A Comparative Study of LSTM, Convolutional, and Transformer Models for Assessing Sentiment on Social Media

Aryan Raj,

Arshdeep Angural and

Dnyanesh Ashok Magare

Dyal Singh Evening College, University of Delhi

Bachelor of Arts Programme (Computer Application with Economics)

Dr. Ram Nayan Verma

May 6, 2025

A Comparative Study of LSTM, Convolutional, and Transformer Models for Assessing Sentiment on Social Media

Abstract

This report summarizes a study comparing three deep learning models—LSTM, CNN, and a transformer-based DistilBERT—for binary sentiment classification of Twitter. We used the Sentiment140 dataset (≈1.6 million tweets) labeled positive/negative. Text preprocessing included relabeling (0→0, 4→1), noise removal (dropping IDs, timestamps, URLs, mentions, hashtags, special characters), tokenization, and padding. Each model was trained under similar conditions (same optimizer and early stopping) to ensure a fair comparison. The LSTM model achieved the highest accuracy (82.56%), followed by CNN (81.93%) and DistilBERT (80.00%). We evaluated performance using accuracy, loss, precision, recall, and F1-score; confusion matrices were plotted to visualize misclassifications. Overall, the results indicate that the sequence-based LSTM performed slightly better than the CNN and transformer models for social media sentiment analysis.

Objective

The study aimed to **evaluate and compare** three deep learning architectures—LSTM, CNN, and DistilBERT—for sentiment analysis of social media text. Specifically, we tested each model's ability to classify tweets as positive or negative, under identical experimental conditions. By benchmarking their accuracy and other metrics, the study sought to determine which architecture is most effective for sentiment classification of Twitter data.

Experimental Setup

- **Data:** The Sentiment140 dataset contains ~1.6 M tweets labeled as positive or negative. Labels originally 0 (negative) and 4 (positive) were converted to binary (0/1) for analysis.

- **Preprocessing:** Tweets were cleaned (removing noise), tokenized, and padded to equal length. Labels were normalized to binary (0/1).
- **Models:** The three architectures tested were:

  (1) **LSTM (Long Short-Term Memory):** An RNN variant that captures long-range dependencies in text sequences. It processes words in order and uses memory cells to learn context over time. In our experiments, the LSTM was set up with word embeddings and trained to recognize sentiment from the sequence of tokens.

  (2) **CNN (Convolutional Neural Network):** Applies convolutional filters over word embeddings to detect local features or n-grams in the text. The CNN learns to identify important phrases or patterns (e.g., "not good" vs "good") that signal sentiment. It generally trains faster than RNNs and is effective at capturing short-range relationships.

  (3) **DistilBERT (Transformer-based):** A smaller, faster version of BERT (Bidirectional Encoder Representations) that uses attention mechanisms. It is pre-trained on large text corpora and then fine-tuned for sentiment classificationfile-rngbhupakesxnbhiiz3v8k. In this study, DistilBERT was loaded via Hugging Face Transformers and fine-tuned on our tweet dataset. Because it is pre-trained, DistilBERT can leverage transfer learning from general language understanding.
- **Training:** All models used the same environment (TensorFlow on Google Colab) and similar optimization settings for a fair comparison. Specifically, LSTM and CNN were trained for 10 epochs (batch size 128, Adam optimizer, with early stopping), while DistilBERT was fine-tuned for 4 epochs.

Results

- **Overall Accuracy:** The LSTM model achieved the highest test accuracy (82.56%). The CNN followed closely at 81.93%, and DistilBERT attained 80.00%.
- **Performance Ranking:** LSTM > CNN > DistilBERT (by accuracy). All three models performed reasonably well (~80% accuracy), indicating that each can learn to classify tweet sentiment effectively.
- **Key Finding:** The LSTM's leading performance suggests it more effectively captured the sequential nature of text in this task. The CNN also performed competitively by extracting local patterns. DistilBERT's slightly lower accuracy may be due in part to limited fine-tuning (4 epochs), but still reflects strong performance given its pre-training on large text corpora.

Analysis

The results indicate that **all three models** are capable of sentiment analysis on social media text, with only modest differences in accuracy. The LSTM's higher accuracy (82.56%) suggests that capturing the order and context of words in a tweet provides useful information. The CNN's strong performance (81.93%) shows that local patterns (like common phrases) are also highly predictive of sentiment. DistilBERT, while slightly behind (80.00%), benefited from pre-training and transferred language understanding; its lower score here may be due to minimal fine-tuning (only 4 epochs).

