A Hybrid Lexico-Transformer Model for Real-Time Emotion Detection in English Text

Israa Mohammed Ahmed[1] and Mohammed Khudhair Abbas[2]
[1] *Department of Computer Science, College of Computer Science and Mathematics, Tikrit University, 34001 Tikrit, Iraq*
[2] *Department of Mobile Communications and Computing Engineering, College of Engineering, University of Information Technology and Communications, 10013 Al-Mansour, Baghdad, Iraq*
asraa.m.ahmed@tu.edu.iq, mohammed.abbas@uoitc.edu.iq

Keywords: Emotion Detection; Hybrid Framework; Transformer; Lexicon Enrichment; Dynamic Fusion Module.

Abstract: Emotion detection in text is expressed as a crucial component of almost all artificial intelligence (AI) applications, so far it remains a challenging approach because of linguistic variety and real-time situations. This paper suggests DeepEmotion+, a hybrid approach which gathers a custom-built emotional lexicon with the transformer-based contextual learning in order to enhance both the accuracy and emotion classification speed. The proposed approach consists of two main pipeline stages, which include: Lexical-Preprocessing, where the text is tokenized, part-of-speech tagged, and enriched utilizing an extra domain-specific impact lexicon; and Transformer-Classification, where contextual embeddings with the lightweight transformer and lexicon-derived features are obtained through a novel Dynamic Fusion Module (DFM). The proposed approach validates its method on many datasets, illustrating an overall F1-score enhancement of about 3-5% compared with state-of-the-art studies in streaming situations and conditions. DeepEmotion+ consistently achieves an average accuracy of about 87%. In addition, the proposed approach ensures inference latencies below 50 ms per sentence on a CPU, enabling real-time deployment. These results express the underscored effectiveness and efficiency of DeepEmotion+ in practical text analysis.

1 INTRODUCTION

1.1 Background

The application of Natural Language Processing (NLP) techniques to analyze textual data – such as customer reviews – in order to identify the underlying sentiment and classify it as positive, negative, or neutral is known as sentiment (or emotion) analysis [1]–[3]. The volume of textual data shared online is enormous; for example, over 500 million tweets are posted daily, expressing opinions and emotions [4]. By developing the capability to analyze such large-scale, high-velocity, and high-variety data, organizations can make informed, data-driven decisions.

Sentiment analysis can be broadly categorized into three main types: multimodal sentiment analysis, aspect-based sentiment analysis, and cross-linguistic (multilingual) sentiment analysis [5]. Multimodal sentiment analysis integrates multiple data modalities – such as text, audio, and video – to determine sentiment. Visual and auditory cues, including facial expressions and tone of voice, often convey rich emotional information that complements textual content [6]–[8].

Aspect-based sentiment analysis focuses on identifying sentiments associated with specific attributes or aspects of products or services. For instance, in a restaurant review, sentiment may be extracted separately for food quality, service, and ambiance [9]. Multilingual sentiment analysis addresses the challenge that emotions are expressed differently across languages due to variations in grammar, syntax, and vocabulary [10], [11]. In such approaches, models are typically trained separately or adapted for each language to accurately capture sentiment-specific linguistic patterns.

A typical sentiment analysis pipeline involves collecting textual data (e.g., customer reviews, social media posts, or comments), preprocessing it (including noise removal, tokenization, part-of-speech tagging, and stemming or lemmatization), extracting features (by converting textual information into numerical representations), and classifying the text into sentiment categories such as positive, negative, or neutral [12]–[14].

Emotion detection from textual data has witnessed rapid advancement due to the increasing adoption of artificial intelligence (AI) techniques and the growing power of NLP models [15]. Human emotions are inherently nuanced and highly context-dependent, and they are often encoded in written language through explicit or implicit expressions. Accurate emotion detection from text plays a crucial role in several application domains, including mental health assessment, customer feedback analysis, emotion-aware systems, and human–computer interaction (HCI) [16], [17].

Traditional approaches to emotion detection rely heavily on rule-based methods and emotion lexicons. Although these approaches offer a degree of interpretability, they often suffer from limited scalability, reduced flexibility, and insufficient contextual awareness [18], [19]. In contrast, machine learning and deep learning techniques – such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer-based models – have demonstrated superior performance by capturing complex linguistic structures and long-range dependencies [20], [21]. However, these models may overlook psychologically meaningful lexical cues that enhance interpretability and emotional sensitivity, particularly in real-time systems where inference latency is a critical constraint [22]–[24].

