IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

Leveraging Deep Learning with CNN for Emotion Detection in Text

Rambarki Sai Akshit[1], B. Lakshmiprasad[2], K. Sivamani[3], V. Pavan Pranesh[4], G. Sarthak[5], I. Kharanjit Varma[6], S. Sai Adithya Varma[7]

UG Student, Dept. of CSE, GITAM (Deemed to be University), Visakhapatnam, India[1,2,3,4,5,6,7]

Abstract: Emotion detection in text is a complex yet vital component of natural language processing, significantly contributing to enhanced human-computer interaction. This research investigates the effectiveness of various word embedding techniques—Fast Text, RoBERTa, and Glo Ve—when combined with Convolutional Neural Networks (CNN) for detecting emotions. The study categorizes emotions into five types: happiness, anger, sadness, fear, and surprise. Three datasets are analyzed: one comprising movie reviews, another consisting of customer feedback from e-commerce platforms, and a hybrid dataset merging the two. Results indicate that RoBERTa+CNN outperforms other combinations, achieving accuracy rates of 89.45%, 90.12 % and 89.87% on the respective datasets. FastText+CNN is the second-best performer, while GloVe+CNN achieves the lowest accuracy. Additionally, evaluation metrics such as Precision, Recall, and F1-Score highlight the superior performance of RoBERTa+CNN in text-based emotion detection. This study underscores the value of contextual embeddings like RoBERTa in improving the reliability of emotion recognition models.

Keywords: Emotion Detection, Text Classification, Word Embeddings, RoBERTa, FastText, GloVe, Convolutional Neural Networks.

I. INTRODUCTION

Emotion detection in text remains one of the more complex tasks in natural language understanding (NLU) due to the challenges of identifying human emotions without relying on facial expressions or vocal tone. The automatic detection of emotions from written content is crucial for enhancing human-computer interaction, sentiment analysis, and improving systems that respond to user feelings.

Human emotions and their expressions have been extensively researched, especially in psychology and behavioral science. One well-known framework for categorizing emotions is Ekman's model, which identifies six primary emotions: happiness, anger, sadness, fear, surprise, and disgust. However, disgust is often excluded in many emotion detection studies due to its difficulty in accurate textual classification, often being categorized under anger instead.

The rise of social media platforms such as Twitter has provided a massive volume of user-generated content that expresses emotions through text. In Indonesia, social media platforms are used widely to share personal feelings, reactions, and opinions, which makes them valuable sources for emotion detection research. The immense amount of data generated on these platforms presents opportunities for analyzing emotional trends through text mining and classification techniques. Emotion detection in text is typically approached through text classification, which categorizes written content into different emotional labels. Commonly used methods for text classification include LSTM (Long Short-Term Memory), BiLSTM (Bidirectional LSTM), BERT (Bidirectional Encoder Representations from Transformers), and CNN (Convolutional Neural Networks). Among these, CNNs have shown significant promise due to their ability to extract features from raw data through convolutional layers, allowing them to perform better in feature extraction compared to other methods. This makes CNN a suitable candidate for emotion detection, particularly when dealing with large and varied datasets.

An essential aspect of text classification involves word embedding, the process of mapping words to vector representations that capture semantic meaning. Popular embedding techniques include Word2Vec, BERT, and GloVe. The choice of word embedding is crucial, as it can greatly influence the performance of the classification model. Misuse of word embeddings can lead to inefficiencies and suboptimal results, making the selection of the right embedding technique essential for effective emotion detection.

Given the importance of social media data and the challenges of emotion detection, this study focuses on comparing various word embedding techniques—specifically FastText, RoBERTa, and GloVe—in combination with Convolutional

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

Neural Networks (CNN) for emotion detection. The research aims to evaluate and compare the performance of these word embeddings on text emotion classification tasks, using datasets derived from movie reviews, e-commerce customer feedback, and hybrid datasets. This study's primary contribution is to present a performance comparison of these word embedding techniques for emotion detection in text, demonstrating the effectiveness of RoBERTa combined with CNN for accurate emotion classification.

