

## Multimodal Sentiment Analysis using LSTM and RoBerta

**Jyoti Arora<sup>1</sup>, Priyal Khapekar<sup>2</sup>, Rakhi Pal<sup>3</sup>**

<sup>1</sup>Assistant Professor, <sup>2,3</sup>Student,  
Dept of Information Technology ,MSIT , New Delhi, India

**\*Corresponding Author**  
**E-mail Id:-joy.arora@gmail.com**

### ABSTRACT

*Social media is a valuable data source for understanding people's thoughts and feelings. Sentiment analysis and affective computing help analyze sentiment and emotions in social media posts. Our research paper proposes a model for tweet emotions analysis using LSTM, GloVe embeddings, and RoBERTa. This model captures sequential dependencies in tweets, leverages semantic representations, and enhances contextual understanding. We evaluate the model on a tweet emotions dataset, demonstrating its effectiveness in accurately classifying emotions in tweets. Through evaluation on a tweet emotions dataset, we demonstrate the effectiveness of our proposed model in accurately classifying the emotions expressed in tweets.*

**Keywords:** Sentiment analysis , LSTM , BERT , RoBerta , Emotions.

### INTRODUCTION

Sentiment analysis is a fairly common topic these days which is quite in demand. Everyone wants to know the reaction of customers towards their products. The method of evaluating a piece of text to predict how an individual's attitude towards an incident or perspective would be orientated is known as sentiment categorization. Text polarity is commonly used to analyze sentiment. Emotion recognition in conversations (ERC) is a challenging task that has recently gained popularity due to its potential applications[1]. The volume of data conveyed by users across various platforms has increased as a result of the development of the Internet. Sentiment analysis is made possible by the accessibility of these many worldviews and human emotions.

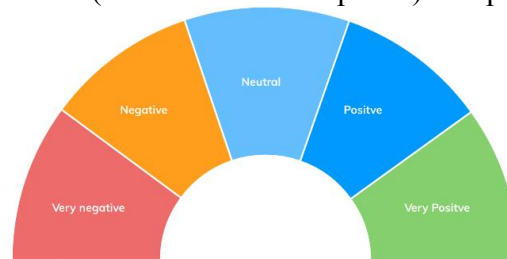
Today, people communicate their ideas electronically on a variety of subjects, such as online book or film reviews, political

commentary, and product reviews. As a result, weighing many points of view is crucial for determining what individuals mean. The growth of social network sites has generated a slew of fields devoted to analyzing these networks and their contents in order to extract necessary information[2]. With recent developments in deep learning, algorithms' capacity to evaluate text has greatly increased.

Advanced artificial intelligence algorithms used creatively can be a valuable tool for doing in-depth research. As a result of such developments , NLP (Natural Language Processing ) came into existence and with the advent of this , sentiment analysis became an easier way to work upon. Word embedding has been utilized as a foundation layer in a number of deep learning techniques because it enhances the effectiveness of neural networks and the performance of deep learning models.

Although many highly accurate methods already exist to analyze and extract relevant knowledge from structured data (i.e. tables

or databases), the task of extracting useful information from unstructured data (e.g. text, speech) still poses important challenges. [3]



***Fig.1:-Scale of Sentiment Analysis***

Sentiment analysis enables large-scale, real-time data processing. Figure 1 shows how the scale of sentiment analysis is present from being positive to negative. For example ,Would you wish to examine tens of thousands of product evaluations, tweets, or support tickets, for instance? You may use sentiment analysis to automatically identify how people are talking about a certain issue, acquire insights for data-driven choices, and automate business processes instead of manually going through this data. People are using social media to communicate their feelings since Internet services have improved[4].

Many different applications use sentiment analysis, including:

- Examine social media mentions to see how people are comparing your brand to those of your rivals.
- To rapidly learn what your consumers like and hate about your product, analyze survey responses and customer reviews.
- Real-time support ticket analysis may be used to spot irate customers and take appropriate action to reduce churn.