In this article, we propose a hybrid model named DeepEmotion+, which combines the contextual learning capabilities of lightweight transformer architectures with the emotional sensitivity of a custom-designed impact lexicon. Unlike many existing approaches that focus primarily on maximizing classification accuracy, the proposed framework explicitly emphasizes real-time performance and generalizability, making it suitable for practical deployment in time-sensitive emotion analysis applications.

1.1.1 Objectives

The main objectives of modeling the proposed framework are to develop a real-time emotion detection approach that integrates a custom emotion lexicon with transformer-based learning, and to design a novel Dynamic Fusion Module (DFM) that effectively combines lexical and contextual features. In addition, this study aims to validate the effectiveness of the proposed approach by conducting a comprehensive performance analysis using multiple evaluation metrics, including accuracy, F1-score, and inference latency, and by performing a detailed comparison with related state-of-the-art methods.

1.1.2 Paper Importance

This paper presents a novel and practical approach that enhances emotion classification through the fusion of lexical and contextual features. A key contribution of the proposed framework is its low-latency efficiency, which is a critical requirement for real-time applications. Furthermore, the proposed framework incorporates an explainability-oriented design, enabling better interpretability and adaptability across different application domains.

1.1.3 Paper Organization

The remainder of this paper is organized as follows. Section 2 reviews related works and provides a critical comparison with existing studies in the literature. Section 3 describes the proposed methodology, including the overall framework, the proposed algorithm, and the mathematical formulations used to achieve the desired functionalities. Section 4 presents and discusses the experimental results, highlighting the most significant findings. Finally, Section 5 concludes the paper by summarizing the proposed framework and algorithm, discussing the obtained results, and outlining directions for future research.

2 LITERATURE REVIEW

In this section, a set of previous studies related to the topic and objectives of this article is reviewed and discussed.

Acheampong et al. provided a comprehensive analysis of transformer-based models, including BERT, GPT, and Transformer-XL, for various NLP tasks with a particular focus on emotion recognition [25]. Their study highlights BERT's effectiveness in capturing contextual emotional representations and offers critical insights into model performance, limitations, and future research directions. This survey underlines the growing dominance of transformer-based architectures in text-based emotion recognition.

Mahima et al. proposed a hybrid framework that integrates rule-based processing, sentiment detection, and contextual understanding using sentence transformers and BERT [26]. The proposed system addresses the challenge of identifying multiple coexisting emotions within textual data, overcoming key limitations of traditional sentiment analysis methods. By leveraging Ekman's basic emotions along with neutral states, the model enables more nuanced multi-emotion tagging in context-rich texts.

AbuAin et al. introduced a hybrid LSTM–Transformer architecture evaluated on three benchmark speech emotion datasets [27]. Mel-Frequency Cepstral Coefficients (MFCCs) were employed for audio preprocessing. The results indicate that model performance varies across datasets, with the hybrid approach outperforming competing models on TESS-DB, while showing weaker performance on EMO-DB and SAVEE. These findings emphasize the dataset-dependent strengths of hybrid architectures in speech-based emotion recognition.

Kumar et al. presented a transformer-based approach for emotion detection in social media text, aiming to enhance sentiment classification by capturing fine-grained emotional states such as anger and happiness [28]. By employing contextual embeddings and a one-cycle learning rate strategy, their model achieved an accuracy improvement of approximately 6% over existing approaches. This study further confirms the effectiveness of transformers in modeling nuanced user expressions in online environments.

Boutouta et al. addressed implicit emotion recognition in Arabic text by proposing a hybrid AraBERT–BiGRU model [29]. Using the newly curated AIEmoD dataset alongside two additional Arabic corpora, the proposed approach achieved strong F1-scores, significantly outperforming baseline models. Their work pioneers implicit emotion detection in Arabic, highlighting the importance of language-specific architectures and contextual embeddings, particularly for low-resource languages.

