Text Emotion Analysis

Anurag singh
CCE,LNM institute of information technology
Jaipur,India
21ucc020@lnmiit.ac.in

Mohit jain
CSE,LNM institute of information technology
Jaipur,India
21ucs131@lnmiit.ac.in

Himanshu Gautam
CSE,LNM institute of information technology
Jaipur,India
21ucs240@lnmiit.ac.in

ABSTRACT:
Emotional recognition has emerged as a crucial area of research, offering valuable insights across various domains, including psychotherapy, customer experience, and human-computer interaction. Emotions manifest through speech, facial expressions, gestures, and written text.This study focuses on detecting emotions in textual data, a challenge rooted in Natural Language Processing (NLP) and deep learning. Leveraging semantic text analysis, word embeddings play a pivotal role in capturing the syntactic and semantic nuances of textual content,enhancing tasks like machine translation and sentiment analysis. We propose a deep learning assisted semantic text analysis (DLSTA) model that achieved high performance metrics, including accuracy, recall, and F1 scores above 91%, outperforming baseline models.Building upon prior advancements, this study delves into dimensional emotion recognition, which maps emotions onto continuous scales such as valence, arousal, and dominance (VAD), enabling a deeper and more precise understanding of emotional states. Leveraging a transformer-based model, fine-tuned using a novel dimensional emotion dataset comprising 75,503 samples, the approach demonstrates significant enhancements over state-of-the-art techniques. The model addresses critical challenges, including heterogeneity in annotations and domain transfer, achieving remarkable Pearson correlation coefficients of 0.90 for valence, 0.77 for arousal, and 0.64 for dominance. By capturing the nuanced and subjective facets of emotional expressions, this methodology complements traditional categorical models, advancing precision, robustness, and applicability in diverse settings. These improvements hold particular promise for applications in therapeutic contexts and interactive systems, where nuanced emotion detection can significantly enhance user experiences.

1. INTRODUCTION

1.1 The Area of Work

Emotional recognition examines human behavior and interactions, focusing on emotions as complex patterns of experiential, behavioral, and physiological reactions. These are crucial for decision-making and communication.

Analyzing text-based emotions in mediums such as social media posts and reviews provides valuable insights for marketing, healthcare, and human-computer interaction. The digital content surge has created vast data for such analysis. Using Natural Language Processing (NLP) and deep learning, this study leverages deep learning-assisted semantic text analysis (DLSTA) and incorporates emotional dimensions like valence, arousal, and dominance to develop scalable systems for accurate emotion understanding.

1.2 Problem Addressed

The rapid growth of digital content calls for efficient methods to analyze human emotions in text. Challenges such as sarcasm, ambiguity, and cultural nuances hinder traditional methods. This project develops a deep learning-based model for detecting and classifying emotions in textual data. It emphasizes precision to capture subtle emotional variations across contexts with reliability.

1.3 Existing System

1.3.1 Deterministic Models

Deterministic models classify emotions into discrete categories (e.g., anger, joy) through supervised learning techniques like Naive Bayes, SVM, and CNNs. Advanced models such as EmoNet and DeepMoji enhance emotion recognition, with EmoNet using GRNNs and DeepMoji leveraging emoji-labeled tweets. However, they struggle with context-dependent and subjective emotions.

1.3.2 Limitations of Categorical Models

Categorical models face challenges in addressing:

1. **Emotional Ambiguity:** Mixed emotional states are hard to represent.
2. **Subjectivity:** Emotional perceptions vary significantly.

1.3.3 Dimensional Models

Dimensional models represent emotions in continuous spaces (e.g., valence, arousal), offering nuanced analyses. Frameworks like Russell's Circumplex Model and datasets such as EmoBank employ VAD dimensions. Recent techniques like Gaussian Mixture Models (GMMs) map emotions as probabilistic distributions to address subjectivity and fuzzy boundaries.

1.4 Advancements in Deep Learning Frameworks

Text classification (TC) categorizes text into predefined groups, relying on rule-based, machine learning, or hybrid methods [1, 2, 3]. Deep learning (DL) architectures like CNN, LSTM, and BiLSTM enhance NLP tasks [6, 7]. Combining CNN and BiLSTM captures spatial and contextual relationships, achieving accuracies of up to 82% [8, 10, 11, 12].

