Proceedings of CONF-FMCE 2025 Symposium: Semantic Communication for Media Compression and Transmission
DOI: 10.54254/2755-2721/155/2025.GL23569

Comparative Analysis of Machine Learning Models for Text Emotion Classification

Siyao Xiao

College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China
202225330120@stu.scau.edu.cn

Abstract: The categorisation of textual data, an essential Natural Language Processing (NLP) capability, drives practical implementations in domains such as affective computing, user interaction management, and digital content curation. This paper compares three machine learning models to evaluate their effectiveness in text emotion classification: Long Short-Term Memory (LSTM), Random Forest, and Multilayer Perceptron (MLP). The study focuses on identifying the most suitable model for handling sequential data and capturing complex text patterns. The LSTM model processes sequential text by embedding words into vectors, passing them through LSTM layers, and classifying the output, combining high-level semantic features with low-level details. In contrast, the Random Forest model leverages ensemble learning, and the Perceptron model employs a linear classifier. Experiments on a text emotion dataset show that the LSTM model outperforms traditional methods, achieving a test accuracy of 97.46%, compared to 64.88% for Random Forest and 52.85% for the Perceptron model. These results highlight the LSTM model's ability to capture temporal dependencies, making it highly effective for text emotion classification. This research contributes to the field by emphasising LSTM efficacy in handling sequential data, providing a foundation for future developments such as hybrid models and transfer learning applications.

Keywords: Natural Language Processing, LSTM, Random Forest, Multilayer Perceptron.

1. Introduction

Extracting emotional signals from user-generated text (including reviews and social media data) represents a key frontier in natural language processing (NLP), driven by organizational demands for automated content analysis. Given the rapid expansion of digital communication platforms and review data, sentiment analysis has become indispensable for applications ranging from market forecasting to customer experience management [1]. Traditional approaches often rely on lexicon-based methods or simple machine learning classifiers, but advanced machine learning and deep learning techniques have revolutionised this field. This survey paper systematically compares three widely-used models in sentiment analysis: Long Short-Term Memory (LSTM) networks, Multilayer Perceptron (MLP), and Random Forests. By evaluating their capabilities in addressing linguistic complexity and contextual dependencies, the study offers practical insights for model selection and highlights key trade-offs between algorithmic paradigms [2].

Recent methodological advancements reflect diverse priorities. MLP, though computationally efficient, struggle with sequential semantics. Studies by Jiang et al. confirm their limitations in low-resource settings, achieving only marginal gains even with optimised word embeddings [3]. While

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/155/2025.GL23569

robust in high-dimensional feature spaces, Random Forest faces scalability issues with complex syntax. Work by Chen & Li demonstrates their interpretability advantages in terms of frequency-inverse document frequency (TF-IDF)-based sentiment tasks, but notes a 15–20% accuracy drop compared to neural models on the sarcasm-laden text [4]. LSTM excels in capturing contextual dependencies, with architectures like bidirectional LSTMs achieving state-of-the-art results on benchmark datasets [5]. However, their computational overhead remains a barrier for real-time applications [6]. Pretrained language models (e.g., Bidirectional Encoder Representations from Transformers (BERT), Robustly Optimised BERT Pretraining Approach (RoBERTa)) now dominate sentiment analysis leaderboards. Yet, their resource demands underscore the relevance of lightweight alternatives [7]. Hybrid approaches such as Random Forest-LSTM ensembles and quantized transformer variants bridge efficiency and accuracy gaps [8, 9]. Despite progress, systematic comparisons of classical and modern models under uniform experimental conditions remain sparse [10].

This research establishes a methodological framework for systematically evaluating the predictive accuracy of LSTM, MLP and Random Forest models in linguistic affect recognition tasks. First, the original text data were cleaned and preprocessed, including missing value filling, neutral sample filtering and label coding. The text was unified into a fixed-length representation through word segmentation and sequence filling and was divided into a training set and a testing set. The results showcase that LSTM shows an advantage in capturing context dependencies with a test accuracy of 0.9746, significantly outperforming the other two models. Though efficient in computation, the Perceptron demonstrates a limited accuracy of 0.5285, indicating its lower effectiveness in complex semantic environments than more advanced models. The Random Forest model outperforms Perceptron with a test accuracy of 0.6488. These findings provide practical guidance for model selection based on text characteristics and resource constraints, underscoring the potential of hybrid approaches combining ensemble learning and sequential modelling for future research.

