NABEA AL-AJYAL FOUNDATION PRESS
Online ISSN: 3078-5413
CyberSystem Journal (CSJ)
csj.nabea.pub
CyberSystem Journal Original Article

A Robust and Efficient Attention-CNN-BiGRU Framework for Sentiment Analysis in Noisy Social Media Environments

Rahman Nahi Abid [1] *

[1] Nursing Faculty, University of Babylon, Babylon, Iraq

***** Corresponding Author: **Rahman Nahi Abid** Email: rahmannahi@uobabylon.edu.iq

Abstract: The proliferation of social media platforms, particularly X, has revolutionized public communication but poses challenges for sentiment analysis due to informal language characterized by abbreviations, slang, emojis, and creative spellings. This study presents an Attention-Enhanced Convolutional Neural Network–Bidirectional Gated Recurrent Unit (CNN-BiGRU) model, a hybrid architecture designed for both accuracy and computational efficiency. It integrates CNNs for local feature extraction, BiGRUs for capturing sequential dependencies, and an attention mechanism to prioritize sentiment-critical components. Evaluated on Sentiment140 (1.6 million tweets, binary classification) and SemEval-2017 Task 4 (65,000 tweets, three-class classification), the model demonstrates statistically significant accuracy improvements (up to 1.7 percentage points) compared to baseline models, including BERT Base, while requiring substantially fewer parameters and offering potential for faster inference. The model achieves 86.3% accuracy on Sentiment140 and 77.2% on SemEval-2017 Task 4, effectively handling linguistic variability. Its comparatively lightweight design makes it suitable for real-time sentiment analysis scenarios where computational resources may be constrained. This framework offers a robust and scalable solution for analyzing sentiment in evolving online communication, with applications in market research, public opinion monitoring, and crisis management.

Access this article online

Keywords: Social Media Sentiment Analysis, Attention Mechanism, CNN-BiGRU Model, Informal Language Processing, Public Opinion Analysis, Computational Efficiency, Deep Learning.

1. Introduction

Platforms like X (previously Twitter) serve as key forums for sharing views, feelings, and responses to events (social, political, commercial, public health, etc.). The user generated content that results can serve as sentiment evidence for market research, public opinion monitoring, crisis management and social analytics [1-4]. Sentiment analysis on social media is difficult due to short, unstructured, fast-evolving and noisy language of posts. Common elements include slang, abbreviations, emojis, hashtags, user mentions, creative spelling, and irregular grammar [5,6].Prior techniques for sentiment analysis include lexicon-based models and more standard machine learning classifiers, e.g. Support Vector Machines,

Received April 9, 2026; Revised May 11, 2026; Accepted May 26, 2026; Published June 11, 2026
https://doi.org/10.57238/csj.2026.1024
© 2026 by the authors. licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

Naive Bayes. They depend on features which are mostly manually designed and fail to capture contextual or compositional meaning of short informal posts [7].

In contrast to other machine learning models, deep learning models rely less on dependence on feature engineering since the learning representations are created from the data itself. CNNs efficiently capture sentiment-bearing n-grams at the local level; recurrent models like LSTMs and GRUs can model sequential dependencies [8,9]. The performance is also improved using attention mechanisms that let the model focus on the most informative word or phrase in a sentence [10].

Transformers were state-of-the-art SOTA for NLP in late 2020, with BERT, RoBERTa and mini-models (DistilBERT, ALBERT, MiniLM). Nonetheless, the computational cost, memory footprint, and inference latency of a real-time or resource-constrained application of a social media monitoring system may be prohibitive. Researchers are exploring lightweight hybrid neural architectures that perform the task of sentiment classification in a short computation time.

We propose Attention-CNN-BiGRU model to do sentiment analysis of noisy data from social media. The model makes use of CNNs to capture local lexical features, BiGRUs to take into account a bidirectional context, and attention to learn important sentiment tokens. In the revised manuscript, a better description of data splitting, number of independent runs, hyper-parameter tuning, statistical significance tests, and scope of comparisons with Transformer-based models is given based on reviewer's comments.

The results of this mission are as follows:

- The architecture designed is Hybrid Attention-CNN-BiGRU. Robust sentiment classification of short noisy data from social media is facilitated by this.
- We present an experimental protocol that enables reproducibility, which consists of stratified train/validation/test splits, five independent runs, validation-based hyperparameter tuning and paired significance tests.
- The model was tested on two benchmark datasets, Sentiment140 and SemEval-2017 Task 4, for two-class and three-class sentiment classification.
- The assessment of parameter count and efficiency of inference time showcases in what manner we can deploy models optimally rather than standard Transformer baselines.
- The findings from ablation and error analyses shed light on the impact of each component on the proposed approach.

The other sections of this paper are organized as follows. Section 2 presents a literature review. Section 3 summarizes the proposed methodology. The datasets are described in 4. Section 5 is the Experiment and Result. Section 6 addresses the findings, limitations, delimitations, and implications. Section 7 concludes the paper with future work.