Each model has trade-offs: the CNN trained faster and with simpler architecture, while LSTM required more training time but captured dependencies. DistilBERT offers the advantage of transfer learning, meaning it can work well even with fewer task-specific training examples. In practice, the choice may depend on resources: for example, if very high accuracy is needed and computational cost is acceptable, one might favor LSTM or a fully fine-tuned transformer. If quick training or online adaptation is needed, a CNN or a distilled transformer could be preferred.

Conclusion

This comparative study showed that LSTM, CNN, and transformer models can all effectively classify tweet sentiment. Under our experimental settings, **LSTM achieved the best performance (82.56% accuracy)**, followed closely by CNN (81.93%) and then DistilBERT (80.00%). The LSTM's sequential nature helped it capture subtle sentiment cues, while the CNN's convolutional filters extracted strong local features. DistilBERT, even with limited fine-tuning, performed competitively, demonstrating the value of pre-trained language models. These results confirm that traditional deep models like LSTM remain highly effective for sentiment tasks, but transformer-based models are also promising, especially when leveraging large pre-training.

Future Work

Potential extensions of this work include:

- **Larger Transformer Models:** Experiment with full-size BERT, RoBERTa, or other state-of-the-art transformers, fine-tuning them for more epochs to potentially boost accuracy.
- **Advanced Preprocessing:** Incorporate linguistic features such as part-of-speech tags or named entities to enrich text representations before the model.
- **Ensemble Methods:** Combine multiple models (e.g., LSTM+CNN or model ensembles) to leverage their complementary strengths.

- **Multilingual Data:** Extend analysis to non-English tweets or multilingual datasets to test model generality across languages.

  These directions can further improve sentiment classification on social media and inform model selection for real-world applications

References

[1] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment Analysis on Twitter Data. *International Journal of Innovative Research in Advanced Engineering*, *2*(1).

[2] Kanimozhi, P. (2019). *Sentiment Analysis Using Twitter Dataset*. https://www.semanticscholar.org/paper/Sentiment-Analysis-Using-Twitter-Dataset-Kanimozhi/88d649a2a134a0750483d928d3d291444bfe9b6d

[3] Chen, L.-C., Lee, C.-M., & Chen, M.-Y. (2020). Exploration of social media for sentiment analysis using deep learning. *Soft Computing*, *24*(11), 8187–8197. https://doi.org/10.1007/s00500-019-04402-8

[4] Wadawadagi, R., & Pagi, V. (2020). Sentiment analysis with deep neural networks: Comparative study and performance assessment. *Artificial Intelligence Review*, *53*(8), 6155–6195. https://doi.org/10.1007/s10462-020-09845-2.

[5] Gandhi, U., Kumar, P. M., Babu, G. C., & Karthick, G. (2021). Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). *Wireless Personal Communications*. https://www.semanticscholar.org/paper/Sentiment-Analysis-on-Twitter-Data-by-Using-Neural-Gandhi-Kumar/96574b0e64da43e9b4f929bdb394841b73f76e81

[6] Srinivas, A. C. M. V., Satyanarayana, Ch., Divakar, Ch., & Sirisha, K. P. (2021). Sentiment Analysis using Neural Network and LSTM. *IOP Conference Series:*

Materials Science and Engineering, *1074*(1), 012007. https://doi.org/10.1088/1757-899X/1074/1/012007

[7] Kastrati, M., Kastrati, Z., Shariq Imran, A., & Biba, M. (2024). Leveraging distant supervision and deep learning for twitter sentiment and emotion classification. *Journal of Intelligent Information Systems*, *62*(4), 1045–1070. https://doi.org/10.1007/s10844-024-00845-0.

[8] Choudhury, R. R., Dey, S., & Paul, P. (2024). *A Comparative Study of LSTM Models on Sentiment Analysis*. 1–6. 2024 6th International Conference on Computational Intelligence and Networks (CINE). https://doi.org/10.1109/cine63708.2024.10881938

[9] Bashiri, H., & Naderi, H. (2024). Comprehensive review and comparative analysis of transformer models in sentiment analysis. *Knowledge and Information Systems*, *66*(12), 7305–7361. https://doi.org/10.1007/s10115-024-02214-3

[10] Islam, M. T., Parvin, F., Sazan, S. A., & Bin Amir, T. (2024). *Comparative Analysis of Sentiment Classification on IMDB 50k Movie Reviews: A Study Using CNN, LSTM, CNN-LSTM, and BERT Models*. 512–517. 2024 IEEE International Conference on Power, Electrical, Electronics and Industrial Applications (PEEIACON). https://doi.org/10.1109/peeiacon63629.2024.10800035

[11] TensorFlow. (2023). *An end-to-end open source machine learning platform*. https://www.tensorflow.org/

[12] Google. (n.d.). *Google Colaboratory*. Google Colab. https://colab.research.google.com/