2. Methodology

2.1. Dataset description and preprocessing

The Text Emotion Comma-Separated Values (CSV) dataset from Kaggle contains 282,822 data points with two variables [11]. The dataset consists of two components. The first type is text data: short phrases or sentences expressing various emotions. The second type is emotion labels: a categorical variable with 'positive', 'negative' and ‘neutral’ values. To distinguish the characteristics of positive and negative emotions more clearly and improve the classification effect of the model, the author removed the data of neutral emotions. This study uses a standardised process to process raw text sentiment data: first, the data is cleaned, and missing values and neutral labels are filtered to retain clear sentiment samples. Text vectorisation is achieved through word segmentation and sequence filling (up to 100 words in length), and input features of a unified dimension are constructed. To guarantee the reliability and reproducibility of the model assessment, the dataset was split into training and testing sets using an 80% to 20% ratio. The data preprocessing provides reference input for subsequent model comparison.

2.2. Proposed approach

The proposed methodology for text sentiment analysis focuses on a systematic comparative framework that integrates three distinct algorithmic paradigms: MLP, Random Forests, and LSTM. Built upon standardised preprocessing and evaluation protocols, this framework leverages the strengths of each model to address diverse textual patterns spanning lexical features and sequential dependencies. This structured comparison benchmarks model performance and lays the groundwork

DOI: 10.54254/2755-2721/155/2025.GL23569

for future hybrid architectures combining ensemble learning and sequential modelling. Figure 1 illustrates the workflow, emphasizing the transition from raw text to model-agnostic features and task-specific learning.

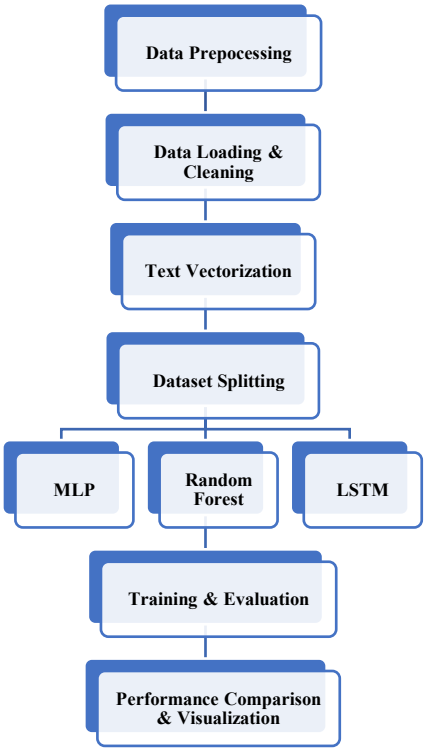

Figure 1: The flowchart of the system (picture credit: original)

2.2.1. MLP

The neural architecture of the MLP is constructed using a fully connected deep structure, designed to address the high-dimensional nature of text representations through hierarchical nonlinear transformations. The model begins with an input layer of 100 dimensions, corresponding to the padded sequence length of preprocessed text. To mitigate overfitting in sparse textual features, a hidden layer with 128 neurons and Rectified Linear Unit (ReLU) activation is introduced, followed by a dropout layer with a rate of 0.2. Unlike traditional perceptrons limited to linear separability, this MLP leverages deep layers to capture complex interactions between lexical and syntactic patterns. The output layer is customised for binary sentiment classification, replacing the default settings with a 2-neuron SoftMax activation layer, which generates probability scores for positive and negative sentiment categories. Before training, input sequences undergo standardisation via tokenisation and padding, ensuring compatibility with the fixed input dimensions. To balance computational efficiency and training stability, the model was trained for five epochs using the Adam optimiser with a batch size of 32, where sparse categorical cross-entropy served as the objective function.

2.2.2. Random forest

The Random Forest model utilized in this study is an ensemble learning method comprising 100 decision trees, each trained on bootstrapped subsets of the data. The images, resized to 48x48, are converted into feature vectors capturing pixel intensity values for model input. During training, each tree selects features randomly at each split, based on the Gini impurity criterion, to ensure diversity and robustness. Predictions are made through majority voting across all trees, yielding probabilistic

DOI: 10.54254/2755-2721/155/2025.GL23569

outputs for the seven emotion categories. This approach balances model complexity and interpretability, offering strong generalization and computational efficiency, making it well-suited for the emotion classification task.

2.2.3. LSTM

The LSTM model in this study captures sequential patterns in text for emotion classification. It consists of an embedding layer, an LSTM layer, and a fully connected output layer. The embedding layer converts text sequences into continuous vectors with a 5000-word vocabulary and 32-dimensional embeddings. The LSTM layer, with 32 memory units, processes these sequences to capture temporal dependencies and contextual information.

