# CSE587 Mid-term Project: An LSTM-Based Emotion detector

Huanshu Zhang
Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16802, USA
hpz5226@psu.edu

## Abstract

*Emotion recognition from textual data plays a crucial role in enhancing human-computer interaction, sentiment analysis, and psychological assessments. This paper presents a deep learning approach utilizing Long Short-Term Memory (LSTM) networks with pre-trained Word2Vec embeddings for emotion detection in text. The dataset is curated using a distant supervision approach, leveraging emotion-related hashtags from social media to generate labeled training data. The proposed model is trained and evaluated on a well-curated emotion dataset, achieving a test accuracy of 93.49% and an area under the ROC curve (AUC) of 1.00 for each emotion category. The findings contribute to advancing emotion recognition research and improving automated sentiment analysis applications.*

## 1. Introduction

Understanding human emotions is an essential aspect of communication, shaping the way individuals interact with one another. Emotion recognition, particularly through textual data, enables machines to gauge the emotional states of individuals, enhancing applications in human-computer interaction[1], sentiment analysis, and psychological assessments[2]. Humans possess a natural ability to infer emotions from text, even when conveyed with minimal context. This ability allows for nuanced interpretations of sentiments, capturing complex feelings embedded within language structures. In computational settings, however, the challenge lies in designing models that can effectively mimic this human-like understanding, particularly when dealing with informal language, slang, and variations in expressions[2]. The integration of deep learning models with word embeddings provides a promising approach to tackling this challenge, ensuring robust and context-aware emotion detection from textual data[3].

Recent advancements in Natural Language Processing (NLP) have made significant strides in recognizing emotions through various computational techniques[4]. Traditional models rely heavily on handcrafted lexicons and statistical approaches, which often fail to generalize well in diverse textual styles. Deep learning architectures, such as Long Short-Term Memory (LSTM) networks, have shown remarkable success in capturing long-range dependencies within text sequences, making them suitable for tasks involving sequential data[5], [6]. When coupled with word embeddings such as Word2Vec, these models learn rich semantic representations of words, allowing for a deeper understanding of contextual relationships.

Here, we present an LSTM-based network with Word2Vec embeddings to achieve emotion detection in text. The model is trained on a well-curated emotion dataset and evaluated based on various performance metrics, with the area under the ROC curve (AUC) achieving a near-perfect score across all six emotion categories. The results indicate that the proposed model effectively captures emotional nuances, offering valuable insights into the interplay between deep learning architecture and linguistic representations.

## 2. Methods

### 2.1. Dataset Curation

The dataset used in this study is derived from a distant supervision approach using social media text annotated with hashtags as weak labels, following the methodology described in the CARER dataset[7]. Specifically, the dataset consists of tweets collected using predefined hashtags that strongly correlate with emotional expressions. These hashtags serve as weak labels, allowing the collection of large-scale emotion-annotated text data without requiring manual annotation.

To construct the dataset, an initial set of 339 emotion-related hashtags was identified, corresponding to eight emotion categories: sadness, joy, love, anger, fear, surprise, trust, and anticipation. The dataset was collected via the Twitter API, with each tweet labeled based on the hashtag it contained. To ensure higher data quality, only tweets where the emotional hashtag appeared at the end of the text were retained, minimizing the likelihood of mislabeling due to context shifts.

Preprocessing steps were applied to normalize the text,

including lowercasing, tokenization, and the removal of URLs, user mentions, and special characters. Additionally, the dataset underwent a filtering process where noisy or ambiguous instances were discarded. The final dataset was randomly split into training (80%) and test (20%) set in the code, ensuring that the model was trained on a diverse yet representative sample of emotion-labeled text.

### 2.2. Word2Vec and LSTM Networks

The Word2Vec model used in this study is the pre-trained "word2vec-google-news-300" model, which is widely recognized in the NLP community for its effectiveness in capturing semantic relationships between words. This model was developed by Google and made available through Gensim API. The model was trained on the Google News dataset, a large-scale corpus comprising approximately 100 billion words extracted from Google News articles. The extensive dataset enables Word2Vec to learn nuanced linguistic patterns, semantic relationships, and syntactic structures that improve generalization across diverse text-based applications.

This Word2Vec employs a Skip-Gram approach with negative sampling, which allows it to predict contextual word representations effectively. The learned word representations are stored in an embedding matrix, which functions as a high-dimensional lookup table. The full pre-trained model includes around 3 million words and phrases, each mapped to a 300-dimensional vector, resulting in an embedding matrix with a shape of approximately 3,000,000-by-300.

For this task, the model loads only the vectors corresponding to words present in the specific vocabulary of the emotion recognition dataset. Regardless of the number of words used, each word's representation remains a fixed 300-dimensional vector, ensuring consistency in the input embeddings.

LSTM networks are a specialized type of recurrent neural network (RNN) designed to model long-term dependencies in sequential data. Unlike standard RNNs, LSTMs incorporate memory cells and gating mechanisms that regulate information flow, mitigating the vanishing gradient problem and enabling effective learning from longer sequences. This makes LSTMs particularly suitable for emotion classification tasks where contextual understanding of textual sequences is critical[8].

