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

Aryan Raj,

Arshdeep Angural

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

1. Abstract

In this study, three different deep learning models, Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and a Transformer-based model (DistilBERT), were investigated to explore the sentiment analysis of social media content. They were evaluated using the Sentiment140 dataset, made up of 1.6 million labeled Twitter tweets to provide a warm purposeful comparison of performance. Preprocessing consisted of normalization of labels (reshaping labels from 0 and 4 to 0 and 1), removal of columns, cleaning the text, tokenizing, and adding padding. All models were run with TensorFlow in the Google Colab environment while keeping the training parameters to be identical for the sake of consistency. LSTM and CNN training was executed for 10 epochs with a batch size of 128 with the Adam optimizer and early stopping, whereas DistilBERT was fine-tuned for 4 epochs with early stopping. The accuracy, loss, and F1-score were used to evaluate performance. Overall, the results concluded that performance was best in LSTM at 82.56% accurate, CNN at 81.93% accuracy, and followed by DistilBERT at 80.00% accurate. These results show that traditional sequence-based deep learning models, such as LSTM, have an advantage over the transformer models. In contrast, it is worthwhile to note that they have the greatest advantage in performing the sentiment classification task on social media data when expressed in a text format.

Keywords

Sentiment Analysis, LSTM, CNN, DistilBERT, Social Media, Deep Learning, Text Classification, Natural Language Processing (NLP), Transformer Models, TensorFlow

2. Introduction

In the modern digital age, social media websites like Twitter and Facebook are now significant spaces where individuals can put forward their views, emotions, and opinions. These websites generate massive amounts of user-generated content daily and give important data to influence public opinion. This has resulted in an increasing application of sentiment analysis, also called opinion mining, an NLP approach applied to label and detect subjective information in text as positive, negative, or neutral [2, 5].

NLP is a subfield of artificial intelligence that allows computers to comprehend, analyse, and react to human language in a meaningful manner. It drives many applications used in real life, including chatbots, machine translation, spell checking, information retrieval, speech recognition, and advertisement targeting [5]. Sentiment analysis is one of the most widely used tasks among these because of its broad applicability in marketing, politics, healthcare, and customer service.

Sentiment analysis not only assists companies in tracking company reputation or customer satisfaction but also aids decision-making based on public sentiment. For instance, knowing the sentiment behind movie or product reviews assists new customers in making sound decisions [2]. Nevertheless, sentiment analysis of social media text is difficult because social media texts have an informal format, slang usage, emojis, abbreviations, and short length. To meet these challenges, sentiment analysis systems usually involve tokenization, where the text is split into phrases or words as initial processing steps [5]

Deep learning is a part of machine learning methods-based artificial neural networks that employ multiple layers for extracting the gigantic-level features of the raw data. Deep learning has been employed in various fields such as signal and image processing. Deep Neural Networks contain multiple neural networks, in which the output of one network serves as an input to the subsequent network. It reads the text data attributes in its own manner, and it also reads about the many layers of features for forecasting that data. The deep learning of the movie review dataset was applied in opinion mining or sentiment analysis.

To handle such complex information successfully, advanced, profound learning approaches like Long Short-Term Memory (LSTM) systems, Convolutional Neural Systems (CNNs), and Transformer-based models like BERT and RoBERTa have shown promising comes about in opinion classification. These models can memorize semantic connections and relevant conditions in content, making them perfect for NLP tasks.

This paper presents a comparative think about of LSTM, CNN, and Transformer-based models for estimation analysis using the Sentiment140 dataset from Kaggle, which contains 1.6 million labeled tweets. This dataset gives a adjusted set of positive and negative estimations and is broadly utilized in benchmarking NLP models.

The aim of this paper is to assess the performance of these models in classifying the sentiments correctly and identify which architecture provides the best outcome on actual Twitter data.