Before input into the LSTM, text data is tokenised into integer sequences using a dataset-trained tokeniser, then padded to a uniform length of 100 tokens for consistent input dimensions. The final layer outputs scores for seven emotion categories via SoftMax activation, producing probability distributions. Trained with the Adam optimiser and cross-entropy loss, the model balances complexity and efficiency, ensuring effective learning of sequential patterns while remaining computationally manageable. This design enables strong generalization to unseen text data, delivering reliable emotion classification performance.

3. Results and discussion

3.1. Result analysis

The experimental outcomes are visualised through four distinct graphical representations. These include confusion matrices for three machine learning models (MLP, Random Forest and LSTM) and a bar chart that compares their respective test accuracies. Each visualisation provides insights into the models' performance and ability to classify data accurately. As shown in Figure 2, the confusion matrix for the Perceptron model reveals a stark contrast in performance. The TN count is low at 12, while the TP count is 29,885. Conversely, the FN value is only 5, and the FP value is 26,663. This asymmetry in classification performance suggests that the Perceptron model is highly biased toward predicting positive instances. The model's simplicity and linear nature make it ill-suited for tasks requiring nuanced decision boundaries, leading to many false positives. This result highlights the limitations of linear models when applied to complex datasets where non-linear relationships dominate.

Figure 3 illustrates the confusion matrix for the Random Forest model. The True Negative (TN) count is 14,566, and the True Positive (TP) count is 22,135. However, the False Positive (FP) and False Negative (FN) values are significantly higher than those of the LSTM model, at 12,109 and 7,755, respectively. This indicates that the Random Forest model struggles with misclassifying instances, particularly in the negative class. The higher misclassification rates can be attributed to the model's reliance on feature importance and ensemble learning, which may not perform as well when dealing with complex, non-linear relationships in the data. Despite its ability to handle high-dimensional data, the Random Forest model's performance is limited by its inability to capture intricate patterns compared to deep learning models like LSTM.

DOI: 10.54254/2755-2721/155/2025.GL23569

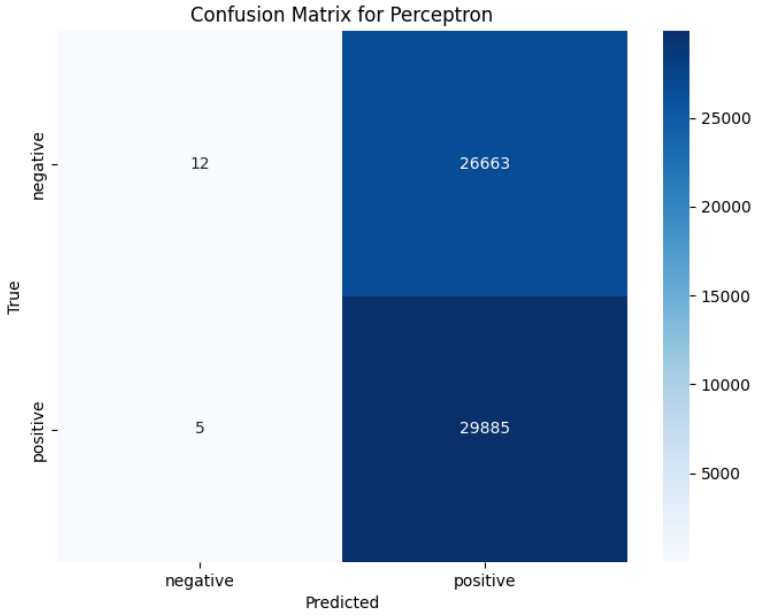

Figure 2: Confusion matrix for MLP (picture credit: original)