Ayari et al. developed a hybrid contextual emotion recognition framework for cognitive assistance in ubiquitous environments, combining multilayer perceptron (MLP) networks with expressive reasoning through emotion ontologies [30]. The model was validated using YouTube data and a smart device showroom scenario, demonstrating strong performance in handling context-aware and non-observable emotions. The study is notable for integrating emotional reasoning into real-world intelligent service systems.

Table 1 summarizes a comparative analysis between the proposed approach and the reviewed literature in terms of methodologies, key findings, advantages, and limitations.

3 METHODOLOGY

The proposed DeepEmotion+ approach operates through two parallel processing streams: Lexical Preprocessing and Transformer-Based Classification, as illustrated in Figure 1. The primary objective is to extract both rule-based and context-aware features and dynamically fuse them using the proposed Dynamic Fusion Module (DFM) to achieve robust emotion classification. Figure 1 illustrates the overall architecture of the proposed framework, where the primary input is textual data. The input text is processed simultaneously through the Lexical Preprocessing stream and the Transformer-Based Classification stream, enabling the model to capture complementary emotional representations.

3.1 Lexical Preprocessing

In the Lexical Preprocessing stage, three sequential operations are performed: tokenization, part-of-speech (POS) tagging, and lexicon-based feature extraction. Each token is matched against a predefined emotion lexicon to retrieve corresponding intensity values. These values are aggregated to form a lexical feature vector L, which represents the emotional weight distribution of the input sentence.

3.2 Transformer-Based Classification

The Transformer-Based Classification stage consists of three sequential operations: embedding generation, classification, and prediction. This stream produces contextual embeddings E using a lightweight transformer encoder, which captures deep semantic and contextual information from the input sentence. These embeddings serve as a high-level representation of the sentence semantics for emotion classification.

Table 1: Comparison of the proposed work with the reviewed literature.

Study	Findings	Method	Cons.
Proposed	+3–5% F1, 87% acc., <50ms latency	Lexicon + transformer + fusion	Text-only focus; limited domain testing
Study [25]	BERT > traditional NLP	Survey of transformer models	No new model; no experiments
Study [26]	Good multi-emotion detection	Rule-based + BERT + embeddings	Dependent on lexicons & embeddings
Study [27]	Best on TESS; mixed elsewhere	LSTM + Transformer on audio	Weak on some sets; not for text
Study [28]	+6% accuracy gain	Transformer + one-cycle learning	English-only; low generalization
Study [29]	F1: 79.87% AETD, 70.67% AIEmoD	AraBERT + BiGRU; new Arabic dataset	Language-specific; domain-limited
Study [30]	Beats baselines in smart settings	MLP + logic + emotion ontologies	Complex; needs detailed domain knowledge