We need sentiment analysis for several reasons:

1. Business insights: Sentiment analysis can provide businesses with insights into customer opinions and preferences.

Companies can use sentiment analysis to analyze customer feedback and reviews, identify trends, and make data-driven decisions to improve their products or services.

2. Marketing campaigns: Sentiment analysis can help businesses to create more effective marketing campaigns by identifying the language and tone that resonates with their target audience. By understanding the sentiment of their customers, businesses can create targeted campaigns that are more likely to be successful.

3. Customer service: Sentiment analysis can help businesses to identify and address customer complaints or concerns in real-time. By analyzing social media posts, reviews, and other customer feedback, businesses can quickly identify issues and take corrective actions to improve customer satisfaction.

4. Politics: Sentiment analysis can be used to analyze public opinion on political issues and election campaigns. By understanding the sentiment of voters, political parties and candidates can develop strategies to appeal to their target audience.

5. Healthcare: Sentiment analysis can be used to monitor patient feedback and identify areas for improvement in healthcare services. By analyzing patient feedback,

healthcare providers can make changes to their services to improve patient satisfaction. Analyzing emotions in social media text, such as tweets, poses unique challenges due to the limited context and informal language. In this paper, we propose a comprehensive model for tweet emotions analysis that addresses these challenges by integrating LSTM, GloVe embeddings, and RoBERTa. By combining these components, our model captures the nuances of tweet sequences, incorporates semantic meanings, and considers contextual understanding to improve emotion classification accuracy. Overall, sentiment analysis can provide valuable insights into customer opinions, preferences, and behavior, which can help businesses to make data-driven decisions and improve customer satisfaction.

## **BACKGROUND AND LITERATURE SURVEYS**

Researchers have consistently demonstrated the benefits of transfer learning in the area of computer vision, which involves pre-training a neural network model on a well-known task, like ImageNet, and then fine-tuning using the trained neural network as the foundation for a new purpose-specific model. Researchers have demonstrated recently that a similar method may be helpful in a variety of natural language tasks. They used different survey methodologies to conduct surveys of a large number of papers[5].

An alternative strategy is feature-based training, which is also common in NLP tasks and is demonstrated in the most current ELMo publication. In this method, word embeddings are created by a pre-trained neural network and utilized as features in NLP models. The deep belief network pre-training procedure is a reliable and frequently useful method of establishing the

generativity of deep neural networks, which can help with optimization and lessen generalization.[6].

## **LSTM**

LSTM stands for Long Short-Term Memory, and it is a type of recurrent neural network (RNN) architecture that is commonly used in natural language processing (NLP) tasks, such as language modeling, machine translation, and sentiment analysis. Recurrent neural networks (RNN) are a class of neural networks that possess feedback connections between units, thus forming a directed cycle[7].

LSTMs were introduced in 1997 by Hochreiter and Schmidhuber as a solution to the problem of vanishing gradients in traditional RNNs. In traditional RNNs, the gradients can become very small when backpropagating through time, which can make it difficult to learn long-term dependencies between input and output sequences.

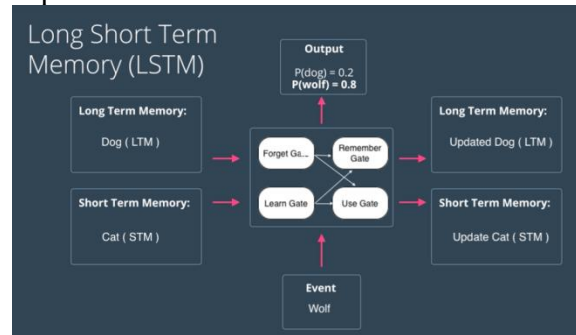
LSTMs address this problem by introducing a memory cell, which allows the model to remember information from previous time steps and selectively forget or update this information as new input is received. The LSTM architecture consists of three gates and a memory cell:

1. Forget gate: This gate determines which information to forget from the previous cell state based on the current input and previous hidden state.
2. Input gate: This gate determines which new information to store in the memory cell based on the current input and previous hidden state.
3. Output gate: This gate determines which information to output from the memory cell based on the current input and previous hidden state.