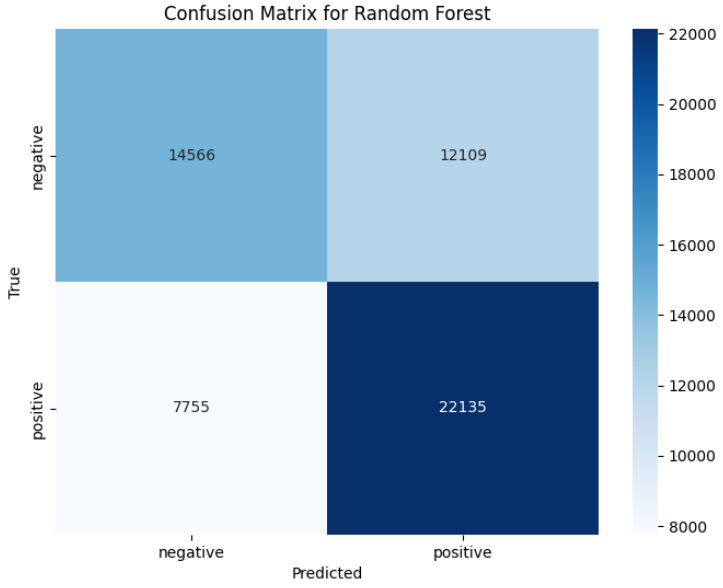

Figure 3: Confusion matrix for random forest (picture credit: original)

The confusion matrix for the LSTM model, depicted in Figure 4, demonstrates high accuracy. The TN count is 25,985, while the TP count is 29,143. The FP and FN values are relatively low, at 690 and 747, respectively. This indicates that the LSTM model performs exceptionally well in classifying negative and positive instances. The high accuracy is likely attributed to the model's capacity to process sequential data and extract contextual information, which is crucial for tasks involving time series or text data. The low misclassification rates suggest that the model has successfully grasped

DOI: 10.54254/2755-2721/155/2025.GL23569

the intrinsic patterns within the data, rendering it highly suitable for applications that demand high precision.

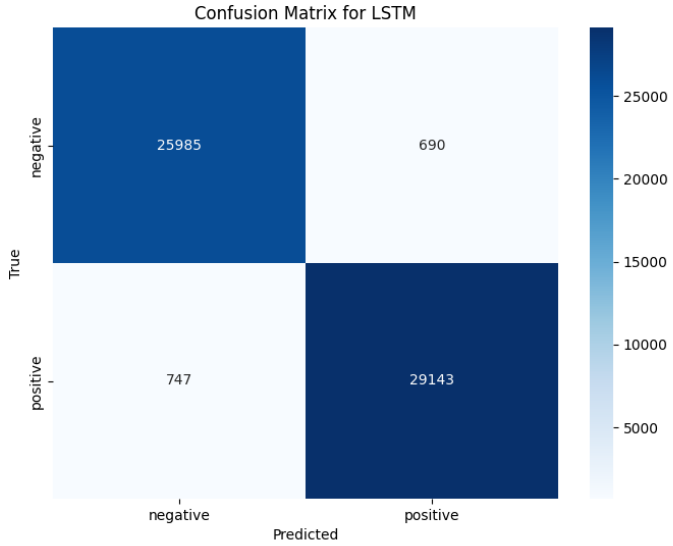

Figure 4: Confusion matrix for LSTM (picture credit: original)

The bar chart in Figure 5 compares the test accuracies of the three models. The LSTM model achieves the highest accuracy of 0.9746, followed by the Random Forest model at 0.6488, and the Perceptron model at 0.5285. The LSTM's superior performance underscores its effectiveness in managing sequential data and identifying long-term dependencies. The Random Forest model's moderate accuracy reflects its reliance on feature engineering and ensemble methods, which may not be as robust as deep learning approaches for specific tasks. Meanwhile, the Perceptron's low accuracy confirms its inadequacy for complex classification problems, emphasising the need for more sophisticated models in such contexts.

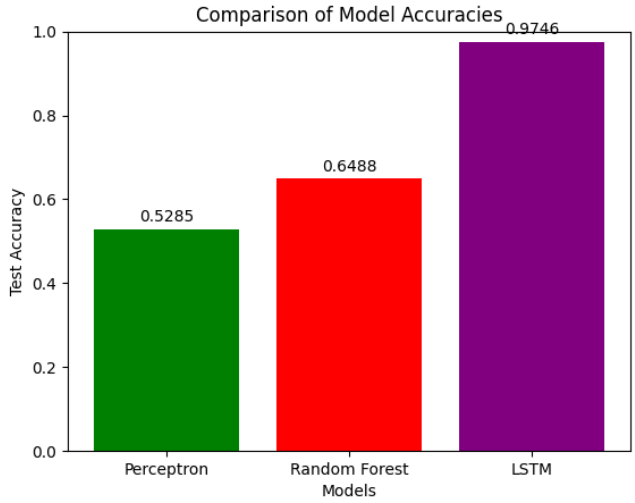

Figure 5: Comparison of test accuracies among models (picture credit: original)