### 2.3. Network structure

The proposed network is specifically designed to capture long-term dependencies in sequential text data, making it well-suited for emotion classification. It begins with a pre-trained embedding layer that transforms words into their corresponding vector representations, allowing the model to understand semantic relationships. This is followed by two stacked LSTM layers, each comprising 128 hidden units, enabling the network to effectively process sequential patterns. To mitigate overfitting, a dropout rate of 0.5 is applied between these layers. A ReLU activation function is incorporated to introduce non-linearity, enhancing the model's ability to capture complex patterns. Finally, a fully connected dense layer maps the processed LSTM output to six distinct emotion categories, ensuring accurate classification.

### 2.4. Training Process

The model is trained using a categorical cross-entropy loss function, which is well-suited for multi-class classification tasks. The Adam optimizer, known for its adaptive learning rate adjustments and efficient convergence, is employed with a learning rate of 1e-3. The training is conducted over five epochs with a batch size of 32.

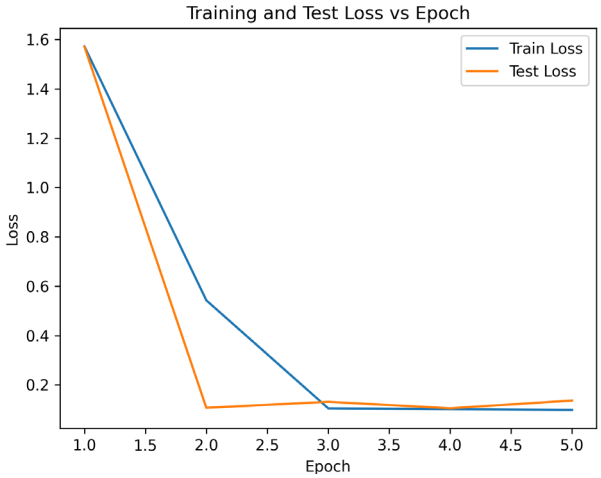


**Figure 1.** The training curve. The low training loss and test loss indicate a well convergence.

The dataset is fed into the network in batches, and gradient updates are performed iteratively. Model performance is monitored using validation loss and accuracy. The training and test loss over the five epochs (Figure 1) shows a sharp decrease within the first two epochs, stabilizing near zero afterward. This suggests that the model converged quickly and effectively learned the emotion representations from the dataset. After 5 epochs, the training loss decreased to 0.0980 and the test loss decreased to 0.1364. The final test accuracy is 93.49%, indicating an accurate detection of emotion.

## 3. Results and Discussion

The evaluation of the proposed LSTM-based emotion detection model demonstrates its strong performance. The overall test accuracy achieved is 93.49%, indicating the model's effectiveness in classifying emotions in textual data. The ROC Curve (Figure 2) illustrates the near-perfect classification performance across all six emotion

categories. The area under the ROC curve (AUC) is 1.00 for each class, suggesting that the model achieves an ideal separation between classes with minimal misclassification.

From the confusion matrix (Figure 3), the best performing emotion is love, which exhibits an exceptionally high recall of 0.99, indicating that nearly all instances labeled as love in the dataset are correctly identified by the model. This suggests that expressions of love are well-captured by the Word2Vec embeddings and LSTM network. The worst performing emotion is fear, which has the lowest recall at 0.85. This may be due to the subtle and context-dependent nature of fear-related expressions, which often share features with anger or sadness.

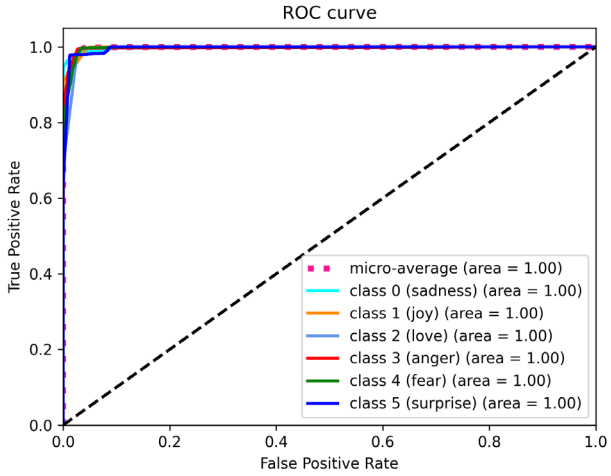


**Figure 2.** The ROC curve.

One of the most prominent observations from the confusion matrix is the misclassification of joy as love, with 2,037 instances falling into this category. The misclassification of joy as love can be attributed to several interrelated factors. One key reason is the semantic overlap between these emotions, as expressions of joy and love often sharing similar linguistic structures and vocabulary. Words conveying happiness, affection, and enthusiasm frequently appear in both contexts, making it challenging for the model to distinguish between them. Additionally, the ambiguity in data labeling, particularly in social media text, contributes to this issue. Posts labeled as joy may contain language commonly associated with love, leading to inconsistencies in the ground truth. Despite the model's architecture, it may still struggle to capture the subtle distinctions between these emotions. The absence of additional contextual cues, such as tone or discourse structure, further limits its ability to differentiate these categories with precision.