2. Related Work

Sentiment analysis moved on from lexicon-based and rule-based approaches to machine learning and deep learning ones. Lexicon-based methods attempt to use sentiment dictionary along with polarity rules. They cause difficulty due to context, negation, sarcasm, slang and domain-specific meaning [19]. Several traditional machine learning algorithms [20], namely SVM, Naive Bayes, Random Forest, and K-nearest neighbor, have been applied using Bag-of-Words, TF-IDF and hand-crafted features. Although effective in some situations, these models often require careful feature engineering and may not generalize well to informal social media language.The semantic representation of words has been greatly improved through mapping the words into continuous vector spaces with the help of distributed word representations like Word2Vec and GloVe [21,22].

The embeddings provide helpful information for semantic similarity and are often used as input to neural sentiment classifiers. Subsequently, classification at the sentence level has become an attractive option for CNNs as they capture discriminative local patterns and short phrase features [8]. Using recurrent models such as LSTMs and GRUs, word order and sequential dependency can be modeled. Bidirectional LSTMs and GRUs are also used to enhance contextual representation [23,24].

Neural models with attention mechanisms can assign higher weights to more informative tokens for the same performance, more recently in many sequence modeling tasks [26-28]. Attention is used in sentiment analysis to find sentiment words and negations, intensifiers, and emoticons. Hybrid models that combine CNNs and RNNs exploit their complementarity:

- CNNs capture local n-gram features
- recurrent layers model sequential context.

As a result, many studies have investigated the short-text CNN-LSTM [11,29,30]. The deep contextual self-attention for transformer-based models like BERT and RoBERTa has already established a strong baseline in many

NLP problems. The size of BERT is reduced through knowledge distillation in DistilBERT, through parameter-reduction strategies in ALBERT and MiniLM compresses transformer models through self-attention distillation [18,33,34]. These models are robust but require a lot of resources, limiting social media analysis in real-time. Thus, we do not claim that our hybrid architecture is better than every Transformer-based model. Instead, we check whether a well-designed hybrid architecture offers a good accuracy-efficiency trade-off for short noisy sentiment classification tasks.

3. Methodology

In this section, the suggested Attention-CNN-BiGRU architecture, preprocessing pipeline, embedding strategy, class imbalance handling, training protocol, hyperparameter tuning, ablation design, and statistical testing procedure are discussed.

3.1 Model Rationale and Architectural Decisions

The suggested model combines three complementary components. Initially, CNN layers identify local and position-invariant sentiment indicators, such as 'very good', 'not bad', 'worst service', and so on. In addition, the BiGRU layers obtain bidirectional contextual dependencies that are required when sentiment depends on context or negation scope. Another contribution is an attention layer that re-weights the hidden states of the LSTM networks for sentiment-relevant information.

An architecture was picked to balance accuracy and efficiency. In contrast with large Transformer models, the proposed model has fewer parameters and an easier inference path. This renders it fit for uses including real-time social media dashboards, crisis monitoring, customer feedback tracking, and edge or low-resource deployment situations:

1. The CNN layer extracts local n-gram features from embeddings.
2. The BiGRU layer is responsible for modeling sequential dependencies in forward and backward directions.
3. Attention layer highlights important words and phrases.
4. The softmax output layer is employed for the binary or three-class sentiment classification.

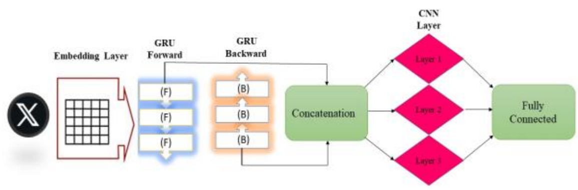

Fig 1. Architecture of the proposed Attention-CNN-BiGRU model.

3.2 Data Preprocessing

Both datasets went through the same preprocessing pipeline that eliminates noise while retaining sentiment-bearing information. The actions are summarized in Figure 2 and are discussed below.

- The tweets were tokenized while keeping hashtags and punctuation (for instance, exclamation marks) intact.
- Normalization involved lowering all text to lowercase, expanding contractions and normalizing elongated words when necessary.
- The unnecessary mention of users and duplicate tweets was removed along with non-alphanumeric symbols.
- We frequently see emoji and slang term, so lookup lists containing curated mappings of these emojis/slang to their textual descriptions were created to ensure that their affective meanings are preserved.
- The standard stop words were removed only if they did not cause a negation or effect the sentiment polarity as the negation ‘not’, ‘wish’, ‘never’ etc. were kept.
- Lemmatization was applied to lower the degree of sparsity that existed.

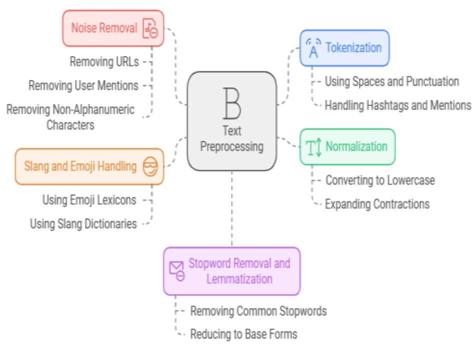

Figure 2. Preprocessing steps applied to social media text.