3. Review of Related Work

- This paper introduces A hybrid method of sentiment analysis of Twitter data that combines both corpus-based and dictionary-based methods has proved successful in tweet classification into positive, negative, and neutral categories. The technique uses a blend of unigram models, tree kernel models, and feature-based models, and adds features such as emoticons, handling negations, and capitalization that are related to sentiment. The aim is to encompass the diverse expressions that are typical of microblogging sites. This method improves the sentiment classification function by resolving informal language and context uncertainty that characterizes social media posts [1].

- This paper analyzes emotional analysis based on the perspective of Twitter dataset, which means that the emotional classification is at the level of the sentence based on algorithms such as K-means clustering, CART, and C4.5. Extracting the research that has been identified as an important aspect of public opinion and the result of providing a sentiment classification is positive, negative, or neutral. The results show the effectiveness of the models applied in sentiment analysis and their contribution to the field of text exploitation and wider analysis [2].

- Chen et al. (2020) carried out an in-depth study targeting the process of document-level sentiment classification with the help of sophisticated neural network models to improve the process of analysis. Specifically, they focused on the Bi-directional Long Short-Term Memory (BiLSTM) models, which have been in the limelight for being highly capable of processing sequences of data with high efficiency. The study involved a range of features that were highly extracted from user comment text, allowing the sentiments to be classified into three different categories: positive, negative, or neutral. The experimental results obtained through their study strongly confirmed that the BiLSTM model not only worked efficiently but was also highly capable of detecting the subtle and highly nuanced expressions of sentiment in the text. This result strongly confirmed its use in analyzing highly complex textual data in real-world applications, thus being a strong foundation for future studies in the field [3].

- The research study reported in this research paper is a scientific and empirical examination of some deep neural network (DNN) architectures that have been specifically created to serve the purpose of sentiment classification tasks. The research study provides an overview of the current status of DNN architectures, along with the performance comparison of some DNN architectures on several datasets of sentiment. In order to further enhance one's insight into the outcomes being obtained, research utilizes approaches such as hyperparameter optimization and data visualization. The findings heavily support the incredible performance of DNNs when utilized to automate sentiment classification tasks, showing a very high rate of accuracy in the classification of varied sentiments [4].

- This paper embarks upon a comprehensive study of the sentiment analysis field of Twitter data with a specific goal of performing word-level sentiment classification based on the power of strong deep learning methodologies. The paper strategically employs two distinct categories of neural network models, ie, Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) models, in order to examine comprehensively various features that, in turn, are derived from the words of tweets. Features include prominent entities like word embeddings and stop words, which are systematically used within the overall deep learning framework of the paper. The dataset used within the scope of the current work includes a total of 50,000 movie reviews fetched from IMDB, and the sentiment classification derived in the process achieves high test accuracies of 87.74% for positive sentiment and 88.02% for negative sentiment, respectively, as observed in reference [5].

- The research is based on document-level sentiment analysis using deep learning models to categorize sentiments in the form of tweets as negative or positive. Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and basic Neural Networks are used in the research, testing how efficient they can be on a 1.6 million tweets dataset made available by Kaggle, where LSTM achieved the highest accuracy mark of 87% in sentiment classification. The research highlights the need for processing long sequence data and how they relate to one another, and it is a significant improvement in sentiment analysis methods [6].

- The research analyzes Twitter posts at document level and relies on distant supervision with emojis to categorize tweets under Ekman's six basic emotions. Through a dataset of 17.5 million tweets and BiLSTM with FastText and attention mechanisms the research reached an F1-score of 70.92% for sentiment classification and 54.85% for emotion detection indicating solid results for positive negative and neutral sentiment detection. The presented study enhances our knowledge of efficient sentiment and emotion detection methods for social media platforms because it emphasizes both vast datasets and sophisticated modeling approaches [7].