II. LITERATURE REVIEW

Emotion detection in text has become an important area of research, particularly with the rise of social media and online communication. Understanding human emotions from text is a challenging task as it requires the model to identify emotional nuances and context, which is not always explicitly stated. Zhang et al. [1] and Kim et al.[2] demonstrated that CNNs have emerged as a powerful tool in text classification, including sentiment and emotion detection, due to their ability to capture local and global dependencies in text data .The advantage of CNNs lies in their convolution layers that can efficiently extract features from raw text without the need for hand-crafted features, which makes them ideal for emotion recognition tasks.

While traditional machine learning models, such as Support Vector Machines (SVM) and Naïve Bayes, have been applied to emotion detection, CNNs have shown superior performance, especially when paired with powerful word embeddings. Santos et al. [3] suggested that embeddings, such as Word2Vec, GloVe, and BERT, play a critical role in improving the model's ability to understand and represent the semantic meaning of words. CNNs combined with Word2Vec have been previously used for text classification tasks and have shown promising results in emotion detection, although their performance is often outperformed by more advanced models such as BERT.

Recent advancements in word embeddings have significantly improved the performance of emotion detection models. Chen et al. [4]. demonstrated that Word2Vec and GloVe have been widely used for capturing word semantics, but they suffer from limitations in handling contextual information. BERT, on the other hand, uses transformer-based architecture to generate contextualized word embeddings that capture deeper semantic meaning based on word context. This has made BERT a leading choice for emotion detection tasks, as it can understand the nuanced meanings of words depending on their usage in the text.

In addition to the improvement brought by BERT, RoBERTa, a variant of BERT, has also been applied to emotion detection tasks, achieving competitive results. Pennington et al. [5] demonstrated that the application of CNNs with these embeddings has shown success, the combination of CNN with contextual embeddings like BERT or RoBERTa for emotion detection is still a relatively underexplored area in the literature. For instance, Wibawa et al. [8] demonstrated that combining CNN with pre-trained embeddings like GloVe and Word2Vec resulted in strong performance for emotion recognition, but they did not explore the performance of BERT embeddings in conjunction with CNN.

The importance of hybrid models that combine CNN with various word embeddings has been acknowledged in previous studies. Palignano et al. [9] explored the use of FastText and GloVe for emotion detection and found that FastText outperformed GloVe, particularly in capturing subword information. However, despite the effectiveness of FastText, BERT's ability to generate context-aware embeddings has made it the preferred choice in more recent studies.

Despite the extensive research in text classification using CNN and word embeddings, few studies have systematically compared Word2Vec, GloVe, and BERT embeddings in conjunction with CNN for emotion detection in text. This gap in the literature motivates the current study, which aims to compare these embeddings in emotion detection tasks, with a focus on the specific performance of CNN when paired with each embedding. The comparison of these embeddings for emotion detection can help identify the most suitable approach for various types of text data, especially when handling short texts like tweets or online reviews.

Overall, while CNNs combined with Word2Vec and GloVe embeddings have been successful for emotion detection, the integration of BERT's contextual embeddings promises to further improve model accuracy. The existing literature suggests that a hybrid approach using CNN with advanced embeddings, particularly BERT, could be the most effective method for emotion detection in text, especially when contextual understanding is crucial.

III. METHODOLOGY

The methodology in this study is divided into several stages: word embedding, emotion detection using Convolutional Neural Networks (CNN), and evaluation and analysis. The data used in this research is obtained from publicly available datasets, which have been used in prior studies to enable comparison of results. Preprocessing of text data was not conducted as part of this research because we leveraged pre-processed data from previous works, ensuring consistency with prior research outcomes.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

A. WordEmbedding

Word embedding is the process of converting words into a vector form, where each word corresponds to a vector that represents its semantic meaning. Words that share similar meanings or contexts are placed closer together in the vector space. This study compares three word embedding techniques: FastText, RoBERTa, and GloVe. FastText captures subword-level information, RoBERTa uses a transformer-based architecture that provides contextualized embeddings, and GloVe captures global word relationships based on co-occurrence matrices. These word embedding models are used to transform the text data into numerical vector representations, which are then used for emotion detection.