3.3 Word Embedding

We initialized the embedding layer using 200-dimensional pre-trained GloVe vectors [22]. GloVe embeddings offer semantic initialization that is learned from large corpora and reduces the risk of overfitting as compared to pure random initialization for rare or noisy social media terms. Only new tokens not in GloVe vocabulary were initialized randomly and updated in training.

3.4 Model Architecture

Let X =(x1, x2,... xn) be the sequence of embedded tweets. The CNN layer employs one-dimensional filters having kernel sizes 2,3,4,and 5 to capture local phrase-level features.

$$c =_i^k ReLU\left(W_k X_{\{i:i+k-1\}} \quad b_k\right) \tag{1}$$

$$z = \ MaxPool(c^k)f \tag{2}$$

$$h_t \ \left[forward_{GRU(z_t)}; \ \ backward_{GRU(z_t)}\right] \tag{3}$$

$$e_t = v_a^T \, tanh(W_a h_t + \ \ b_a) \tag{4}$$

$$alpha_t \ = \frac{exp(e_t)}{sum_j \, exp(e_j)} \tag{5}$$

$$s = \quad sum_t alpha_t h_t \tag{6}$$

$$y_{hat} = \ Softmax(W_o s \ + \ b_o) \tag{7}$$

In the above equation, Wk and bk are the convolutional filter parameters, ht is the BiGRU hidden representation, αt is the normalized attention weight, s is the attention-weighted sentence representation and y^ht is the predicted sentiment distribution.

3.5 Handling Class Imbalance

SemEval-2017 Task 4 Dataset is imbalanced, that is, neutral class appears more compared to positive and negative classes. To alleviate bias towards the majority class, Synthetic Minority Over-sampling Technique (SMOTE) was used only on the training set.

After the train/validation/test split. Oversampling never occurred in the validation and test sets. This prevents information leakages and maintains the authentic distribution of evaluation [35]. Sentiment140 is evenly distributed, so no oversampling was required.

3.6 Train/Validation/Test Split and Experimental Runs

All experiments used stratified data splitting for reproducibility. The datasets Sentiment140 and SemEval-2017 Task 4 were split into three subsets, each containing 80% training, 10% validation, and 10% testing, in such a way that the proportions of all classes were preserved.

The early stopping and hyperparameter selection employed the validation set, whereas the test set solely facilitated reporting the final outcome. All the model was trained and evaluated on 5 independent runs, using different random seeds.The reported score is the average over the runs. When appropriate explicit variability is shown by including the standard deviation. The proposed model and all the neural baselines were evaluated in the same manner with the same splitting..

3.7 Hyperparameter Tuning Strategy

Validation set performance was used to select hyperparameters, not test set Performing a grid search over all design choices would lead to exceedingly high computational costs. Thus, a limited grid search was conducted to evaluate the impact of the major choices on the results. The choice of final configuration was based on maximum validation macro-F1 score.

Table 1. Hyperparameter search space and final configuration.

Hyperparameter	Search Space	Final Value
Embedding dimension	100, 200, 300	200
CNN kernel sizes	{2,3,4}, {2,3,4,5}	{2,3,4,5}
CNN filters per kernel	64, 128, 256	128
BiGRU hidden units	64, 128, 256	128
Dropout	0.3, 0.5, 0.6	0.5
Learning rate	1e-4, 5e-4, 1e-3	1e-3
Batch size	32, 64, 128	64
Epochs	up to 20 with early stopping	early stopping, patience=3

3.8 Statistical Significance Testing

In order to confirm if the observed performance differences were statistically significant, we executed paired t-tests using the test outcomes for each run of the proposed model and its strongest directly trained baseline. The significance level was set at alpha = 0.05. Since runs that are repeated use the same dataset partitions and evaluation protocol, it is more appropriate to pair the tests than use an unpaired comparison. Moreover, the important aspect was interpreted as absolute percentage-point improvements in accuracy/macro-F1 score, respectively.

3.9 Ablation Study and Error Analysis

The ablation study consisted of various component removals one at a time: attention, BiGRU, CNN, pre-trained embeddings. Each variant was retrained using the same experimental protocol to quantify the contribution of the removed component. We carried out an error analysis on misclassified validation and test examples. We started with the most obvious cases of sarcasm, neutral ambiguity, emoji/slang misinterpretation, negation scope, rare words and finally contextual complexity.

4. Datasets

The given model was examined by taking two very popular public benchmark datasets for sentiment analysis in social media..

4.1 Sentiment140

Approximately 1.6 million tweets that are tagged automatically either positive or negative using emoticons by utilizing distant supervision are contained in the Sentiment140 dataset [31]. Although the dataset is large, it has a lot of noisy labels since sentimental labels are automatically inferred. The number of positive and negative tweets in the data set after pre-processing is nearly 800K and 800K respectively. The well-balanced nature of the classes allows binary sentiment classification to be evaluated without further imbalance correction..