This research extends these frameworks to informal microblogging platforms (e.g., Twitter, Facebook), where brevity, linguistic variations, and contextual ambiguities challenge emotion detection [13]. By incorporating dimensional emotion models and blending deterministic and probabilistic methods, this study aims to create adaptable, scalable emotion recognition systems for modern NLP applications.

2. LITERATURE REVIEW

2.1 Introduction

Text classification and sentiment analysis are foundational tasks in Natural Language Processing (NLP) that have witnessed substantial improvements through the adoption of deep learning architectures. These advancements have enabled researchers to address complex emotion detection challenges, particularly in multilingual and informal text datasets. By combining state-of-the-art techniques like Convolutional Neural Networks (CNNs), Bidirectional Long Short-Term Memory networks (Bi-LSTMs), and attention mechanisms, recent studies have achieved significant milestones in classification accuracy and robustness.

2.2 Summary of Relevant Research

Recent research in text emotion analysis spans deterministic and dimensional approaches, focusing on hybrid deep learning architectures and probabilistic models. These studies have significantly advanced the field by improving classification accuracy and addressing the challenges of subjective and ambiguous emotional expressions in textual data.

2.2.1 Deterministic and Hybrid Models

Deterministic and hybrid models have been foundational in advancing text emotion analysis. Deterministic approaches categorize text into fixed emotion labels, while hybrid models combine multiple deep learning techniques to achieve higher accuracy and robustness. These methodologies address challenges in emotion classification by integrating local feature extraction, temporal dependencies, and attention mechanisms, providing a strong baseline for emotion detection tasks.

2.2.1.1 Hybrid Bi-LSTM and Attention Mechanisms:

A study leveraging a Bi-LSTM architecture combined with Word2Vec embeddings, CNNs,and attention mechanisms achieved remarkable performance, with an F1-score of 90.1% on a 13k-instance dataset [12]. By assigning attention weights to significant parts of the text and utilizing Word2Vec embeddings for semantic enrichment, the model demonstrated the advantages of integrating multiple methodologies.

2.2.1.2 Character Embeddings and Multilingual Sentiment Classification:

Salur and Aydin [10] presented a hybrid CNN-BiLSTM model that outperformed standalone models by incorporating character embeddings and pre-trained embeddings like GloVe and FastText. This architecture, tested on Turkish GSM-related tweets, achieved an accuracy of 82.14%, underscoring the value of hybrid deep learning approaches for languages with complex morphological structures.

2.2.1.3 Deep Learning for Low-Resource Languages:

Naqvi et al. [11] addressed sentiment analysis in Urdu, utilizing CNNs for feature extraction and RNNs for capturing temporal dependencies. With a Bi-LSTM ATT model achieving an accuracy of 77.9% and an LSTM model with Samar embeddings reaching 85.16% precision, the study highlighted the potential of tailored embeddings for low-resource languages.

2.2.2 Dimensional Emotion Models:

Dimensional models provide a continuous representation of emotions, moving beyond fixed categories to capture emotional nuances in multidimensional spaces. These frameworks, such as Valence-Arousal models, focus on representing the intensity and pleasantness of emotions, offering a more flexible and subjective analysis. Emerging approaches use probabilistic distributions to address fuzzy boundaries and overlapping emotional states.

2.2.2.1 Challenges with Deterministic Models:

Traditional emotion analysis systems classify emotions into fixed categories, which limits their ability to handle subjective variations and fuzzy emotional boundaries. For example, texts like ”The virus is spreading” can evoke fear, sadness, or both, depending on the context and reader’s perception.

2.2.3 Context-Aware Models:

Ghafoor [14]. (2023) developed the TERMS model, which employs Gaussian Mixture Models (GMMs) for soft assignment of emotional perceptions in valence-arousal space. Evaluated on 4,000 Twitter messages, TERMS achieved Pearson correlation coefficients of 0.6 for valence and 0.3 for arousal, outperforming prior models like DeepMoji and SemEval-2018 The TERMS (Textual Emotion Recognition in Multidimensional Space) model addresses these challenges by mapping emotions into continuous Valence-Arousal dimensions using Gaussian Mixture Models (GMMs). By representing emotions as probability distributions, TERMS captures overlapping and ambiguous emotional states, making it particularly effective for analyzing informal and opinionated data from social media platforms.