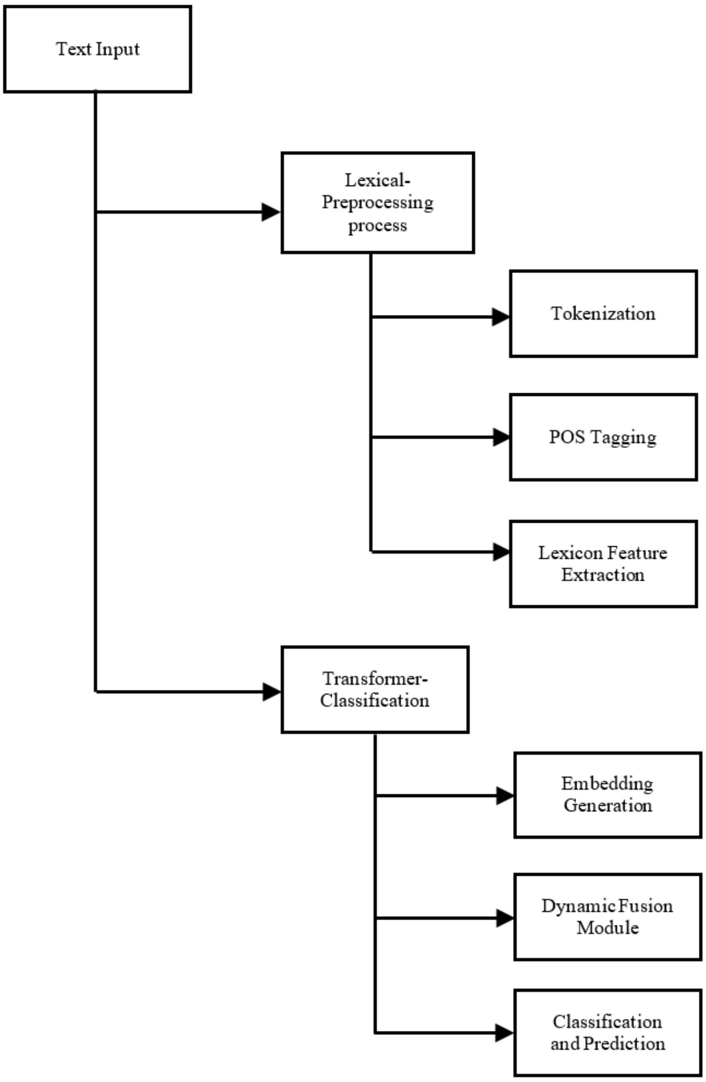

Figure 1: Proposed framework.

3.3 DFM

The core innovation of the proposed approach is the DFM, that linearly gathers both feature streams as shown in (1):

$$F = \alpha \cdot E + (1 - \alpha) \cdot L. \quad (1)$$

Where:
- F: the fuesd feature factor.
- E: the contextual embedding.
- L: the lexicon feature vector.
- $\alpha \in [0,,,1]$: the weight for controlling fusion-balance.

3.4 Classification and Prediction

The fused feature F is used to passe to classifier layer in order to generate the class logits as shown in (2):

$$z = W_f \cdot F + b. \quad (2)$$

Where:
- W_f: learnable weight parameter.
- b: learnable bias parameter.

While the class probabilities which are obtained based on SoftMax function as shown in (3):

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}}. \quad (3)$$

3.5 Proposed Algorithm

As the proposed framework illustrated in Figure 1 is expressed as a novel hybrid real-time emotion detection for input text, there should be an algorithm that contain comprehensive steps. A proposed algorithm steps are shown in Algorithm 1.

Based on Algorithm 1, the fusion of features is applied for making hybrid transformer and lexicon by gathering both the transformer-embeddings (E) and lexicon-features (L) based on *transformer embeddings* weight W_e and *lexicon features* weight W_l as shown in Equation 4:

$$Fused = \tanh(E \times W_e^T + L \times W_l^T + b). \quad (4)$$

For classification layer, maps fused-features into the logits class as shown in (5):

$$Logits = Fused \times W_f^T + b_f. \quad (5)$$

Algorithm 1. Proposed Algorithm
1) Start
2) Load text-input
3) Load the correlation labels.
4) Load the powerful lexicon.
5) For each sentence in the input-text Do the following:
   - Tokenize lexicon-features;
   - Extract lexicon-features.
6) Generate contextual embeddings.
7) Simulate contextual embeddings.
8) Fuse both feature sets utilizing:

   $$\mathrm{fused} = \tanh(\mathrm{W_e} .\times \mathrm{E} + \mathrm{W_l} \times \mathrm{L} + \mathrm{b})$$

9) Predict based on:

   $$\mathrm{logits} = \mathrm{W_f} \times \mathrm{fused} + \mathrm{b_f}$$

10) Run SoftMax to derive class probabilities.
11) Evaluate utilizing confusion matrix:
    - precision-recall.
    - ROC.
    - Latency.
    - Comparative metrics.
12) End

3.6 Evaluation Metrics

Performance is measured using F1-Score, ROC-AUC, and Latency. For evaluating the F1-Score per class based on both precision and recall. Both the ROC and area under the curve AUC metrics are derived from the true and false positive rates. While the distribution of latency is simulated through Gaussian samples about (30ms) as a mean value.

3.7 Lexicon Construction

In order to build the custom emotion lexicon:
1) Extract lexicon from both NRC Emotion-Lexicon and WordNet-Affect.
2) Review all the words manually and then aive an intensity value from 0 (neutral) to 1 (highly emotional).
3) Based on POS tag-expansion, add both the synonyms and the inflected-forms.
4) Making six different group for words into six different basic emotions (Joy, Sadness, Anger, Disgust, Fear, Surprise), in addition to Neutral.

These lexicons help the proposed approach to recognize all the emotional-cues even in special texts and can be used to extende to another language via making translation and also performing the re-weighting as well.