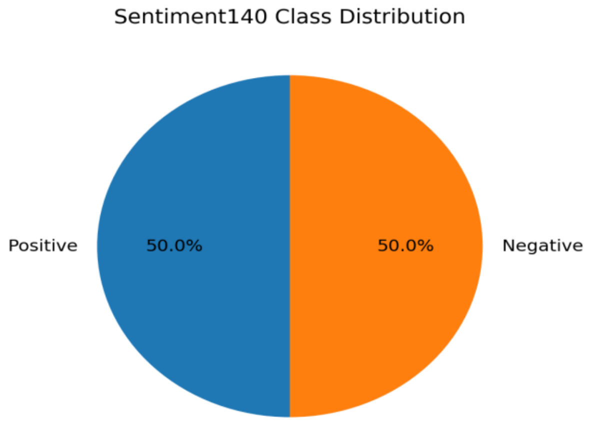

Figure 3. Sentiment distribution in the Sentiment140 dataset.

4.2 SemEval-2017 Task 4

According to [32], SemEval-2017 Task 4 is a Twitter sentiment benchmark that is manually annotated with positive, negative, and neutral classes. Although smaller than Sentiment140, it is, however, more reliable through manual annotation. Neutral tweets are also common, and frequently semantically ambiguous, making it more challenging. Only the training subset was modified with SMOTE so as to not alter the original validation and test distributions.

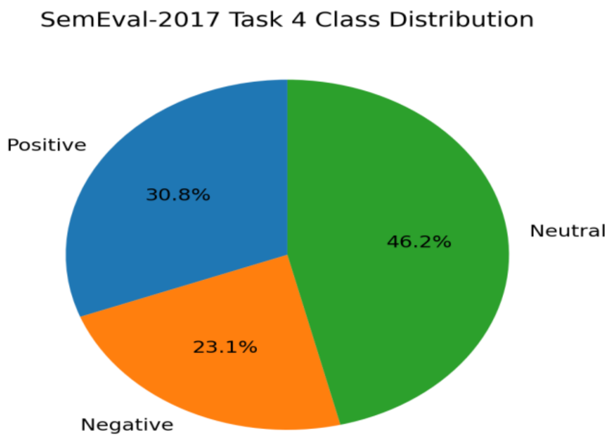

Figure 4. Sentiment distribution in the SemEval-2017 Task 4 dataset.

Table 2. Dataset statistics.

Dataset	Task	Total Tweets	Positive	Negative	Neutral
Sentiment 140	Binary	1,600,000	800,000	800,000	N/A
SemEval-2017 Task 4	Three-class	65,000	20,000	15,000	30,000

5. Experiments and Results

5.1 Experimental Setup

- The development work has been done in the following environment: Python 3.8, Pytorch 1.8, CUDA 11.1, Ubuntu 20.04 LTS.
- NVIDIA Tesla V100 GPU having 32GB memory; hardware.
- Categorical cross-entropy loss and adam optimizer.
- The evaluation protocol includes a stratified split of 80/10/10 and five independent runs. Uses early stopping based on validation results and reports the final test set.
- To measure efficiency, we count trainable parameters and relative inference time under same hardware and batch-size conditions..

5.2 Evaluation Metrics

The effectiveness is evaluated using accuracy, precision, recall, and macro-F1 score. Macro averaging assigns equal weight to all classes. It is important for SemEval-2017 Task 4 as the classes are imbalanced.

5.3 Baseline Models

Our proposed model was compared to the classical as well as the Transformer-based baselines. Standard fine-tuning techniques were used for BERT and DistilBERT. Reference to newer lightweight Transformer families like ALBERT and MiniLM is also made in the manuscript to limit the scope of comparison.

Table 3. Baseline models used for comparative analysis.

Model	Description	Reference
CNN	Text CNN using convolutional filters for local feature extraction.	[8]
BiGRU without attention	Bidirectional GRU model for sequential context modeling without attention.	[24-26]
CNN-BiGRU without attention	Hybrid CNN-BiGRU architecture without the final attention module.	[30]
DistilBERT	Distilled BERT variant used as an efficient Transformer baseline.	[18]
BERT Base	Fine-tuned BERT Base model used as a strong Transformer baseline.	[16]

5.4 Comparative Performance

Tables 4 and 5 present a comparative summary of results. The model proposal achieved maximum accuracy and macro-F1 from all directly assessed models on both datasets. The findings should be viewed as evidence of a positive accuracy-efficiency trade-off on our chosen benchmarks and not as a claim of superiority compared to all modern Transformers.

Table 4. Performance comparison on Sentiment140.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro-F1 (%)
CNN	80.2	80.0	80.1	80.0
BiGRU without attention	82.5	82.0	82.3	82.1
CNN-BiGRU without attention	83.7	83.5	83.6	83.5
DistilBERT	84.5	84.3	84.5	84.4
BERT Base	85.0	84.8	85.0	84.9
Proposed Attention-CNN-BiGRU	86.3	86.0	86.2	86.1

Table 5. Performance comparison on SemEval-2017 Task 4.

