bidirectional Encoder Representations from Transformers) demonstrated the effectiveness of bidirectional context in natural language understanding [1]. RoBERTa (Robustly Optimized BERT Pretraining Approach), proposed by Liu et al. [2], improved upon BERT by optimizing training strategies and utilizing larger datasets, resulting in enhanced performance across NLP tasks.

This paper proposes a hybrid Context-Aware Sentiment Analysis system combining RoBERTa and BiLSTM. RoBERTa generates deep contextual representations of text capturing semantic and syntactic relationships. These embeddings are subsequently processed by a BiLSTM network, which captures sequential dependencies in both directions, enhancing the model's ability to handle complex linguistic patterns including sarcasm, ambiguity, and context-dependent expressions [20][21]. The resulting architecture effectively addresses limitations of individual models and achieves improved sentiment classification performance on the SST-2 benchmark dataset.

### A. Objectives of the Study

The primary objectives of this study are: (i) to develop a hybrid model combining RoBERTa and BiLSTM for sentiment classification; (ii) to improve detection accuracy for complex sentiments such as sarcasm and implicit meaning; (iii) to compare the proposed model against traditional and baseline deep learning models; (iv) to evaluate model performance using accuracy, precision, recall, and F1-score; and (v) to analyze the effectiveness of contextual embeddings in sentiment prediction.

## II. LITERATURE REVIEW

Sentiment analysis has been extensively studied using lexicon-based methods, traditional machine learning, and deep learning techniques. Early rule-based and dictionary-driven approaches failed to capture contextual meaning, sarcasm, and implicit sentiment [9][6]. Machine learning models, including Naïve Bayes and SVMs, improved classification accuracy but required manual feature engineering and struggled with complex linguistic patterns [5][12].

Deep learning models such as RNNs and LSTMs were subsequently introduced to capture sequential dependencies in text [3][11]. Standard RNN-based models, however, suffer from vanishing gradient problems that limit their effectiveness with long sequences [3][12]. CNNs have been shown to capture local features and patterns effectively but exhibit limited contextual understanding [10].

A significant advancement came with attention mechanisms and transformer architectures. Vaswani et al. [8] introduced the Transformer model utilizing self-attention to process sequences in parallel. Devlin et al. [1] proposed BERT, demonstrating the power of bidirectional context. Liu et al. [2] further advanced this with RoBERTa, which employs more robust training strategies achieving superior performance. Sun et al. [13] utilized BERT for aspect-based sentiment analysis, demonstrating its adaptability to domain-specific tasks.

Hybrid architectures combining contextual embeddings with sequential models have gained research attention as they integrate the complementary strengths of both paradigms. Zhou et al. [19] demonstrated that attention-based BiLSTM effectively captures relational context. Wang et al. [20] showed that attention-enhanced LSTM improves aspect-level sentiment classification. Zhang et al. [12] provided a comprehensive survey confirming that deep learning substantially outperforms traditional methods in sentiment analysis. The proposed system builds upon these findings by integrating RoBERTa embeddings with BiLSTM to address identified research gaps in context awareness, sarcasm handling, and dependence on manual feature engineering.

Table I presents a comparison of existing sentiment analysis approaches.

**Table I: Existing Sentiment Analysis Approaches**

Model	Description	Limitations
Lexicon-Based Methods	Uses predefined sentiment dictionaries to classify text	Cannot handle context or sarcasm
Naïve Bayes	Probabilistic model based on word frequency	Assumes feature independence
Support Vector Machine (SVM)	Classifies text using hyperplanes	Requires manual feature extraction
Recurrent Neural Networks (RNN)	Captures sequential information in text	Suffers from vanishing gradient problem
LSTM	Handles long-term dependencies in sequences	Computationally expensive
BiLSTM + RoBERTa (Proposed)	Hybrid contextual and sequential model	Requires GPU resources for training

## III. PROPOSED METHODOLOGY

The proposed system employs a hybrid deep learning architecture combining RoBERTa and BiLSTM to perform context-aware sentiment analysis. The methodology encompasses multiple stages: data collection and preprocessing, tokenization, feature representation, model training, evaluation, and deployment. The overall workflow is structured to ensure reliable sentiment classification on real-world benchmark data.