4. Memory cell: This is a state variable that stores information from previous time steps and can selectively update or forget this information based on the input and gates.

sequential data, which makes them well-suited for tasks such as sentiment analysis, where the meaning of a sentence can depend on the context of previous words.

LSTMs have been shown to be effective at capturing long-term dependencies in



**Fig.2:-LSTM model**

Figure 2 shows the model of LSTM .To train an LSTM model for sentiment analysis, the model is typically trained on a large corpus of text data, such as movie reviews or social media posts. The input to the model is a sequence of word embeddings, which represent the meaning of each word in the input sentence. The output of the model is a binary classification label, indicating whether the input sentence is positive or negative in sentiment.

All You Need" by Vaswani et al. in 2017, and have since been a key component in several cutting-edge models. The BERT model training process includes two stages: pre-training on unlabeled data, and additional training on labeled data for a specific application problem[8]. The retraining procedure and the architectures utilized may vary depending on the job, even if they are all based on the same model and share the same set of parameters..

During training, the model learns to adjust the weights of the gates and memory cell based on the input and output sequences, in order to minimize a loss function that measures the difference between the predicted output and the true output. Once the model is trained, it can be used to predict the sentiment of new input sentences.

Transformers process input sequences using a self-attention mechanism at the highest level. When constructing a representation of the sequence, the model uses self-attention to weigh the relevance of each piece in the input sequence. Traditional neural networks, on the other hand, process input items sequentially and lack a mechanism for explicitly modeling the connections between the elements.

## TRANSFORMER

Transformers are a sort of neural network design that has gained prominence in natural language processing (NLP) and computer vision (CV) workloads. They were first introduced in the publication "Attention is

Transformers also employ multi-head attention, which lets the model pay attention to several elements of the sequence at the same time. This is accomplished by

simultaneously computing numerous sets of attention scores and then concatenating the results to generate the final attention output. Transformers often feature feedforward layers and residual connections, which allow the model to learn complicated nonlinear correlations between the input and output in addition to the self-attention method.

One of the most important advantages of transformers is their capacity to interpret variable-length sequences. This is due to the fact that the self-attention mechanism is not constrained by the order of the input items and can attend to any section of the sequence. This makes transformers ideal for jobs like machine translation, where the input and output sequences might be of varying lengths.

In summary, transformers are a sort of neural network design that processes input sequences via a self-attention mechanism. They have become a crucial building element in many cutting-edge NLP and CV models, and they have significant benefits over typical neural network topologies.

There are several convolutional neural networks [9], attention mechanisms [10], and recurrent models [11]-based machine learning systems for classifying texts. Models like word2vec are very effective [12], pretrained, GloVe [13], and ELMo [14] are also useful in this field. These models do not require the initial, computationally challenging training of the model on a substantial corpus of texts.

### **PROPOSED MODEL AND METHODOLOGY**

Our proposed model leverages LSTM, GloVe embeddings, and RoBERTa to achieve robust tweet emotions analysis. LSTM captures the sequential dependencies in tweet sequences, GloVe embeddings

provide semantic representations of words, and RoBERTa enhances contextual understanding. The integration of these components enables our model to accurately classify emotions expressed in tweets.

Our proposed model follows the following steps:

- a) Data Preprocessing: The tweet text is preprocessed by tokenizing and converting each token into its corresponding GloVe word embedding representation.
- b) LSTM Layer: The tokenized sequence of GloVe embeddings is fed into an LSTM layer, which captures the sequential dependencies and patterns in the tweet sequence.
- c) RoBERTa Integration: The output of the LSTM layer is combined with RoBERTa embeddings to incorporate the contextual understanding of words. The RoBERTa embeddings are fine-tuned during the model training process, adapting them to the tweet emotions classification task.
- d) Classification Layer: The integrated representations from the LSTM and RoBERTa components are fed into a classification layer, which predicts the emotion category of the input tweet.