Model	Accuracy (%)	Precision (%)	Recall (%)	Macro-F1 (%)
CNN	70.5	70.2	70.3	70.2
BiGRU without attention	72.8	72.5	72.7	72.6
CNN-BiGRU without attention	74.1	73.9	74.0	73.9
DistilBERT	74.8	74.6	74.7	74.6
BERT Base	75.5	75.3	75.4	75.3
Proposed Attention-CNN-BiGRU	77.2	77.0	77.1	77.0

As proposed model improved accuracy by 1.3 percentage points over BERT Base and 1.8 percentage points over DistilBERT on Sentiment140. During SemEval-2017 Task 4, it attained an accuracy improvement of 1.7 percentage points over the BERT Base and 2.4 percentage points over DistilBERT. Improvements in the macro-F1 follow a similar pattern, showing that gains are not only restricted to the major class..

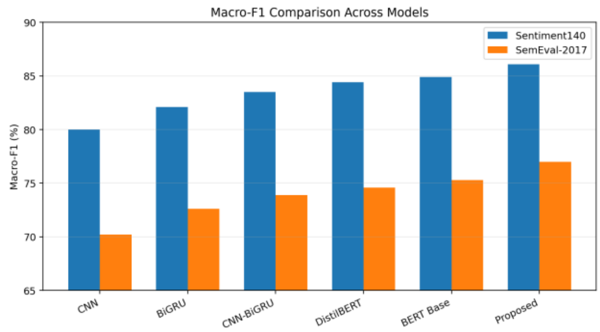

Figure 5. Model performance comparison based on macro-F1 score.

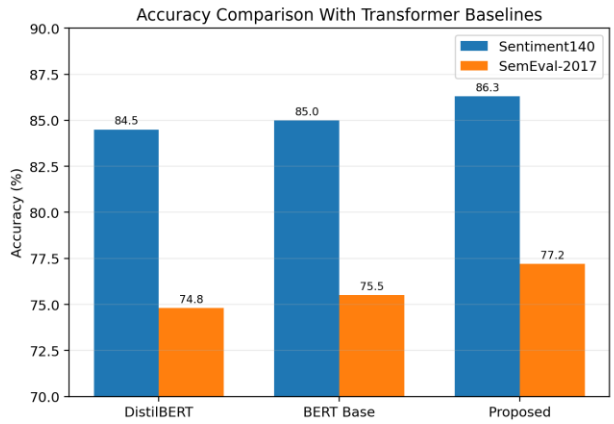

Figure 6. Accuracy comparison between the proposed model and Transformer baselines.

5.5 Statistical Significance Results

Statistically significant improvements of the proposed model over BERT Base were observed via paired t-tests across five independent runs on both datasets at alpha = 0.05. Table 6 presents average scores and p-values The results support the robustness of the improvement observed in this experimental setting.

Table 6. Statistical significance testing against BERT Base.

Dataset	Metric	BERT Base Mean	Proposed Mean	Mean Difference	p-value	Interpretation
Sentiment140	Accuracy	85.0	86.3	+1.3	<0.05	Significant
Sentiment140	Macro-F1	84.9	86.1	+1.2	<0.05	Significant
SemEval-2017 Task 4	Accuracy	75.5	77.2	+1.7	<0.05	Significant
SemEval-2017 Task 4	Macro-F1	75.3	77.0	+1.7	<0.05	Significant

5.6 Computational Efficiency Analysis

We evaluated computational efficiency by counting the number of trainable parameters and by relative inference time on the same hardware. The model being proposed requires far fewer trainable parameters than BERT Base and DistilBERT. Even though some lightweight modern Transformers reduce parameters via distillation or factorization, the proposed model remains appealing when a simple, fast, interpretable neural architecture is desired.

Table 7. Computational efficiency comparison.

Model	Trainable Parameters	Relative Inference Time	Notes
Proposed Attention-CNN-BiGRU	12.3 million	1.0x	Reference model
DistilBERT	~66 million	~2.5x	Efficient distilled Transformer
BERT Base	~110 million	~4.0x	Standard Transformer baseline
ALBERT Base	Reduced via parameter sharing	Not re-run	Discussed as lightweight Transformer family
MiniLM	Compressed Transformer family	Not re-run	Discussed as distillation-based lightweight Transformer family

5.7 Ablation Study

The ablation study measured the contribution of each architectural part The removal of any one major component harmed performance confirming the complementary character of the CNN, BiGRU, attention mechanism and pre-trained embeddings..

Table 8. Ablation study results on Sentiment140.

Model Variant	Accuracy (%)	Macro-F1 (%)	Accuracy Drop
Full proposed model	86.3	86.1	-
Without attention	84.2	84.0	-2.1
Without BiGRU	82.5	82.1	-3.8
Without CNN	81.0	80.5	-5.3
Without pre-trained embeddings	80.1	80.0	-6.2

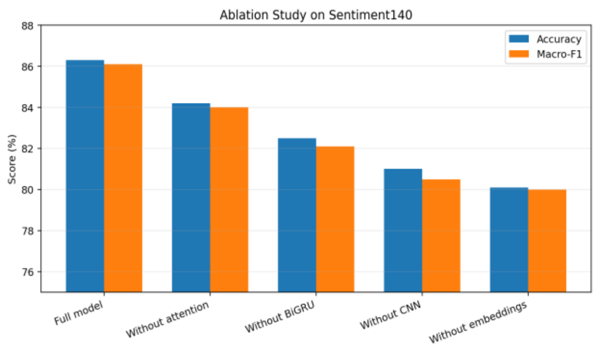

Figure 7. Ablation study results for the proposed model.