- This thesis provides a comparison of various types of Long Short-Term Memory (LSTM) based architectures for sentiment analysis, showing their ability to capture long sequences of text. As part of this work, we explored the new models Context and Aspect Memory Network (CAMN) , developed for aspect-level sentiment analysis, and Enhanced Transformer-BiLSTM, which uses the Transformer mechanism with Bidirectional LSTM to improve feature extraction. The models are discussed across four datasets and domains (laptop, restaurant, Twitter) and by showing that hybrid models consistently improved accuracy and context for sentiment classification compared to baseline methods [8].

- The manuscript provides a thorough survey of the methods used for sentiment analysis emphasising the development and use of Transformer-based models. The study used the most current architectures BERT, RoBERTa, XLNet, ELECTRA, DistilBERT, ALBERT, T5 and GPT to focus on their underlying principles and performance applied to sentiment classification tasks. The study carried out a vast set of experiments on 22 benchmark datasets and concluded that T5 in particular performs favourably compared to other models modelling sentiment analysis across multiple domains. Also, XLNet has good performance on the finer-grained sentiment expressions, such as identifying irony and opinions about products. This provides a good example of the individual strengths between Transformer architectures used in sentiment analysis [9].

- The research paper provides a comparative study of sentiment classification of IMDB 50k movie reviews dataset with the focus being on document-level sentiment classification, and the performance of all of the following deep learning architectures namely, CNN, LSTM, CNN-LSTM, and BERT, in classifying sentiment as either positive, negative, or neutral. The article suggests the utility of these deep learning architectures extracting deeper meaning from text and understanding the perception of the audience based off of the patterns learning from the text overall. This is seen as informative information for industry planners and helps applied research for tactical planning in something like targeted marketing [10].

A summary of the reviewed studies is provided in Table 1

Literature Review Table (Table 1)					
S.No.	**Paper Title**	**Authors**	**Year**	**Key Findings**	**Citation Count**
1.	Sentiment Analysis on Twitter Data	Sahayak, Varsha Shete, Vijaya Pathan, Apashabi	2015	used tree kernels and POS-specific features in machine learning to categorize Twitter sentiments with the goal of minimizing the	183

				need for manual feature engineering.	
2.	Sentiment Analysis Using Twitter Dataset	Kanimozhi, P.	2019	Used Naive Bayes for sentence-level sentiment analysis on Twitter data; classified text and emoticons into positive, negative, and neutral categories with good accuracy.	3
3.	Exploration of social media for sentiment analysis using deep learning	Chen, Liang Chu Lee, Chia Meng Chen, Mu Yen	2020	Proposed a sentiment analysis model using Bi-LSTM with a custom military sentiment dictionary; achieved higher accuracy and F1-score compared to using standard dictionaries alone.	83
4.	Sentiment analysis with deep neural networks: comparative study and performance assessment	Wadawadagi, Ramesh Pagi, Veerappa	2020	Carried out empirical comparison of deep neural network models for sentiment classification; studied performance impacts of hyperparameter tuning on various sentiment datasets.	62
5.	Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)	Gandhi, Usha Devi Malarvizhi Kumar, Priyan Chandra Babu, Gokulnath Karthick, Gayathri	2021	Applied CNN and LSTM models with word2vec features for sentiment analysis; achieved high accuracy (~88%) on IMDB data, showing potential for real-time tweet classification.	100