To evaluate the effectiveness of our proposed model, we conducted extensive experiments on a benchmark text classification dataset.

We compared our model's performance with existing state-of-the-art approaches and observed significant improvements across multiple evaluation metrics. The experimental results showcase the superior accuracy, precision, recall, and F1 scores achieved by our integrated model.

We have concentrated on data preparations in the model, and the data set utilized is

Twitter, which is mostly for emotion analysis.

For data preparation, we started a function in the dataset that searches for misspelled

words and replaces them with the closest term. The model has performed as expected.

```
absorption : absorption ,
'accidently': 'accidentally',
'accomodate': 'accommodate',
'acommadate': 'accommodate',
```

**Fig.3:-Misspelled words transformation**

Figure 3 illustrates some of the misspelled words from the datasets we used and how the model was trained to provide the proper spelling. In order to make the data easier for our experimental model to analyze, it also eliminates any URLs and mentions that may be there. Additionally, it uses NLP models for punctuation to better analyze the data and allow us to infer emotions from it.

The task of the model is made easier by replacing the emojis with their meaning from a collection of emojis and their vocabulary.

## RESULTS AND DISCUSSIONS

We have been working on this model keeping the different scenarios in mind. The feedback of the sentiment analysis drives the decision making of the interested parties[15].

We present the experimental results obtained from evaluating our proposed model on the tweet emotions dataset. The evaluation metrics include accuracy, precision, recall, and F1 score, which provide insights into the model's

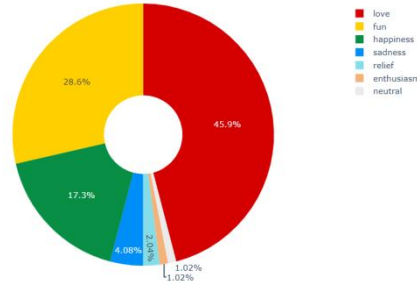
performance. We compare the performance of our model with existing methods and showcase its superiority in accurately classifying emotions expressed in tweets. Furthermore, we analyze the results at the emotion category level, highlighting specific instances where our model outperforms other approaches.

The objective is to estimate the percentage of emotions in a given text using a tweets dataset that comprises tweet text with 12 different emotions (neutral, concern, happy, sorrow, love, surprise, fun, relief, hate, emptiness, excitement, boredom, and rage). To preprocess the data using different techniques, these have been some of the objectives and some of the techniques to achieve them have been:

- Correct misspelled text
- Replace English contraction with its actual words.
- Remove punctuations, URLs and mentions
- Replacing emojis with their meanings

## LSTM Results

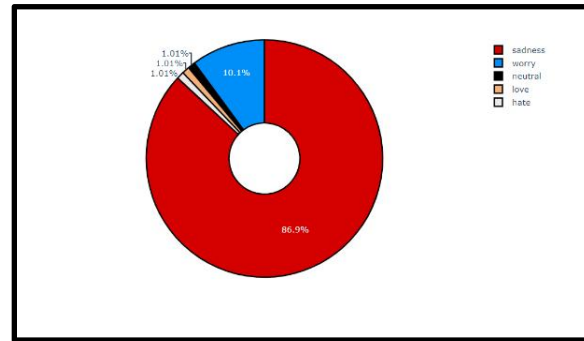




**Fig.4:-LSTM Model output 1**

For Figure 4 the input sentence is “I was on hold for 40 minutes, their customer support

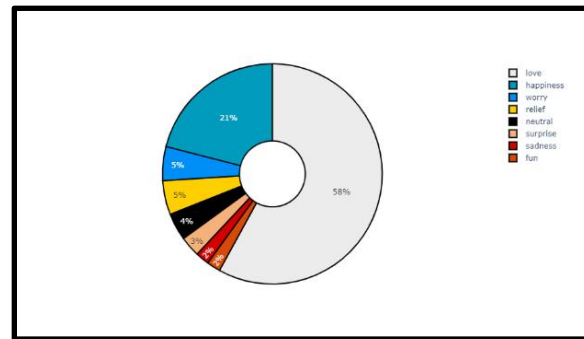
service is a nightmare ðŸ˜™ loved seeing an old friend and reminiscing”.