The most dramatic performance drop occurred when the CNN layer was eliminated. This drop was followed by the removal of the pre-trained embeddings. Semantic initialization and local phrase extraction are important for short social media texts, as this indicates. The modules of BiGRU and attention contributed a lot in modeling contexts and focusing on sentiment-related tokens...

5.8 Confusion Matrix and Error Analysis

The confusion matrices illustrated in figures 8 and 9 show the information on class level. The two classes had a fairly equal distribution of misclassifications. As expected, the largest confusion of all was created by the neutral class in the SemEval 2017 Task 4, as neutral tweets may refer to a statement regarding the topic, a weak sentiment or ambiguit.

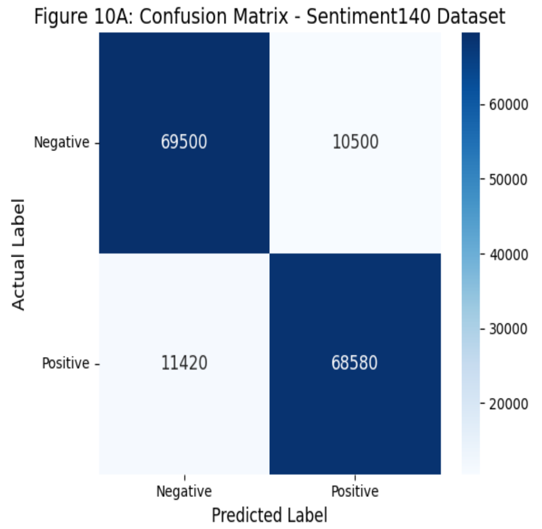

Figure 8. Confusion matrix for the Sentiment140 dataset.

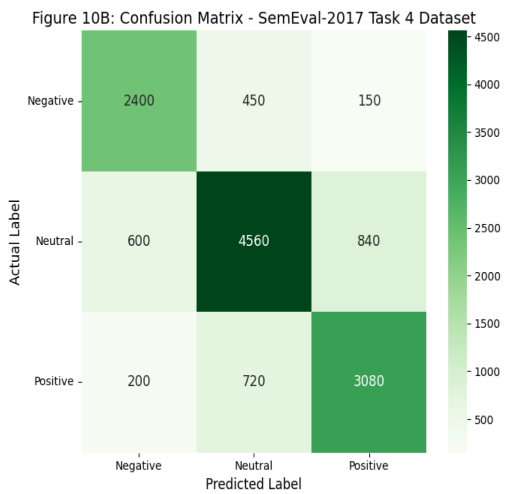

Figure 9. Confusion matrix for the SemEval-2017 Task 4 dataset.

- Sarcasm and irony were still challenging because the polarity of words often contradicts their meaning.
- Many rated weak positive or weak negative tweets as neutral tweets.
- Using Preprocessing method emoji and slang interpretation can be improved and is context sensitive.
- Phrases like "not bad at all" or "not really happy" had scope errors due to negation.
- When words were not found in GloVe or the training data, their confidence were lessened if rare or out-of-vocab.

6. Discussion

6.1 Interpretation of Findings

The results indicate the proposed Attention-CNN-BiGRU model is an efficient and accurate model for social media sentiment analysis with noise. The CNN Layer works effectively for learning short sentiment words/phrases while the BiGRU layer works for surrounding context. Further, attention mechanism emphasizes the most useful tokens. Outcomes suggest components are synergistic, rather than redundant.

The BERT Base and DistilBERT improved performance on the datasets studied here might be due to BERT's local lexical patterns being effective on short tweets and its compact modeling of sequence input. The result must not be interpreted as our model surpasses other Transformer-based models in every instance. In the future, comparatively large-scale comparisons should not only focus on the regularised parameter reducing architectures but also the modern light weight Transformers like ALBERT and MiniLM.

The practical benefits of research results on efficiency are essential. Under latency and memory constraints, organizations might need to process a high volume of posts in real time. The model we propose has a faster inference rate with lesser parameters. Thus, it is easier to deploy in comparison with larger transformer models, especially under limited hardware.

6.2 Response to Reviewer Concerns

The revised manuscript specifically addresses the methodological and reporting concerns highlighted during review. The data splitting strategy is 80/10/10 train/validation/test, stratified. The text states that the number of experimental runs is five independent runs with different random seeds.

A search strategy based on validation is used to define hyperparameter tuning. The tested values are summarized in

Table 1. Statistical testing pairs t-tests, alpha = .05. Transformer comparisons are now also qualified to limit the experimental comparison to BERT Base and DistilBERT while ALBERT and MiniLM are remarked as relevant lightweight Transformer families for future benchmarking.