2.3 Comparative Studies:

A recent survey by Al Maruf [15]. (2024) consolidated insights into text-based emotion recognition. They noted the predominance of categorical models and identified gaps in dataset diversity and annotation consistency. Their analysis underscored the potential of pooled datasets and dimensional approaches for improved emotion modeling.

2.4 Critical Analysis of Existing Research:

The landscape of emotion recognition in text has seen substantial advancements, especially with the integration of deep learning models such as CNN, BiLSTM, and transformers. However, despite these strides, several limitations remain, particularly in terms of the complexities inherent in emotional expression and the adaptability of current models across varied domains.

One prominent limitation is the over-reliance on categorical emotion models, which limit the representation of emotions to predefined labels. This approach is common in systems like Woebot (Gabriels[16]), which categorizes emotional states but fails to capture the continuous and multifaceted nature of human emotions. The categorical model restricts the system’s capacity to recognize subtle variations in emotions, which is critical for applications like cognitive behavioral therapy (CBT). On the other hand, some advanced systems, such as the work of Ghafoor [14] have begun incorporating dimensional models, specifically Valence-Arousal-Dominance (VAD) spaces, to better represent the complexity of emotions. These models, which predict emotional states on a continuous scale rather than discrete labels, hold great promise in addressing the nuanced nature of emotional expression.

Furthermore, many existing emotion recognition models suffer from challenges related to dataset biases. For example, Al Maruf [15] highlighted issues of unbalanced datasets and a lack of standardized annotations, which can hinder the performance and generalizability of emotion recognition models. The lack of domain adaptability and inconsistency in dataset structures remain significant challenges, especially when models trained on specific domains are applied to new, unrelated domains.

Additionally, existing models often neglect the integration of contextual information, which is vital for effective emotion recognition, particularly in settings like therapeutic interventions. For instance, systems like Youper and Wysa, although effective, utilize discrete emotional categories without leveraging contextual understanding of emotional progression over time .

Context-aware models, like TERMS (Ghafoor [14]), aim to solve this issue by incorporating context-specific classifiers that track emotions across interactions and adapt to the emotional dynamics of a conversation.

2.5 Connection to the Current Research

This study builds upon the strengths of hybrid models while addressing their limitations by exploring dimensional frameworks. By integrating deterministic CNN-BiLSTM models with the context-aware and probabilistic capabilities. The current research builds upon these findings by introducing a transformer-based model that captures a broader range of emotional dimensions, specifically targeting valence, arousal, and dominance (VAD). Unlike previous categorical systems, this model offers continuous emotion tracking, allowing for finer-grained emotional understanding. By leveraging this transformer-based model, the research seeks to overcome the limitations of fixed emotion categories, providing more precise emotion recognition that is essential for applications like cognitive behavioral therapy (CBT) where emotional subtleties are critical.

3. DESIGN AND SIMULATION

The dataset, labeled with sentiment classifications, underwent testing with multiple algorithms,including Neural Networks, LSTM, BiLSTM, and CNN. Before initiating the testing phase,an extensive preprocessing pipeline was implemented to ensure data integrity and relevance.This process involved the removal of elements such as numerical values, hashtags, URLs,emojis, and punctuation to refine the dataset. Following the cleaning phase, feature extraction techniques were applied, which included tokenization and encoding to represent the text effectively.

The data was subsequently partitioned into subsets and processed through padding techniques to standardize sequence lengths, ensuring compatibility with the modeling framework. A comparative analysis was conducted to evaluate the test results against the proposed hybrid model.

Initially, a simple BiLSTM model was deployed on the curated datasets to establish baseline performance. The structure and methodology of the BiLSTM model are outlined in the subsequent sections.

Layer	Output Shape	Parameters
Embedding Layer	(None, 200, 100)	7,530,300
BiLSTM Layer	(None, 256)	234,496
Dense Layer	(None, 6)	1,542

TABLE 3.1: Summary of BiLSTM Model Layers

The construction of the hybrid CNN-BiLSTM model began with the phases of data labeling,preprocessing, and feature extraction, as outlined previously. To represent text data, we utilized an embedding method that converts textual information into low-dimensional numeric vectors,with a padding size fixed at 300. This approach deviates from earlier methodologies to enhance representation and compatibility.