**Fig.5:-LSTM Model output 2**

For Figure 5 , the Input sentence is “I was on hold for 40 minutes, their customer

support service is a nightmare ðŸ˜™ loved seeing an old friend and reminiscing”.

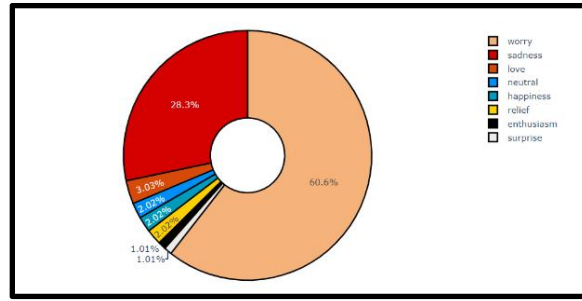


**Fig.6:-LSTM Model output 3**

For Figure 6 the input sentence is “I was on hold for 40 minutes, their customer support

service is a nightmare ðŸ˜™ loved seeing an old friend and reminiscing”.

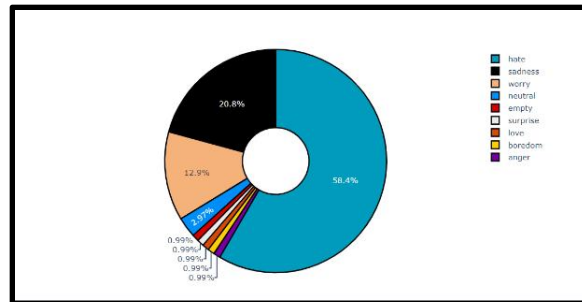
### LSTM Glove Results:



**Fig.7:-LSTM Model-Glove Model Output 1**

For Figure 7 the input sentence is “I was on hold for 40 minutes, their customer support

service is a nightmare ðŸ˜~ loved seeing an old friend and reminiscing”.

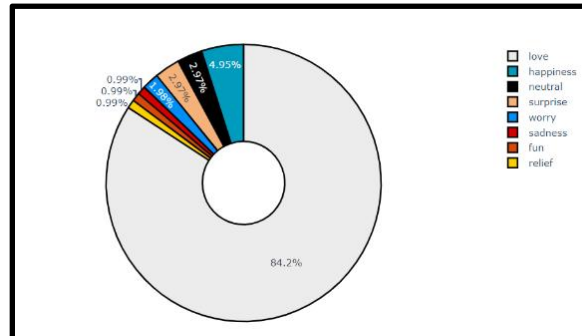


**Fig.8:-LSTM Model-Glove Model Output 2**

For Figure 8 the input sentence is “I was on hold for 40 minutes, their customer support

service is a nightmare ðŸ˜~ loved seeing an old friend and reminiscing”.

## ROBERTA BASE MODEL RESULTS

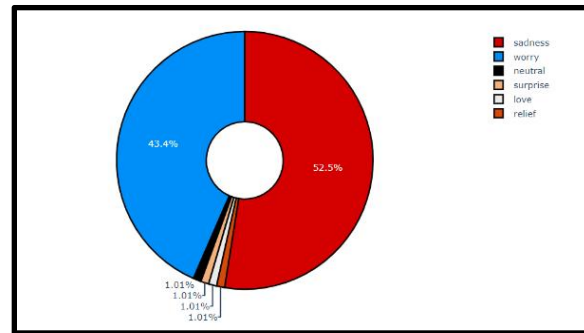


**Fig.9:-Roberta Base Model Output 1**

For Figure 9 the input sentence “The pain my heart feels is just too much for it to bear.