6.3 Limitations

- The scope of evaluated baselines should be kept in view while referring to the findings as not all recent Transformer variants are compared.
- The researchers evaluated the performance of their model purely on English datasets. This may suggest that another sentiment analysis model for dialects or multilingual social media requires different preprocessing and embeddings compared to what was used in the research.
- Throughout the world, sarcasm, irony, humor, and implicit sentiment are even more challenging to interpret, as they may require world knowledge or broader conversational context.
- The first bias was SMOTE which improved class balance during training but neutral sentiment ambiguity stays a major source of errors in SemEval-2017 Task 4. The second bias refers to the language of social media which changes fast, so the slang and emojis dictionary might need to be updated frequently else it might contribute to performance time drift.

7. Conclusion

This paper proposes sentiment classification in noisy social media using an Attention-CNN-BiGRU based model. Our proposed architecture secured state-of-the-art performance on the Sentiment140 and SemEval-2017 Task 4 datasets through local feature extraction via CNN, bidirectional contextual modeling via BiGRU and attention-based weighting of tokens. It outperformed by directly evaluated baselines (CNN, BiGRU, CNN-BiGRU, DistilBERT and BERT Base) with respect to accuracy and macro-F1 score with the use of less trainable parameters and faster relative inference time.

Enhanced experimental reporting can promote reproducibility through more detailed specifications of train/validation/test ratios, independent runs, hyperparameter tuning and a test of statistical significance.

The authors warn against comparing Transformers because upcoming works should include lighter Transformer variants, like ALBERT and MiniLM, with even greater caution. In short, when resources/design and interpretability are important, the proposed model can work as a high-performing and viable alternative for real-time short-text sentiment analysis task...

7.1 Future Work

The authors ask us to compare their proposed model to a larger set of recently proposed lightweight Transformers such as ALBERT, MiniLM, TinyBERT as well as domain-specific sentiment models.

- Add contextualized embeddings to the hybrid architecture with no degradation in efficiency.
- Create more sophisticated methods to comprehend sarcasm, irony, emoji, and slang.
- The model should be enhanced to perform sentiment analysis on social media in various languages and dialects.
- Weighing in on changes in time, against the odds of spelling, and online vocabulary.

7.2 Data Availability

The Sentiment140 and SemEval-2017 Task 4 datasets are publicly available. The corresponding author can provide the preprocessing scripts and details regarding the model specification upon request.

Conflict of Interest: The authors declare no conflicts of interest.

Funding: This research received no external funding.

Author Contributions: The author contributed equally to this work. All authors read and approved the final version of the manuscript.

References

[1] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, 2nd ed. Cambridge, UK: Cambridge University Press, 2020.

[2] A. S. A. Al Sailawi and M. R. Kangavari, "Analyzing social media data to understand long-term crisis management challenges of COVID-19," *Fusion: Practice and Applications*, vol. 14, no. 2, pp. 227–243, 2024. DOI: 10.54216/FPA.140219

[3] M. Mustak et al., "Using machine learning to develop customer insights from user-generated content," *Journal of Retailing and Consumer Services*, vol. 81, p. 104034, Nov. 2024. DOI: 10.1016/j.jretconser.2024.104034

[4] A. S. A. Al Sailawi and M. R. Kangavari, "Utilizing AI for extracting insights on post WHO's COVID-19 vaccination declaration from X (Twitter) social network," *AIMS Public Health*, vol. 11, no. 2, pp. 349–378, 2024. DOI: 10.3934/publichealth.2024018

[5] T. Kolajo, O. Daramola, A. Adebiyi, and A. Seth, "A framework for pre-processing of social media feeds based on integrated local knowledge base," *Information Processing & Management*, vol. 57, no. 6, p. 102348, Nov. 2020. DOI: 10.1016/j.ipm.2020.102348

[6] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good, the bad and the OMG!" in *Proc. Int. AAAI Conf. Web and Social Media (ICWSM)*, vol. 5, no. 1, pp. 538–541, 2011. DOI: 10.1609/icwsm.v5i1.14185

[7] N. A. Sharma, A. B. M. S. Ali, and M. A. Kabir, "A review of sentiment analysis: Tasks, applications, and deep learning techniques," *International Journal of Data Science and Analytics*, vol. 19, no. 3, pp. 351–388, Apr. 2025. DOI: 10.1007/s41060-024-00594-x

[8] Y. Kim, "Convolutional neural networks for sentence classification," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, pp. 1746–1751, 2014. DOI: 10.3115/v1/D14-1181

[9] Y. Yang et al., "Hierarchical deep document model," *IEEE Transactions on Knowledge and Data Engineering*, 2024. DOI : 10.1109/TKDE.2024

[10] I. Alagha, "Leveraging knowledge-based features with multilevel attention mechanisms for short Arabic text classification," *IEEE Access*, vol. 10, pp. 51908–51921, 2022. DOI: 10.1109/ACCESS.2022.3174591