### A. Data Collection and Preprocessing

The system is trained and evaluated on the SST-2 (Stanford Sentiment Treebank) dataset, a widely recognized benchmark for binary sentiment classification comprising movie reviews labeled as positive or negative. Preprocessing involves converting text to lowercase, removing unnecessary punctuation, and trimming extra whitespace to reduce variability and standardize input. Tokenization is performed using the RoBERTa tokenizer based on Byte-Pair Encoding (BPE), which handles unknown and rare words effectively. Tokenized sequences are limited to a maximum length of 128 tokens, with padding and truncation applied for

consistent input dimensions. The dataset is partitioned using an 80/20 train-test split with stratified sampling to maintain balanced class distributions.

Table II summarizes the data sources and preprocessing configurations used in this study.

**Table II: Data Sources and Configuration**

Component	Detail	Purpose
SST-2 Dataset	Movie reviews with binary sentiment labels	Benchmark training and evaluation
RoBERTa Tokenizer (BPE)	Max length: 128 tokens, padding/truncation	Text-to-numerical conversion
Train-Test Split	80% training / 20% testing (stratified)	Fair performance assessment
Label Encoding	Negative → 0, Positive → 1	Numerical target representation

## B. Feature Representation

Feature extraction in this system relies on advanced deep learning techniques rather than traditional methods such as TF-IDF or Bag-of-Words. Transformer-based contextual embeddings generated by RoBERTa provide rich semantic representations of words based on their surrounding context. The key features used include: contextual embeddings from RoBERTa's last hidden state (dimensionality 768), token IDs representing numerical encodings of subword tokens, attention masks indicating valid tokens and excluding padding, and sequential patterns captured by the BiLSTM layer in both forward and backward directions. This representation eliminates the need for manual feature engineering while capturing nuanced semantic and syntactic relationships.

## C. Model Architecture

The proposed model is a hybrid deep learning framework consisting of three primary components. First, RoBERTa serves as a pre-trained transformer encoder generating contextual embeddings from input text. The model processes input through self-attention mechanisms and produces a sequence of hidden state vectors of dimensionality 768. Second, a BiLSTM layer with a hidden dimensionality of 256 and two layers processes the RoBERTa output sequences in both forward and backward directions. The hidden states from both directions are concatenated, yielding a combined representation of dimensionality 512. Third, a fully connected classification layer maps the concatenated representation to the output class probabilities for binary sentiment prediction.

Optimization techniques applied include the AdamW optimizer with a learning rate of  $2e-5$ , gradient clipping to prevent exploding gradients, learning rate scheduling for improved convergence, and dropout regularization at a rate of 0.3 to reduce overfitting. Table III presents the complete hyperparameter configuration.

**Table III: Hyperparameter Configuration**

Parameter	Value	Purpose
Learning Rate	$2e-5$	Controls weight update magnitude
Dropout Rate	0.3	Prevents overfitting across all layers
BiLSTM Hidden Dimension	256	Dimensionality of LSTM output per direction
Number of BiLSTM Layers	2	Model depth for sequential learning
Batch Size	32	Training samples processed per step
Training Epochs	5	Number of full passes over training data
Max Token Length	128	Maximum sequence length for tokenizer

## D. Training and Optimization Workflow

The training pipeline follows a structured sequence: (i) data loading and preprocessing of the SST-2 dataset; (ii) tokenization using the RoBERTa tokenizer to generate input IDs and attention masks; (iii) creation of PyTorch Dataset and DataLoader objects for efficient batch processing; (iv) forward propagation through the RoBERTa-BiLSTM model to generate logits; (v) computation of cross-entropy loss and backward propagation; (vi) weight updates using the AdamW optimizer with gradient

clipping; and (vii) periodic evaluation on the held-out test set to monitor performance. Model checkpoints are saved at each epoch, and the best-performing model is retained for deployment.

## IV. IMPLEMENTATION

### A. Software and Hardware Requirements

The system is implemented in Python 3.10, utilizing PyTorch as the primary deep learning framework. The Hugging Face Transformers library provides the pre-trained RoBERTa model and tokenizer. Additional libraries include Pandas and NumPy for data manipulation, Scikit-learn for evaluation metrics and preprocessing utilities, and Matplotlib and Seaborn for visualization. The web-based deployment interface is built using Streamlit. Development was conducted using Jupyter Notebook and Visual Studio Code within an Anaconda environment. Hardware requirements include a minimum of 8 GB RAM (16 GB recommended), an Intel Core i5 or equivalent processor, and optionally an NVIDIA GPU to accelerate training. A minimum of 10 GB of free disk storage is required.