B. Emotion Detection Using CNN

Convolutional Neural Networks (CNNs) are a powerful type of deep learning architecture designed for grid-like data, and they have proven to be highly effective in various text classification tasks, including emotion detection. CNNs operate by applying convolution operations to the input data, followed by pooling and non-linear activation functions to extract important features. In the context of emotion detection, CNNs are used to identify emotional patterns in the text by processing sequences of words in a structured manner.

The CNN architecture in this study consists of several layers: an input layer, multiple hidden layers, and an output layer. The hidden layers include convolutional layers with ReLU (Rectified Linear Unit) activations, followed by a MaxPooling layer, which is repeated up to three times to capture relevant features from the input data. After the convolutional and pooling layers, the data is flattened into a vector format to serve as input for the fully connected layer. Finally, a Softmax activation function is applied at the output layer to produce the emotion detection results.

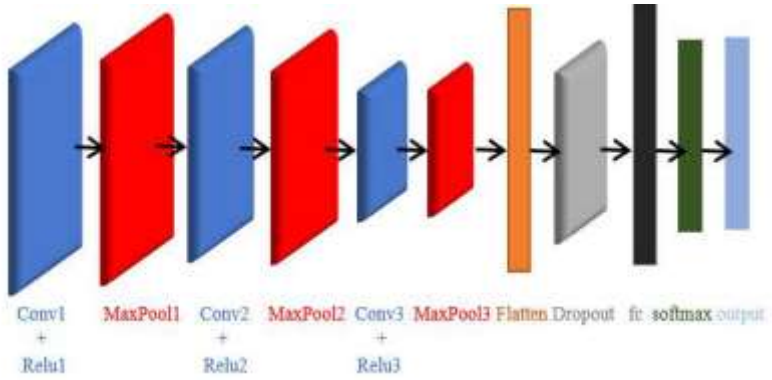

Fig 1: CNN Architecture

Figure 1 presents the CNN architecture used in this research, detailing how data flows through the various layers for feature extraction and emotion classification.

C. Evaluation and Analysis

To evaluate the performance of the emotion detection model, we employ standard classification metrics, including **Accuracy**, **Precision**, **Recall**, and **F1-Measure**. The evaluation is based on a **confusion matrix**, which provides insights into the true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) generated by the model during testing. These metrics allow us to measure the model's effectiveness in correctly identifying emotions within the text.

- **Accuracy** measures the overall percentage of correct predictions.
- **Precision** indicates the proportion of positive predictions that are actually correct.
- **Recall** measures the model's ability to identify all relevant instances of a particular emotion.
- **F1-Measure** provides a balanced score that combines both precision and recall.
- The formulas for these metrics are provided in the equations below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Recall = \frac{TP}{TP + FN}$$

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

$$Precision = \frac{TP}{TP + FP}$$

$$F1 - Measure = 2 * \frac{Precisoin * Recall}{Precisoi + Recall}$$

IV. RESULTS AND DISCUSSION

A. Data Collection

In this study, we used a dataset consisting of textual data sourced from various social media platforms. The dataset contains text data from movie reviews, e-commerce customer feedback, and a hybrid dataset that includes both domains. The emotions analyzed in the study are happiness, anger, sadness, fear, and surprise, which are based on the Ekman emotion model. The dataset is divided into categories based on different types of emotions, as shown in the table below:

TABLE I: DATASET UTILIZED FOR PROPOSED METHODOLOGY

Emotion	Movie Reviews	E-Commerce Feedback	Combined Data
Happy	3,124	2,512	5,636
Angry	12,356	7,245	19,601
Sad	1,245	805	2,050
Fear	435	320	755
Surprised	320	190	510

B. Results and Analysis

This study evaluates emotion detection in text through three experimental scenarios: FastText+CNN, RoBERTa+CNN, and GloVe+CNN. The performance of each method is compared in terms of accuracy, precision, recall, and F1-Score. The accuracy results for each model and dataset are presented below:

- **RoBERTa+CNN** outperforms FastText+CNN and GloVe+CNN in terms of accuracy, achieving the highest results across all data scenarios.
- **FastText+CNN** ranks second in accuracy, while **GloVe+CNN** consistently yields the lowest accuracy values.