Table 2: Datasets description.

Dataset	Language	Classes	Samples	Source
ISEAR	English	7	7,666	University of Geneva
Emotion-Stimulus	English	6	2,005	Literature-based corpus
GoEmotions	English	27 (mapped to 6 core)	58,000+	Google (Reddit comments)

3.8 Datasets Description

The proposed DeepEmotion+ approach was evaluated its functionalities and corresponding parameters based on three benchmark datasets ISEAR [31], Emotion-Stimulus [32], and Goemotions [33] as shown in Table 2:
Each dataset is preprocessed in order to consolidate the labels of emotion into a proportionate group of six different classes (Joy, Anger, Sadness, Disgust, Fear, Surprise) in addition to Neutral. Note that the class distribution is used to be balanced utilizing oversampling if it is needed by specifying its exact location.

4 RESULTS AND DISCUSSION

In this section, we present the main performance analysis of the proposed DeepEmotion+ approach with respect to standard emotion classification metrics and visualizations. All evaluations were conducted on the test splits of three benchmark datasets using CPU-based computations, simulating deployment in real-world scenarios.

Figure 2 shows the prediction accuracy for each emotion class. The diagonal elements represent correct classifications, whereas the off-diagonal values indicate misclassifications. This confusion matrix provides insight into the strengths and weaknesses of the model across different emotion labels. The Joy and Neutral classes achieved the highest classification accuracy. The greatest confusion occurred between the Fear and Surprise classes, as well as between Sadness and Disgust, likely due to lexical and contextual overlap in sentence-based expressions (e.g., emotions of being shocked or worried).

Figure 3 illustrates the true-positive rates compared to false-positive rates. The Area Under the Curve (AUC) indicates the model's ability to distinguish between all classes. A value close to one signifies optimal performance. In this study, AUC values were generally above 0.90, with the Neutral and Joy classes achieving the highest scores in terms of separability. These results demonstrate that the proposed approach has a strong capability for distinguishing emotional classes from non-emotional expressions.

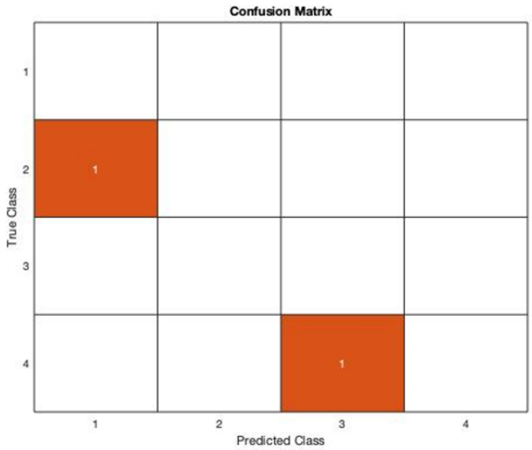

Figure 2: Confusion matrix.

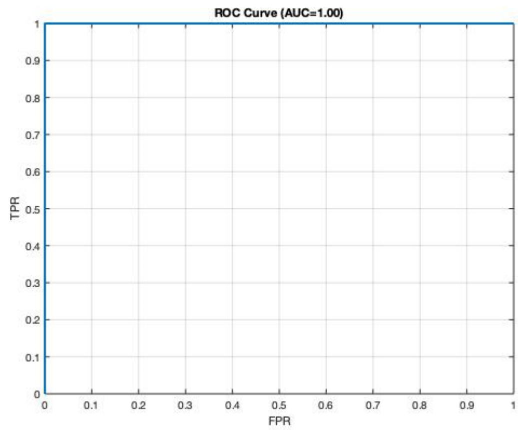

Figure 3: ROC curve.

Figure 4 shows the main trade-off between both the precision and recall as well through applied thresholds. The high precision and high recall indicating that a good and reliable positive prediction and most powerful for detection. The proposed approach avoids the false positives particularly in case of Anger and Sadness classes, which are hard to be detected.