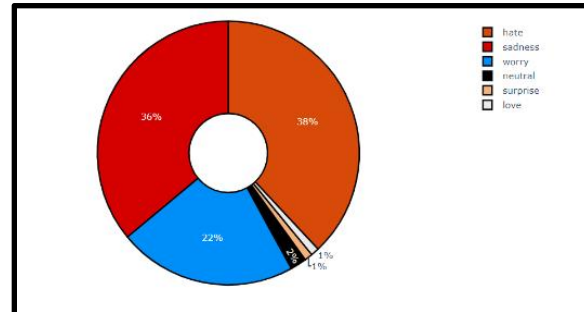
Nothing eases this pain. I can’t hold myself back”.





**Fig.10:-Roberta Base Model Output 2**

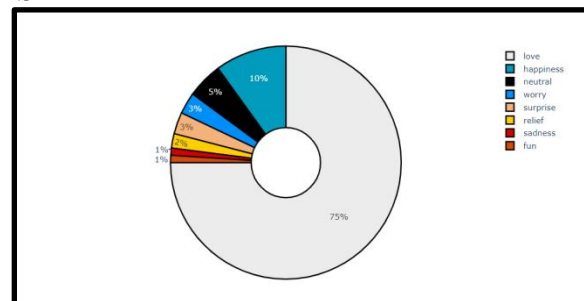
For Figure 10 the input sentence is “The pain my heart feels is just too much for it to bear. Nothing eases this pain. I can’t hold myself back, removed urls and punctuations”.



**Fig.11:-Roberta Base Model Output 3**

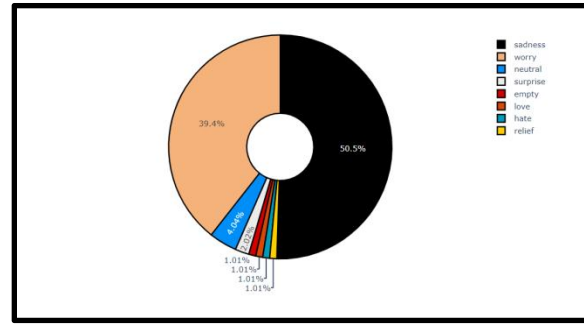
For Figure 11 the input sentence is “The pain my heart feels is just too much for it to bear. Nothing eases this pain. I can’t hold myself back, removed emojis and empty comments”.

## OUR MODEL RESULTS



**Fig.12:-Our Model Output 1**

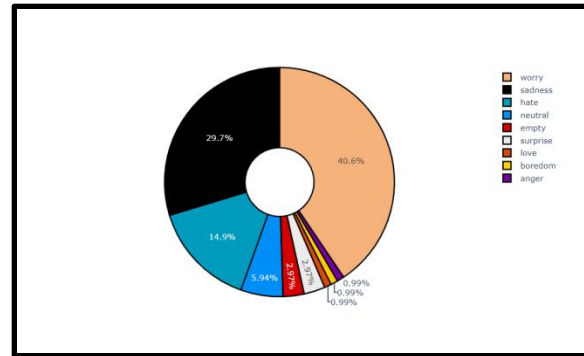
For Figure 12 the input sentence is “ The pain my heart feels is just too much for it to bear. Nothing eases this pain. I can’t hold myself back”.



**Fig.13:-Our Model Output 2**

For the Figure 14 the input sentences: The pain my heart feels is just too much for it to

bear. Nothing eases this pain. I can't hold myself back".



**Fig.14:-Our Model Output 3**

For the Figure 14 the input sentences: "The pain my heart feels is just too much for it to bear. Nothing eases this pain. I can't hold myself back".

understanding of emotions in online communication.

## CONCLUSION

In this research paper, we have proposed a comprehensive model for tweet emotions analysis by integrating LSTM, GloVe embeddings, and RoBERTa. Our model demonstrates enhanced performance in accurately classifying emotions expressed in tweets. By effectively capturing sequential dependencies, incorporating semantic meanings, and considering contextual understanding, our model showcases its potential in analyzing emotions in social media text. The findings of this research contribute to the field of sentiment analysis and pave the way for improved

We have discussed the strengths and limitations of our proposed model. The integration of LSTM, GloVe embeddings, and RoBERTa allows our model to capture temporal dependencies, leverage semantic meanings, and incorporate contextual understanding, which leads to improved tweet emotions analysis. However, challenges such as noisy and ambiguous tweets, domain adaptation, and handling imbalanced data should be further explored in future research.