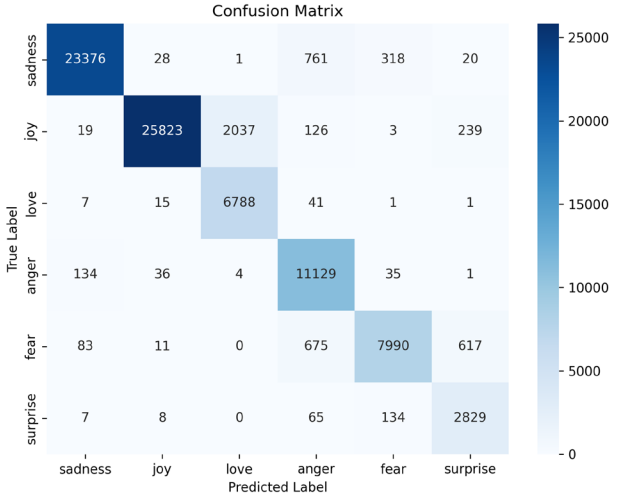


**Figure 3.** The confusion matrix.

## 4. Conclusion

This project provided insights into emotion detection using deep learning techniques. One of the key lessons learned was the importance of high-quality, well-annotated datasets. The confusion between similar emotions, such as joy and love, highlighted the need for refined labeling strategies and more context-aware embeddings. Additionally, working with pre-trained word embeddings like Word2Vec is beneficial, as it allowed the model to leverage prior linguistic knowledge, but it also showed limitations in handling nuanced emotional expressions that depend heavily on context.

Another takeaway was the effectiveness of LSTM networks in capturing sequential dependencies in textual data. However, while the model demonstrated high accuracy, the need for contextual embeddings like BERT or GPT became evident. Implementing an attention mechanism in future work may further improve classification by enabling the model to focus on the most informative parts of a sentence.

From a practical standpoint, optimizing hyper-parameters such as dropout rates and learning rates was crucial in preventing overfitting and ensuring model generalization. The training process reinforced the importance of early stopping and regularization techniques to maintain robustness in performance.

Finally, evaluating the model's performance using multiple metrics, including confusion matrices and ROC curves, was instrumental in identifying weaknesses and guiding improvements. The experience gained in handling multi-class classification and analyzing misclassifications will be valuable for future research in sentiment analysis and emotion-aware applications. Future work should focus on integrating contextual embeddings, refining dataset curation, and experimenting with hybrid architectures to

enhance the reliability of emotion detection models.

## References


[1] R. Cowie *et al.*, “Emotion recognition in human-computer interaction,” *IEEE Signal Process. Mag.*, vol. 18, no. 1, pp. 32–80, Jan. 2001, doi: 10.1109/79.911197.

[2] Computer Science and Engineering Department, Guru Gobind Singh Indraprastha University, New Delhi, India, A. Saxena, A. Khanna, and D. Gupta, “Emotion Recognition and Detection Methods: A Comprehensive Survey,” *J. Artif. Intell. Syst.*, vol. 2, no. 1, pp. 53–79, 2020, doi: 10.33969/AIS.2020.21005.

[3] Z. Zhu and K. Mao, “Knowledge-based BERT word embedding fine-tuning for emotion recognition,” *Neurocomputing*, vol. 552, p. 126488, Oct. 2023, doi: 10.1016/j.neucom.2023.126488.

[4] J. Deng and F. Ren, “A Survey of Textual Emotion Recognition and Its Challenges,” *IEEE Trans. Affect. Comput.*, vol. 14, no. 1, pp. 49–67, Jan. 2023, doi: 10.1109/TAFFC.2021.3053275.

[5] L. Chao, J. Tao, M. Yang, Y. Li, and Z. Wen, “Long Short Term Memory Recurrent Neural Network based Multimodal Dimensional Emotion Recognition,” in *Proceedings of the 5th International Workshop on Audio/Visual Emotion Challenge*, Brisbane Australia: ACM, Oct. 2015, pp. 65–72. doi: 10.1145/2808196.2811634.

[6] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, “Spatial–Temporal Recurrent Neural Network for Emotion Recognition,” *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019, doi: 10.1109/TCYB.2017.2788081.

[7] E. Saravia, H.-C. T. Liu, Y.-H. Huang, J. Wu, and Y.-S. Chen, “CARER: Contextualized Affect Representations for Emotion Recognition,” in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Brussels, Belgium: Association for Computational Linguistics, 2018, pp. 3687–3697. doi: 10.18653/v1/D18-1404.

[8] Y. Yu, X. Si, C. Hu, and J. Zhang, “A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures,” *Neural Comput.*, vol. 31, no. 7, pp. 1235–1270, Jul. 2019, doi: 10.1162/neco_a_01199.