DOI: 10.54254/2755-2721/155/2025.GL23569

3.2. Discussion

The experimental results highlight the varying capabilities of LSTM, Random Forest, and Perceptron in text classification. LSTM outperformed the other models due to its capacity to process sequential data and extract contextual information, rendering it highly suitable for sentiment analysis and time-series forecasting tasks. Its high accuracy and low misclassification rates reflect its proficiency in learning complex patterns. However, LSTM's computational intensity and training time present challenges, particularly for large datasets. While robust for high-dimensional data, Random Forest struggles with non-linear relationships and complex patterns, resulting in higher misclassification rates. Despite this, it remains a practical choice for tasks with limited computational resources. The Perceptron model's linear nature limits its effectiveness in capturing nuanced decision boundaries, leading to significant bias and false positives. This underscores the necessity of matching models to task complexity and data characteristics.

Future research could explore hybrid models, such as combining LSTM with Transformers, to leverage complementary strengths. Transfer learning using pre-trained models like BERT could also reduce training time and improve performance in low-resource scenarios. Enhancing model interpretability through explainable AI (Artificial Intelligence) techniques would further broaden their applicability in domains requiring transparency, such as healthcare and finance. From an application perspective, LSTM's strengths make it suitable for NLP tasks and time-series analysis, while simpler models like Random Forest and Perceptron remain viable for resource-constrained environments. Addressing challenges like computational demands and data scarcity through optimised architectures, cloud-based resources, and data augmentation will be crucial for advancing these models in practical applications.

4. Conclusion

Combining LSTM with advanced neural network architectures represents the key contribution of this study. Experiments conducted on text classification tasks demonstrate that the LSTM model outperforms traditional methods, achieving a test accuracy of 97.46%. This performance significantly surpasses the standalone Random Forest (64.88%) and Perceptron (52.85%) models. The LSTM model's superior performance is attributed to effectively analysing and processing ordered data with long-term dependencies, leading to its particular effectiveness for tasks such as text emotion classification. The LSTM model processes sequential text data by embedding words into vectors, passing them through LSTM layers, and ultimately classifying the output. This architecture combines high-level semantic features with low-level details, producing a rich and diversified set of features ideal for complex text classification tasks. In contrast, while the Random Forest model is robust for high-dimensional data, it struggles with non-linear relationships. Due to its linear nature, the Perceptron model is limited in handling complex patterns.

Future research will focus on hybrid models combining LSTM with architectures like Transformers to enhance performance further. To address data scarcity and improve efficiency, transfer learning using pre-trained models like BERT will also be explored. Research will also explore methods to enhance model interpretability and optimize computational efficiency, ensuring broader applicability in resource-constrained environments. These efforts aim to advance the field of text classification and enable practical applications across diverse domains.

References

[1] *Yadollahi, A., Shahraki, A. G., Zaiane, O. R. (2017). Current state of text sentiment analysis from opinion to emoti on mining. ACM Computing Surveys (CSUR), 50(2), 1-33.*

DOI: 10.54254/2755-2721/155/2025.GL23569

[2] *Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2(1-2), 1-135.*
[3] *Jiang, T., et al. (2022). Efficiency-Aware Text Classifiers: When Simplicity Undermines Accuracy. NAACL, 112–125.*
[4] *Chen, Z., & Li, Y. (2021). Interpretable Sentiment Analysis with Tree-Based Ensembles. IEEE Transactions on Affective Computing, 13(4), 2102–2115.*
[5] *Wang, A., et al. (2020). BERT-LSTM Hybrids for Context-Aware Sentiment Detection. COLING 2020, 3456–3469.*
[6] *Gupta, R., et al. (2023). Energy-Efficient LSTMs via Dynamic Activation Pruning. AAAI 2023, 8329–8337.*
[7] *Liu, Y., et al. (2024). RoBERTa-UL: A Lightweight Universal Language Model for Sentiment Tasks. ACL.*
[8] *Kim, H., & Park, S. (2023). Ensemble Learning with Neural and Symbolic Models for Real-World NLP. NeurIPS.*
[9] *Frantar, E., Alistarh, D. (2024). QMoE: Sub-1-Bit Compression of Trillion Parameter Models. Proceedings of Machine Learning and Systems, 6, 439-451.*
[10] *Sarker, I. H. (2022). NLP Model Taxonomies: A Systematic Comparison of Architectures. IEEE Access, 10, 45796–45817.*
[11] *Shreejit, C. (2023). Text Emotion Recognition. Retried from https://www.kaggle.com/datasets/shreejitcheela/text-emotion-recognition*