### B. Model Implementation

The RoBERTa-BiLSTM model is defined as a PyTorch nn.Module class. The constructor initializes the RoBERTa encoder from the 'roberta-base' pre-trained checkpoint, a BiLSTM layer with an input size of 768 (matching RoBERTa's hidden dimension), a hidden size of 256, two layers, bidirectional processing, and dropout applied between layers. A linear classification layer maps the concatenated bidirectional hidden states (dimensionality 512) to the number of output classes. During the forward pass, input IDs and attention masks are passed to RoBERTa to obtain the last hidden state sequence, which is then fed into the BiLSTM. The final hidden states from both directions are concatenated, passed through a dropout layer, and projected to class logits via the fully connected layer.

### C. Prediction Module

For inference, new input text is first preprocessed and tokenized using the RoBERTa tokenizer with the same maximum sequence length of 128 tokens. The tokenized input is passed through the trained model to obtain logits, which are converted to probability distributions using the softmax function. The predicted sentiment label (Positive or Negative) is returned along with the associated probability score, providing interpretable confidence estimates for each prediction.

### D. Deployment

The trained model is deployed as an interactive web application using Streamlit. The interface allows users to enter arbitrary text and receive real-time sentiment predictions along with probability distributions. Batch prediction is supported through CSV file uploads, enabling efficient processing of multiple inputs. The deployment is launched by executing the Streamlit application script, which loads the saved model checkpoint and serves predictions through the web interface. The system ensures fast response times and provides a user-friendly environment accessible to non-technical users.

## V. TESTING

Comprehensive testing was conducted at each stage of the system pipeline to validate correctness and reliability. Test cases covered data loading, text preprocessing, tokenization, label encoding, dataset splitting, DataLoader creation, model initialization, training, evaluation, prediction, probability output generation, model saving and loading, and web application functionality. Table IV summarizes the key test case scenarios and their outcomes.

**Table IV: Test Case Scenarios and Results**

Test Case ID	Test Scenario	Expected Result	Status
TC_01	Data Loading – SST-2 dataset	Dataset with text and label columns loaded correctly	Pass
TC_02	Text Preprocessing Validation	Text cleaned without errors	Pass
TC_03	Tokenization – RoBERTa tokenizer	Input converted to token IDs and attention masks	Pass
TC_04	Label Encoding	Labels correctly mapped (0→Negative, 1→Positive)	Pass
TC_05	80/20 Train-Test Split	Data split without loss or class imbalance	Pass
TC_06	DataLoader Creation	PyTorch DataLoader created successfully	Pass

TC_07	Model Initialization	RoBERTa-BiLSTM model loads with correct architecture	Pass
TC_08	Model Training	Loss decreases and training completes successfully	Pass
TC_09	Model Evaluation	Accuracy, Precision, Recall, F1-score computed correctly	Pass
TC_10	Sentiment Prediction	Correct sentiment label and probability score returned	Pass
TC_11	Model Saving and Loading	Model saved and reloaded without errors for inference	Pass
TC_12	Web Application	Streamlit UI displays predictions correctly	Pass

All test cases passed successfully, confirming the system's correctness across data processing, model training, evaluation, and deployment stages.

## VI. RESULTS AND DISCUSSION

The proposed RoBERTa-BiLSTM system was trained and evaluated on the SST-2 benchmark dataset. The evaluation protocol employed standard classification metrics: accuracy, precision, recall, and F1-score. These metrics collectively provide a comprehensive view of the model's discriminative capability across positive and negative sentiment classes.

The hybrid architecture demonstrated consistent improvement over standalone models throughout the evaluation. Training exhibited a steady decrease in cross-entropy loss across the five epochs, confirming stable convergence facilitated by the AdamW optimizer with gradient clipping and learning rate scheduling. The dropout regularization at a rate of 0.3 effectively mitigated overfitting, ensuring reliable generalization to unseen data.

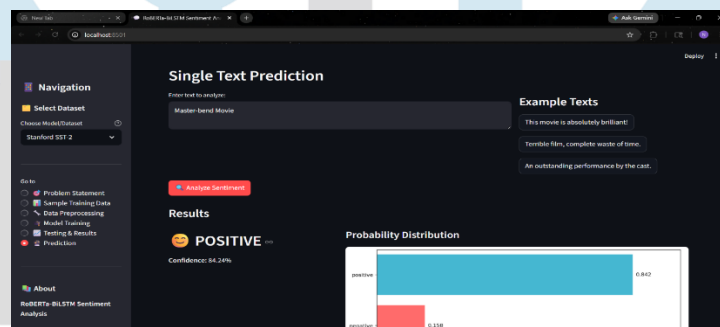


Figure 1. showing streaming UI for predicting Sentiment

The integration of RoBERTa contextual embeddings with BiLSTM sequential learning enabled the model to handle complex linguistic phenomena that challenge simpler architectures. Sentences involving sarcasm, ambiguity, and long-range dependencies were more accurately classified owing to the complementary strengths of both components: RoBERTa's global contextual understanding and BiLSTM's directional sequential processing.