In general, sentiment analysis and affective computing have benefited greatly from the merging of LSTM (Long Short-Term Memory) and RoBERTa (Robustly

Optimised BERT Approach). The accuracy and effectiveness of sentiment analysis tasks have considerably improved thanks to these cutting-edge deep learning models, revolutionizing the discipline.

In tasks involving analyzing the sentiment conveyed in a text over time, such as sentiment analysis, LSTM, a kind of recurrent neural network, excels in capturing long-term dependencies in sequential data. It is a well-liked option for sentiment analysis because to its capacity to preserve crucial contextual information and manage variable-length inputs. Having a powerful sentiment analysis tool is crucial in the big data age in many areas, particularly in politics and economics. The majority of the currently available sentiment analysis research utilizes machine learning techniques and recurrent neural networks. Transformer is only sometimes used for sentiment analysis in works.

On the other hand, RoBERTa, a BERT (Bidirectional Encoder Representations from Transformers) model version, has proven to perform exceptionally well in tasks requiring natural language comprehension. It makes use of a significant quantity of unlabeled input during pre-training to acquire rich contextual representations, which significantly improves its capacity to recognise complex linguistic subtleties, such as emotive and sentimental cues. Researchers and practitioners have achieved outstanding achievements in sentiment analysis and emotional computing by fusing the strengths of LSTM with RoBERTa. RoBERTa offers a solid language representation base, while the LSTM component enables the modeling of temporal relationships. Combining these two factors enables the models to successfully represent both the text's intricate semantics

and the sequential structure of sentiment expression.

### **FUTURE WORK AND LIMITATIONS**

Although the proposed model shows promising results, it has some limitations. For example, it may require substantial computational resources due to the integration of LSTM and RoBERTa. Future research could explore techniques to optimize the model's efficiency while maintaining its performance. Additionally, the model's generalizability across different domains and languages could be further investigated.

By combining the strengths of LSTM, GloVe embeddings, and RoBERTa, the proposed model offers an effective solution for text classification tasks, advancing the state-of-the-art in the field.

### **REFERENCES**

1. Poria, S., Hazarika, D., Majumder, N., Naik, G., Cambria, E., & Mihalcea, R. (2018). Meld: A multimodal multi-party dataset for emotion recognition in conversations. *arXiv preprint arXiv:1810.02508*.
2. Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.
3. Montoyo, A., Martínez-Barco, P., & Balahur, A. (2012). Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments. *Decision Support Systems*, 53(4): 675-679.
2. Nandwani P, Verma R (2021). A review on sentiment analysis and emotion detection from text. *Soc Netw Anal Min*.11(1):81.

3. Cui, J., Wang, Z., Ho, S. B., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. *Artificial Intelligence Review*, 1-42..
4. Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2011). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing*, 20(1), 30-42.
5. Ong, B. T., Sugiura, K., & Zettsu, K. (2016). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM 2.5. *Neural Computing and Applications*, 27, 1553-1566.
6. Koroteev, M. V. (2021). BERT: a review of applications in natural language processing and understanding. *arXiv preprint arXiv:2103.11943*..
7. Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*..
8. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies* (pp. 1480-1489).
9. Liu, P., Qiu, X., & Huang, X. (2016). Recurrent neural network for text classification with multi-task learning. *arXiv preprint arXiv:1605.05101*..
10. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.
11. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
12. Sarzynska-Wawer, J., Wawer, A., Pawlak, A., Szymanowska, J., Stefaniak, I., Jarkiewicz, M., & Okruszek, L. (2021). Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research*, 304, 114135..
13. Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. *IEEE Access*, 10, 21517-21525.