6.	Sentiment Analysis using Neural Network and LSTM	Srinivas, Akana Chandra Mouli Venkata Satyanarayana, Ch. Divakar, Ch. Sirisha, Katikireddy Phani	2021	Applied LSTM, CNN, and Neural Networks for Twitter sentiment analysis; achieved 87% accuracy with LSTM, and suggested real-time analysis via REST API for scalability.	29
7.	Leveraging distant supervision and deep learning for twitter sentiment and emotion classification	Kastrati, Muhamet Kastrati, Zenun Shariq Imran, Ali Biba, Marenglen	2024	Used BiLSTM with FastText and attention mechanism for sentiment and emotion classification; achieved F1-scores of 70.92% for sentiment and 54.85% for emotion detection, outperforming other models.	11
8.	A Comparative Study of LSTM Models on Sentiment Analysis	Choudhury, Ritabrata Roy Dey, Soumik Paul, Prithwineel	2024	LSTM and Convolutional LSTM models perform well in classifying positive sentiments with high precision, recall, and F1-score but struggle with negative sentiment classification, both achieving 97% accuracy overall.	10
9.	Comprehensive review and comparative analysis of transformer models in sentiment analysis	Bashiri, Hadis Naderi, Hassan	2024	T5 outperforms others in sentiment analysis, while XLNet excels in irony and product sentiment; BERT and DistilBERT struggle with complex tasks.	14
10.	Comparative Analysis of Sentiment Classification on IMDB 50k Movie Reviews:	Islam, Md Touhidul	2024	This study compares sentiment analysis models on movie	12

	A Study Using CNN, LSTM, CNN-LSTM, and BERT Models	Parvin, Farjana Sazan, Saad Ahmed Bin Amir, Tariq		reviews, finding RoBERTa+DNN to be the most effective with 92% accuracy and top F1-score.	

4. Methodology

4.1 Purpose

The purpose of this study is to build comparisons of three of the best performing deep learning models: Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNN), and DistilBERT - in terms of using them to perform sentiment analysis in a social media context. The goal is to assess the effectiveness of these models' ability to classify polarities of sentiment (positive or negative) in tweets.

4.2 Experimental Setup and Implementation Overview

The experiment was conducted using the **Sentiment140** dataset, which originally labeled sentiments as 0 (negative) and 4 (positive). To streamline binary classification, these labels were relabeled to 0 and 1. Before training, a series of data cleaning procedures were performed. Irrelevant columns such as tweet IDs, timestamps, and user metadata were removed to reduce noise. The distribution of sentiment labels was visualized using a pie chart to gain insights into class balance. The tweets were then cleaned to remove unwanted components such as URLs, user mentions (e.g., @username), hashtags, and special characters. Following the cleaning process, the text data was tokenized and padded to standardize input length, ensuring compatibility with the input requirements of deep learning models.

Three deep learning models were implemented and trained for sentiment classification: **LSTM (Long Short-Term Memory)**, **CNN (Convolutional Neural Network)**, and **DistilBERT**, a transformer-based model. The LSTM model was employed to capture long-range dependencies in sequential data, while the CNN model focused on extracting local patterns from the text. The DistilBERT model, a lighter and faster version of BERT, was fine-tuned using the Hugging Face Transformers library for sentiment classification tasks.

The training configurations for the models are summarized in **Table 1** below:

Table 1. Training Configuration for Implemented Models

Model	Epochs	Batch Size	Optimizer	Early Stopping
LSTM	10	128	Adam	Yes
CNN	10	128	Adam	Yes
DistilBERT	4	—	Adam	Patience = 1

Model performance was evaluated using two key metrics: **accuracy** and **loss**. Accuracy was used to measure the proportion of correctly predicted samples, providing an overall measure of model correctness. Cross-entropy loss was employed to quantify the difference between predicted and actual sentiment labels, serving as a learning objective for the classification models.

All model development and experimentation were conducted using **Google Colaboratory**, a cloud-based Python development environment. The project made extensive use of libraries such as **TensorFlow** and **Keras** for implementing the LSTM and CNN models, and the **Hugging Face Transformers** library for fine-tuning DistilBERT. Additional Python libraries including **NumPy**, **Pandas**, **Matplotlib**, **Seaborn**, and **Scikit-learn** were used for data preprocessing, visualization, and evaluation tasks.

The full implementation and source code are publicly available at the following GitHub repository:
https://github.com/Maiaryanraj/Sentiment-Analysis.git

5. Results and Discussion

This section presented a comparison of the performance of three different deep learning architectures: LSTM, CNN and DistilBERT, on the task of sentiment analysis on the Sentiment140 dataset. The models were compared based on accuracy, precision, recall, F1-score, and confusion matrices to measure its effectiveness at classification.