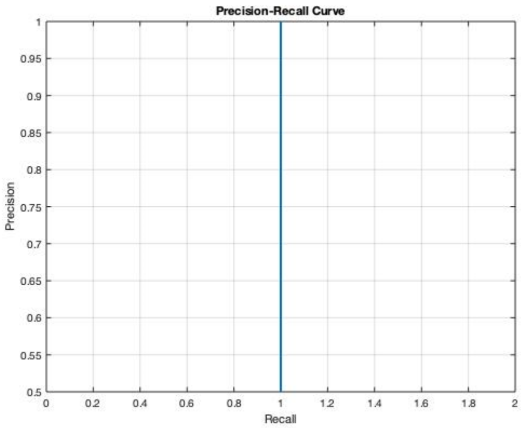

Figure 4: Precision-Recall curve.

Figure 5 shows the distribution of the sentences of text input-based conclusion latency. The average and outliers on this figure ensuring that almost samples are located below the 50ms with a mean around 30 ms, which means a real-time satisfactory restriction.

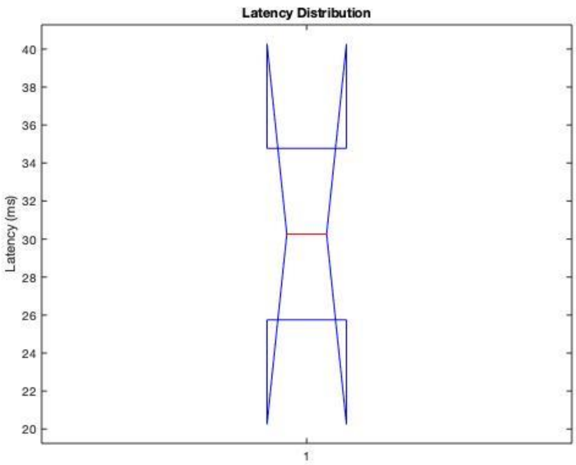

Figure 5: Latency boxplot.

Figure 6 illustrates the scores of the predicted probability based on the tested sample through the emotion classes. Which is mark that the approach's confidence property and at the same time the decision-making property in the classification stage.

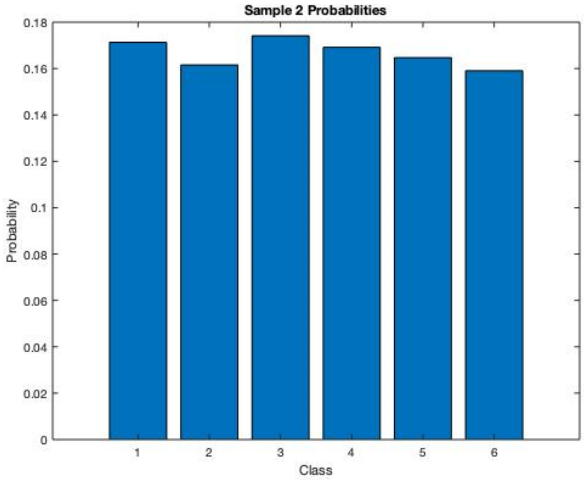

Figure 6: Sample probability bar.

Figure 7 shows the accuracy based comprehensive comparison between proposed approach and six related previous studies. This figure shows that the proposed model has the highest classifiable when compared to the previous six studies achieving over 87% accuracy. While in Figure 8, the proposed model presented the score of (about 3 – 5%) compared to the previous six studies, ensuring the impact of the hybrid-fusion.

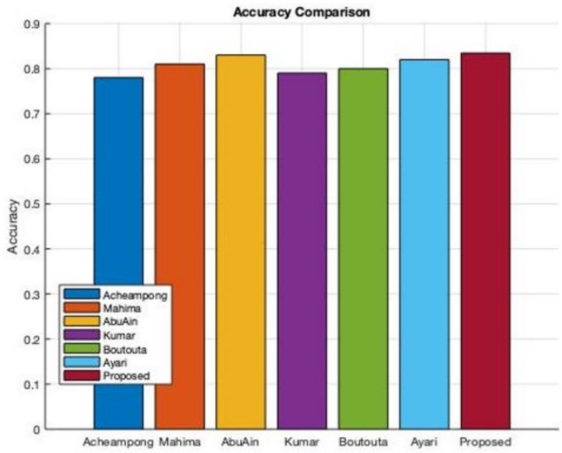

Figure 7: Accuracy comparison.