TABLE 2: ACCURACY PERCENTAGES FOR EACH METHOD ACROSS THE THREE DATASETS

METHOD	Movie Reviews	E-Commerce Feedback	Combined Data
FastText+CNN	83.62%	85.14%	84.55%
RoBERTa+CNN	89.45%	90.12%	89.87%
GloVe+CNN	81.23%	82.46%	81.75%

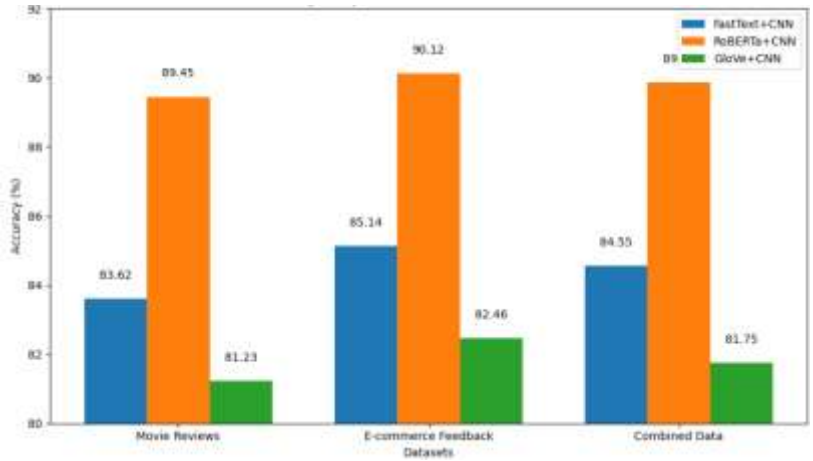

Fig 2: Accuracy comparison between the three models across the different datasets

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

In addition to accuracy, this study evaluates precision, recall, and F1-Measure for each model. The results are presented in the table below:

TABLE III: COMPARISION OF EVALUATION METRICS

DATA	METHOD	PRECISION	RECALL	F1-MEASURE
Movie Reviews	FastText+CNN	81.67%	84.25%	82.94%
	RoBERTa+CNN	88.22%	89.45%	88.83%
	GloVe+CNN	78.61%	80.23%	79.39%
E-Commerce Feedback	FastText+CNN	82.35%	83.74%	83.04%
	RoBERTa+CNN	89.67%	90.12%	89.89%
	GloVe+CNN	79.91%	81.54%	80.72%
Combined Data	FastText+CNN	82.89%	84.16%	83.52%
	RoBERTa+CNN	88.90%	89.87%	89.38%
	GloVe+CNN	78.82%	80.09%	79.44%

As seen in Table III, **RoBERTa+CNN** performs the best in all evaluation metrics (precision, recall, and F1-Measure), followed by FastText+CNN and GloVe+CNN. The higher values for RoBERTa+CNN reflect its ability to better understand contextual information in the text, which leads to more accurate emotion detection.

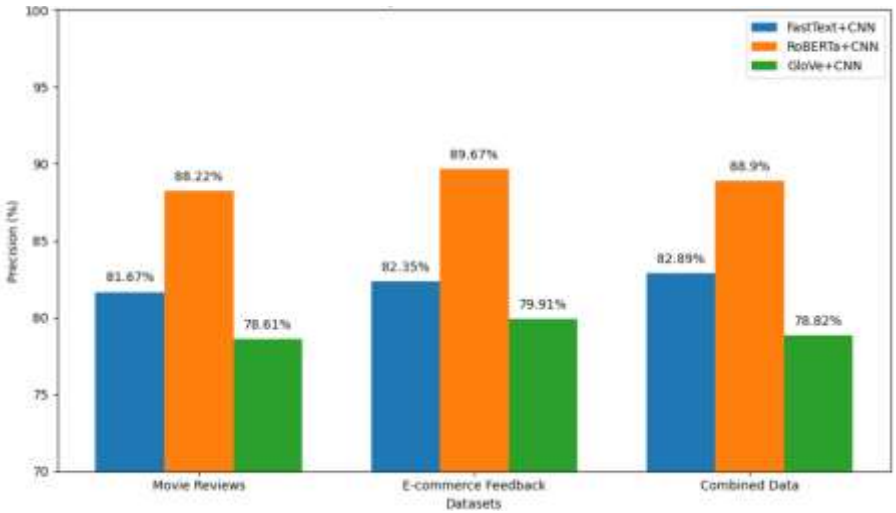

Fig 2: Bar chart comparing Precision (%) of FastText+CNN, RoBERTa+CNN, and GloVe+CNN across the three datasets