Layer	Output Shape	Parameters
Embedding Layer	(None, 200, 100)	7,530,300
Convolution1D Layer	(None, 200, 64)	19,264
Pooling Layer	(None, 100, 64)	0
BiLSTM Layer	(None, 256)	197,632
Dense Layer	(None, 6)	1,542

TABLE 3.2: Summary of CNN+BiLSTM(1) Model Layers

And also one more model as follows:

Layer	Output Shape	Parameters
Embedding Layer	(None, 200, 100)	7,514,600
Convolution1D Layer	(None, 198, 200)	60,200
BiLSTM Layer	(None, 198, 128)	135,680
Dropout	(None, 198, 128)	0
BiLSTM Layer	(None, 128)	98,816
Dense	(None, 50)	6,450
Dense	(None, 50)	2,550
Flatten	(None, 50)	0
Dense	(None, 100)	5,100
Dense	(None, 6)	606

TABLE 3.3: Summary of CNN+BiLSTM(2) Model Layers

5.RESULTS ANALYSIS

5.1 Evaluation of Our Proposed Model

5.1.1 Accuracy with Different Datasets and Embeddings

The performance of different models with various datasets and embeddings is summarized in
Table 4.1.

Model	Dataset	Embedding	Accuracy
ML (SVM)	1	Tf-idf vectorizer	72.71%
ML (SVM)	2	Tf-idf vectorizer	87.11%
ML (Logistic Regression)	3	Tf-idf vectorizer	73.4%
BiLSTM	1	Word2Vec (Wikipedia file)	72.68%
BiLSTM	2	GloVe (Twitter)	92.26%
CNN-BiLSTM (1)	1	Word2Vec (Wikipedia file)	73.36%
CNN-BiLSTM (1)	2	GloVe (Twitter)	93.46%
CNN-BiLSTM (2)	1	Word2Vec (Wikipedia file)	73.16%
CNN-BiLSTM (2)	2	GloVe (Twitter)	94%
CNN-BiLSTM (2)	3	Word2Vec	83.6%
BERT	1	WordPiece	82.67%
BERT	2	WordPiece	97.21%
BERT	3	WordPiece	87.21%

TABLE 4.1: Model Performance with Different Datasets and Embeddings

5.1.2 Classification Metrics

The precision, recall, and F1-score for different emotion categories are provided in Table 4.2.

Emotion	Precision	Recall	F1-Score
Anger	0.94	0.95	0.94
Fear	0.94	0.86	0.90
Joy	1.00	0.91	0.95
Love	0.77	1.00	0.87
Sadness	0.98	0.97	0.97
Surprise	0.75	0.99	0.85

TABLE 4.2: Emotion Classification Metrics

5.2 Evaluation of Dimensional Emotion Recognition Model

The evaluation demonstrates that the DL-based approach outperforms the rule-based method in emotion recognition tasks. Leveraging metrics such as Mean Squared Error (MSE) and the Pearson correlation coefficient (r), the DL model consistently achieves lower MSE values and higher correlation scores across all datasets. This indicates its superior capability to capture the underlying emotional dimensions. Its advantage lies in training on a combined dataset,allowing it to generalize effectively to unseen data during split evaluations.

Precision, recall, and F1 scores on the combined dataset reveal the DL model's significant superiority over the rule-based approach, achieving higher macro, micro, and average F1 scores.

Metric	DL-Group			Rule-Group		
	Precision	Recall	F1	Precision	Recall	F1
Empty	0.87	0.66	0.75	0.26	0.04	0.07
Threatened	0.87	0.90	0.89	0.31	0.29	0.05
Tranquil	0.79	0.56	0.65	0.19	0.24	0.21
Excited	0.77	0.92	0.84	0.52	0.40	0.07
Rooted	0.89	0.90	0.89	0.61	0.94	0.74
	Macro F1	**Micro F1**	**Average F1**	**Macro F1**	**Micro F1**	**Average F1**
Total	**0.80**	**0.86**	**0.86**	**0.23**	**0.59**	**0.48**

6.Conclusions and Future Work

We successfully demonstrate the effectiveness of a hybrid CNN-BiLSTM model for text emotion analysis, particularly in the domain of sentiment analysis. The model, combining convolutional and bidirectional long short-term memory networks, exhibits high performance across various datasets. It captures both spatial and temporal features of the text, leading to precise emotion classification with accuracy rates up to 94% using Glove embeddings. The model excels in accurately classifying emotions including joy, sadness, anger, fear , disgust, neutral and surprise.