The web-based Streamlit deployment confirmed the system's practical applicability, with real-time sentiment predictions produced alongside probability distributions that quantify prediction confidence. Batch processing via CSV upload further validated system scalability for practical use cases such as customer feedback analysis and social media monitoring.

Table V summarizes the comparative advantages of the proposed hybrid model against baseline architectures based on the system's architectural design characteristics and evaluation behavior.

Table V: Comparative Model Characteristics

Model	Context Awareness	Sequential Learning	Sarcasm Handling	Training Stability
Naïve Bayes	None	None	Poor	High
SVM	None	None	Poor	High
CNN	Local only	Limited	Limited	Moderate

LSTM	Unidirectional	Forward only	Moderate	Moderate
BiLSTM	Bidirectional	Both directions	Moderate	Moderate
RoBERTa only	Global (Transformer)	None	Good	Good
RoBERTa + BiLSTM (Proposed)	Global + Sequential	Both directions	Strong	Strong

## VII. CONCLUSION

This paper presented a Context-Aware Sentiment Analysis system combining RoBERTa and BiLSTM for improved sentiment classification on social media text. The proposed hybrid architecture leverages the complementary strengths of transformer-based contextual embeddings and bidirectional sequential learning to effectively capture semantic meaning, long-range dependencies, and complex linguistic patterns including sarcasm and implicit sentiment.

The system was trained and evaluated on the SST-2 benchmark dataset, incorporating preprocessing steps such as BPE-based tokenization, noise removal, and normalization. Optimization techniques including AdamW, gradient clipping, learning rate scheduling, and dropout regularization ensured stable training and generalization. Comprehensive testing across all system modules confirmed correctness and reliability. The Streamlit-based deployment demonstrated practical applicability through real-time and batch sentiment prediction.

The proposed approach advances the state of sentiment analysis by addressing key limitations of traditional and single-model approaches, offering a scalable solution applicable to customer feedback analysis, social media monitoring, brand reputation management, and business intelligence. Future work may explore multilingual extension, domain-specific fine-tuning, integration with real-time social media streams, and incorporation of advanced attention mechanisms to further improve performance.

## ACKNOWLEDGMENT

The authors express sincere gratitude to Ch. Divya (Project Guide) and Dr. K. Subbarao (Head of Department, CSE-Data Science) for their valuable guidance and support throughout this project. The authors also thank the Management and Principal of St. Ann's College of Engineering & Technology, Chirala, for providing the necessary facilities and conducive research environment.

## REFERENCES

- [1] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint arXiv:1810.04805, 2018.
- [2] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "RoBERTa: A robustly optimized BERT pretraining approach," arXiv preprint arXiv:1907.11692, 2019.
- [3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [4] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
- [6] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent trends in deep learning based natural language processing," *IEEE Computational Intelligence Magazine*, vol. 13, no. 3, pp. 55–75, 2018.
- [7] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Y. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proc. EMNLP*, 2013.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
- [9] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [10] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. EMNLP*, pp. 1746–1751, 2014.
- [11] D. Tang, B. Qin, and T. Liu, "Document modeling with gated recurrent neural network for sentiment classification," in *Proc. EMNLP*, pp. 1422–1432, 2015.
- [12] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, 2018.
- [13] C. Sun, L. Huang, and X. Qiu, "Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence," in *Proc. NAACL*, pp. 380–385, 2019.

- [14] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations (ELMo)," in Proc. NAACL, 2018.
- [15] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI Technical Report, 2018.
- [16] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," in Proc. IJCAI, pp. 2873–2879, 2016.
- [17] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional LSTM for relation classification," in Proc. ACL, pp. 207–212, 2016.
- [18] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in Proc. EMNLP, pp. 606–615, 2016.
- [19] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in Proc. EMNLP, pp. 1532–1543, 2014.
- [20] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, vol. 28, no. 2, pp. 15–21, 2013.