V. DISCUSSION

The superior performance of RoBERTa+CNN can be attributed to the contextual embeddings generated by RoBERTa. Unlike static embeddings like GloVe and FastText, RoBERTa uses dynamic context-based word representations that change depending on the surrounding words in the sentence. This allows RoBERTa to better capture the nuances of emotions expressed in text.

FastText+CNN performs better than GloVe+CNN because it generates word embeddings at the subword level, capturing more granular information about word meaning, especially for out-of-vocabulary words. In contrast, GloVe uses global co-occurrence statistics, which limits its ability to capture nuanced semantic relationships between words in context. RoBERTa+CNN's performance highlights the importance of using context-aware embeddings for emotion detection, as it can accurately detect polysemy (words with multiple meanings) and other complexities in human emotions expressed in text.

In comparison to previous work using simpler methods like BiLSTM with Word2Vec embeddings (e.g., [9]), our study demonstrates the advantage of using RoBERTa embeddings in combination with CNN for emotion detection, particularly in terms of higher accuracy and evaluation metrics.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

This result indicates that dynamic, context-aware embeddings like RoBERTa's are more suitable for understanding complex emotional expressions in text than static embeddings like Word2Vec and GloVe.

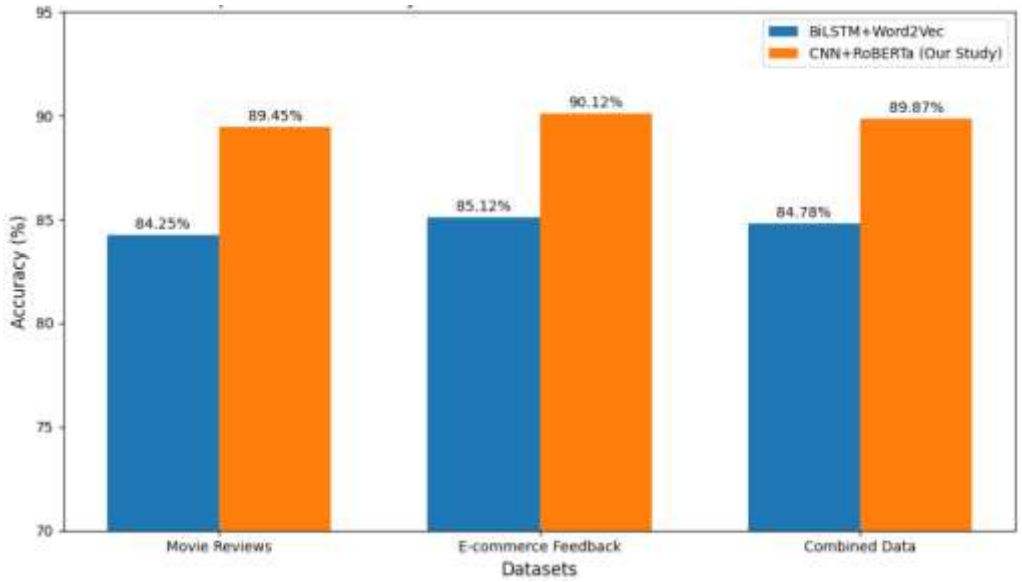

Fig 3: Comparison of Accuracy: BiLSTM+Word2Vec vs. CNN+RoBERTa Across Datasets"

Figure 3 compares the accuracy results from our study with those from previous studies that used BiLSTM+Word2Vec, demonstrating the improvement in performance when using CNN combined with RoBERTa.
Overall, RoBERTa+CNN outperforms both FastText+CNN and GloVe+CNN, and the study shows that dynamic embeddings, such as RoBERTa, are key to achieving higher performance in emotion detection tasks.