We have successfully developed a deep learning approach for dimensional emotion recognition using the transformer-based ALBERT model. By fine-tuning the model to predict emotions across the Valence, Arousal, and Dominance (VAD) dimensions, we demonstrated the power of transformers to accurately capture emotions in text. Our model generalizes well across diverse emotional expressions, benefiting from balanced dataset merging. ALBERT's lightweight architecture enables high performance and computational efficiency, making it practical for real-world applications. This work advances emotion recognition by enabling continuous VAD prediction, opening doors for applications in sentiment analysis and emotional health monitoring.

Future work in emotion recognition could explore multimodal approaches, combining text, speech, and facial expressions for more accurate emotion detection. Models could be finetuned for specific domains, like healthcare or finance, and adapted for multilingual and crosscultural contexts. Real-time emotion recognition for interactive systems, personalized emotional profiles, and expanding dimensional emotion recognition (beyond VAD) would further enhance applications. Ethical considerations and model biases should be addressed, while integration with VR/AR could improve user experiences. Additionally, improving accuracy in noisy, informal text and advancing hybrid models with explainability could make emotion recognition systems more reliable and impactful in diverse settings.

7. REFERENCES

[1] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed, A. Mehmood, and M. T. Sadiq,“Document-level text classification using single-layer multisize filters convolutional neural network,” IEEE Access, vol. 8, pp. 42689–42707, 2020.

[2] A. Wahdan, S. Hantoobi, S. A. Salloum, and K. Shaalan, “A systematic review of text classification research based on deep learning models in arabic language,” International Journal of Electrical and Computer Engineering, vol. 10, no. 6, pp. 6629–6643, 2020.

[3] Y. Zhang, T. Wang, and J. Li, “Object classification by effective segmentation of tree canopy using u-net model,” Journal of Advances in Information Technology, vol. 15, no. 3, pp. 424–432, 2024.

[4] W. Fang, H. Luo, S. Xu, P. E. D. Love, Z. Lu, and C. Ye, “Automated text classification of near-misses from safety reports: An improved deep learning approach,” Advanced Engineering Informatics, vol. 44, p. 101060, 2020.

[5] Z. Liu, C. Lu, H. Huang, S. Lyu, and Z. Tao, “Hierarchical multi-granularity attention-based hybrid neural network for text classification,” IEEE Access, vol. 8, pp. 149362–149371, 2020.

[6] Q. L. et al., “A survey on text classification: From shallow to deep learning,” IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 11, pp. 1–21, 2020.

[7] F. Zaman, M. Shardlow, S. Hassan, and N. Radi, “Htss: A novel hybrid text summarisation and simplification architecture,” Information Processing and Management, vol. 57, no. 6, p. 102351, 2020.

[8] K. Pasupa, T. Seneewong, and N. Ayutthaya, “Thai sentiment analysis with deep learning techniques: A comparative study based on word embedding, pos-tag, and sentic features,” Sustainable Cities and Society, vol. 50, p. 101615, 2019.

[9] S. Kumar, C. Akhilesh, K. Vijay, and B. Semwal, “A multibranch cnn-bilstm model for human activity recognition using wearable sensor data,” Visual Computing, vol. 38, no. 12, pp. 4095–4109, 2021.

[10] M. U. Salur and I. Aydin, “A novel hybrid deep learning model for sentiment classification,” IEEE Access, vol. 8, pp. 58080–58093, 2020.

[11] U. Naqvi, A. Majid, and S. A. L. I. Abbas, “Utsa: Urdu text sentiment analysis using deep learning methods,” IEEE Access, vol. 9, pp. 114085–114094, 2021.

[12] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, “Bi-lstm model to increase accuracy in text classification: Combining word2vec cnn and attention mechanism,” Applied Sciences, vol. 10, no. 17, p. 5841, 2020.