5.1 Overall LSTM, CNN, and DistilBERT Model Performance

The results from the classification accuracy for each model were presented in Table 1.

Table 1: Model Accuracy Comparison

Model	**Accuracy**
LSTM	82.56%
CNN	81.93%
DistilBERT	80.00%

The results demonstrate that the LSTM model performed better than the other models with the highest accuracy of 82.56%. The LSTM model demonstrates an ability to capture temporal dependencies in sequential text data. This ability is important as it gives it insight into the surrounding text and its contextual sentiment. Moreover, the CNN model also performed competitively with an accuracy of 81.93%. The static nature of CNN architecture may leave out evidence of historical local n-gram features, but it does use convolutional filters to apply n-gram level features to extract the most valuable. Similarly, the DistilBERT model is a transformer-based model that was pre-trained on a large corpus of text and achieved slightly lower accuracy of 80.00%. However it was still able to achieve reasonable performance given that this model was fine-tuned for a short amount of time (number of epochs = 4), and also continued to show strong generalization.

5.2 Comprehensive Evaluation Metrics

All 3 models were evaluated on a larger set of metrics including precision, recall, and F1-score using classification reports. The classification reports allowed us to evaluate the models above and beyond accuracy because they provide deeper understanding of performance. It is also crucial to point out that we plotted confusion matrices to represent the true positives, false positives, false negatives, and true negatives for each model.

The LSTM and CNN models were well balanced with high recall and precision across the positive as well as the negative sentiment classes. DistilBERT showed slightly lower recall for the minority class but additional monitoring is needed to ensure it can perform better with further fine tuning extant training or not.

5.3 Visual Data

As mentioned above, to further support the quantitative data, we provided visual data for the models results to include both confusion matrices and performance comparison visualizations. These provided the audience with perspective of the strengths and weaknesses of each model's performance in the context of sentiment classification and to easily develop an understanding of misclassifications.

5.4Study Overview

Overall, the experiments establish that all 3 models can perform sentiment analysis on social media text data, using our defined problem, at fairly good accuracy. The LSTM model performed the best, closely followed by CNN and DistilBERT. DistilBERT is an excellent tool, especially when tasked with minimal computational resources and/or labelled data amongst other processes in the analysis pipeline, but additional detail and thorough testing and fine-tuning material is needed to achieve maximum optimization.

6. Conclusion

In examining objects, this research set out to perform a comparative study of the three most widely used deep learning architectures for sentiment analysis of social media text with the Sentiment140 dataset—LSTM, CNN, and DistilBERT. Each model was implemented, trained, and evaluated according to standard preprocessing and training parameters to provide a fair comparison.

The results indicated:

- LSTM achieved an accuracy of 82.56% topping the comparison of state of the arts and demonstrating its ability to better capture sequential dependencies in text data.
- CNN trailed closely with an accuracy of 81.93% following suit of leveraging its ability to uncover local patterns with convolutional filters.
- Even with few epochs of fine-tuning DistilBERT achieved an accuracy of 80.00% contributing to the knowledge of the impact of transformer-based pre-trained models on performing sentiment classification tasks.

Overall, although LSTM produced the best results, it is recognized that each model would be beneficial, depending on the use case. CNN provided faster training with model robustness and less complexity, and DistilBERT provides the added benefit of transfer learning with less task-specific data required.

Aspects for Future Work

Future work could examine a variety of avenues:

Tuning large transformers (e.g. BERT, RoBERTa) for a better prediction performance.

Integrating more sophisticated preprocessing (e.g. part-of-speech tagging and named entity recognition).

Utilizing ensemble methods to combine models to take advantage of various model strengths. Broadening to multilingual sentiment analysis datasets.

This study has laid the groundwork for a strong comparison and will be useful in future work related to social media sentiment analysis.

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/