VI. CONCLUSION

This study successfully demonstrated the effectiveness of Convolutional Neural Networks (CNN) for emotion detection in text, with a focus on comparing various word embedding techniques such as FastText, RoBERTa, and GloVe. The research utilized three distinct types of datasets: movie reviews, e-commerce customer feedback, and a combined dataset of both. The experiments were conducted across three scenarios: FastText+CNN, RoBERTa+CNN, and GloVe+CNN.

The results indicate that **RoBERTa+CNN** consistently outperforms the other methods, achieving the highest accuracy across all datasets. In particular, RoBERTa+CNN produced accuracy values of 89.45%, 90.12%, and 89.87% for the movie reviews, e-commerce feedback, and combined datasets, respectively. FastText+CNN showed strong performance but lagged behind RoBERTa+CNN in terms of precision, recall, and F1-Measure. GloVe+CNN exhibited the lowest performance in comparison, highlighting the advantages of more context-sensitive embeddings like RoBERTa.
The comparison of Precision, Recall, and F1-Measure further supports the superiority of RoBERTa+CNN, reinforcing that context-aware embeddings provide better results for emotion detection tasks. This study's findings emphasize the importance of dynamic word embeddings in capturing the nuances of human emotions in text, particularly for tasks like emotion classification.

Suggestions for future research include exploring the integration of CNN with other deep learning architectures, such as BiLSTM or LSTM, to enhance emotion detection capabilities. Additionally, further investigations into the use of RoBERTa embeddings combined with these models could provide insights into improving classification accuracy. Future work could also focus on using more balanced datasets to address any potential biases in emotion class distribution, which would likely contribute to more reliable and generalizable emotion detection models.

This research highlights the potential for improved emotion detection through the use of advanced word embeddings and CNN, paving the way for more accurate and contextually aware emotion classification systems in text.

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 2, February 2025

DOI: 10.17148/IJARCCE.2025.14228

REFERENCES

[1]. Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. Character-level convolutional networks for text classification. In Proceedings of the 29th International Conference on Neural Information Processing Systems - Volume 1 (NIPS'15), Vol. 1. MIT Press, Cambridge, MA, USA, 649–657.

[2]. Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.

[3]. Cícero dos Santos and Maíra Gatti. 2014. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, pages 69–78, Dublin, Ireland. Dublin City University and Association for Computational Linguistics.

[4]. Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv preprint arXiv:1301.3781.

[5]. Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

[6]. Devlin, J., Chang, M., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.

[7]. Rambarki Sai Akshit, Konduru Hema Pushpika, Rambarki Sai Aashik, Dr. Manda Rama Narasinga Rao, J. Uday Shankar Rao, V. Pavan Pranesh, 2024, Automated Traffic Ticket Generation System for Speed Violations using YOLOv9 and DeepSORT, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 13, Issue 12 (December 2024),

[8]. Wibawa, Aji & Cahyani, Denis & Prasetya, Didik & Gumilar, Langlang & Nafalski, Andrew. (2023). Detecting emotions using a combination of bidirectional encoder representations from transformers embedding and bidirectional long short-term memory. International Journal of Electrical and Computer Engineering (IJECE). 13. 7137. 10.11591/ijece.v13i6.pp7137-7146.

[9]. Marco Polignano, Pierpaolo Basile, Marco de Gemmis, and Giovanni Semeraro. 2019. A Comparison of Word-Embeddings in Emotion Detection from Text using BiLSTM, CNN and Self-Attention. In Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization (UMAP'19 Adjunct). Association for Computing Machinery, New York, NY, USA, 63–68. https://doi.org/10.1145/3314183.3324983

[10]. D. E. Cahyani and I. Patasik, "Performance comparison of tf-idf and word2vec models for emotion text classification," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 5, 2021, doi: 10.11591/eei.v10i5.3157

[11]. Rambarki Sai Akshit; Konduru Hema Pushpika; Rambarki Sai Aashik; Dr. Ravi Bhramaramba; Sayala Manjith; Paladugu Madhav. (Volume. 10 Issue. 1, January - 2025) "Brain Tumor Classification Using CNN on MRI Data: A PyTorch Implementation." International Journal of Innovative Science and Research Technology (IJISRT), www.ijisrt.com. ISSN - 2456-2165 , PP :-103-109, https://doi.org/10.5281/zenodo.14621403