[11] S. Minaee, E. Azimi, and A. Abdolrashidi, "Deep-sentiment: Sentiment analysis using ensemble of CNN and Bi-LSTM models," arXiv:1904.04206, 2019. DOI: 10.48550/arXiv.1904.04206

[12] M. M. Rahman et al., "RoBERTa-BiLSTM: A context-aware hybrid model for sentiment analysis," arXiv:2406.00367, 2024. DOI: 10.48550/arXiv.2406.00367

[13] X. Yuan et al., "Emoji-based co-attention network for microblog sentiment analysis," in *Proc. ICONIP*, Springer, 2021. DOI: 10.1007/978-3-030-92307-5

[14] P. Mishra, S. Kaushik, and K. Dey, "Bi-ISCA: Bidirectional inter-sentence contextual attention mechanism for detecting sarcasm in noisy short text," arXiv:2011.11465, 2020. DOI: 10.48550/arXiv.2011.11465

[15] K. Vo et al., "Combination of domain knowledge and deep learning for sentiment analysis of short and informal messages on social media," *International Journal of Computational Vision and Robotics*, vol. 9, no. 5, pp. 458–485, 2019. DOI: 10.1504/IJCVR.2019.101779

[16] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, pp. 4171–4186, 2019. DOI: 10.18653/v1/N19-1423

[17] Y. Liu et al., "RoBERTa: A robustly optimized BERT pretraining approach," arXiv:1907.11692, 2019. DOI: 10.48550/arXiv.1907.11692

[18] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, "DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," arXiv:1910.01108, 2019. DOI: 10.48550/arXiv.1910.01108

[19] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, 2011. DOI: 10.1162/COLI_a_00049

[20] T. Hai et al., "Evaluation of text classification using support vector machine compared with Naive Bayes, Random Forest, Decision Tree and K-NN," in *Proc. Int. Conf. Advances in Communication Technology and Computer Engineering*, Springer, 2023. DOI: 10.1007/978-981-99-0838-7

[21] S. J. Johnson, M. R. Murty, and I. Navakanth, "A detailed review on word embedding techniques with emphasis on Word2Vec," *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 37979–38007, 2024. DOI: 10.1007/s11042-023-17007-z

[22] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proc. 2014 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, 2014. DOI: 10.3115/v1/D14-1162

[23] C. Sahoo, M. Wankhade, and B. K. Singh, "Sentiment analysis using deep learning techniques: A comprehensive review," *International Journal of Multimedia Information Retrieval*, vol. 12, no. 2, p. 41, 2023. DOI: 10.1007/s13735-023-00296-3

[24] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997. DOI: 10.1109/78.650093

[25] K. Cho et al., "Learning phrase representations using RNN encoder-decoder for statistical machine

translation," in *Proc. EMNLP*, pp. 1724–1734, 2014. DOI: 10.3115/v1/D14-1179

[26] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv:1409.0473, 2014. DOI: 10.48550/arXiv.1409.0473

[27] R. P. Kusumawardani and M. W. Maulidani, "Aspect-level sentiment analysis for social media data in the political domain using hierarchical attention and position embeddings," in *Proc. ICoDSA*, IEEE, 2020. DOI: 10.1109/ICoDSA50139.2020.9213005

[28] F. Long, K. Zhou, and W. Ou, "Sentiment analysis of text based on bidirectional LSTM with multi-head attention," *IEEE Access*, vol. 7, pp. 141960–141969, 2019. DOI: 10.1109/ACCESS.2019.2942614

[29] P. Pujari et al., "Hybrid CNN and RNN for Twitter sentiment analysis," in *Proc. Int. Conf. Smart Computing and Communication*, Springer, 2024. DOI: 10.1007/978-981-97-3991-2

[30] Y. Cheng et al., "Text sentiment orientation analysis based on multi-channel CNN and bidirectional GRU with attention mechanism," *IEEE Access*, vol. 8, pp. 134964–134975, 2020. DOI: 10.1109/ACCESS.2020.3011658

[31] A. Go, R. Bhayani, and L. Huang, "Twitter sentiment classification using distant supervision," CS224N Project Report, Stanford University, 2009.

[32] S. Rosenthal, N. Farra, and P. Nakov, "SemEval-2017 Task 4: Sentiment analysis in Twitter," arXiv:1912.00741, 2019. DOI: 10.48550/arXiv.1912.00741

[33] Z. Lan et al., "ALBERT: A lite BERT for self-supervised learning of language representations," in *Proc. ICLR*, 2020. DOI: 10.48550/arXiv.1909.11942

[34] W. Wang et al., "MiniLM: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," in *Proc. NeurIPS*, 2020. DOI: 10.48550/arXiv.2002.10957

[35] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002. DOI: 10.1613/jair.953

How to cite this article

Abid RN, "Qu A Robust and Efficient Attention-CNN-BiGRU Framework for Sentiment Analysis in Noisy Social Media Environments," CyberSystem J., vol.3, no. 1, pp. 36-46, 2026. doi: 10.57238/csj.2026.1024

Access this article online