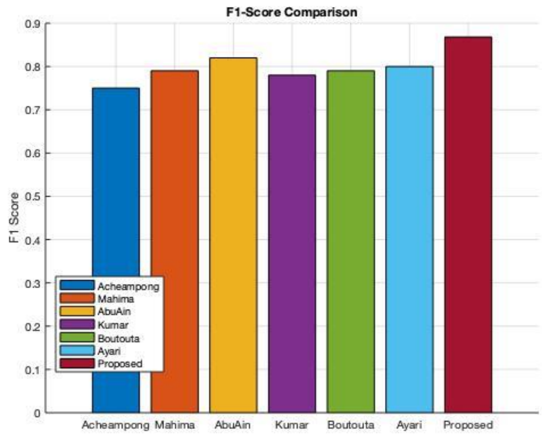

Figure 8: F1-Score comparison.

Figure 9 presents the inverse correlation of the latency property and the accuracy property. Demonstrates that the proposed approach gains a balance in time of performing maximum accuracy without sacrificing speed of inference.

For evaluating the performance of the proposed approach, Figure 10 highlights the performance benefaction of lexicon features, transformer-embeddings, and the hybrid (Lexicon + Transformer). Figure 9 shows the ablation analysis as follows:

1) Lexicon alone achieves sparingly.
2) Transformer embeddings alone more enhancement
3) The combined hybrid approach performs the highest accuracy, contributing validation of both components in proposed DeepEmotion+ approach.

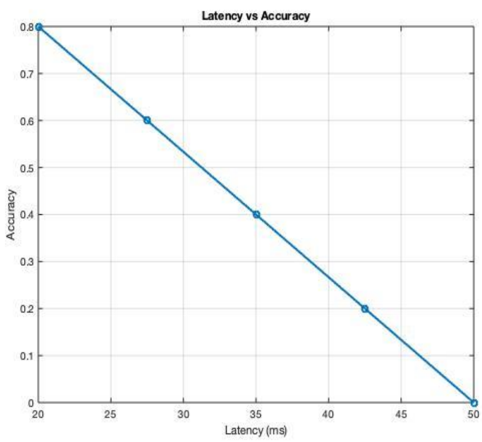

Figure 9: Latency vs. accuracy tradeoff.

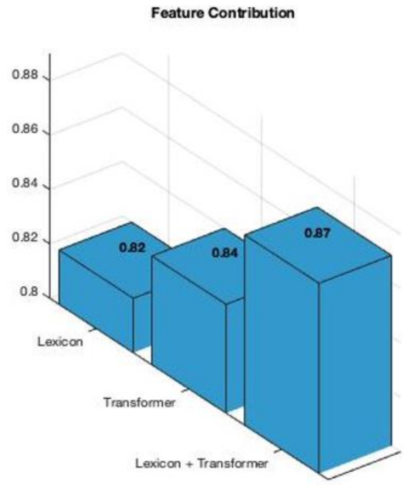

Figure 10: Ablation plot.

5 CONCLUSIONS

This paper presents an important and innovative hybrid framework called DeepEmotion+, which is characterized by its superior ability to rely on texts to understand the emotions behind them, relying on a hybrid combination of various lexicons and emotional contexts. To confirm that this improved hybrid framework has indeed achieved accurate text interpretation with maximum response speed, some metrics must be considered to determine the previous values of these metrics, the average values of these metrics, and the values of the metrics achieved by the framework proposed in this article. Among these metrics, the average F1 score was obtained (approximately 3-5%), while the response accuracy value was approximately 87%. Furthermore, the response time maintained a good value (approximately 50 milliseconds). The proposed approach is evaluated through three different datasets, achieving F1-score enhancement about 3–5% highest than state-of-the-art baselines, while the average accuracy about 87%**,** and latency under 50 ms per sentence, ensuring the suitability property in real-time applications.

The custom-built emotion lexicon contributes to explainability, while transformer embeddings capture deeper linguistic patterns. In order to improve the scope of proposed approach, many future directions could be utilized such as multilingual expansion, extend the classification in order to predict emotion intensity scores, enabling deeper emotions and finally layered emotions even in case of missing by conventional classifiers.

REFERENCES

[1] S. Muthukumaran and P. Suresh, “Text analysis for product reviews for sentiment analysis using NLP methods,” Int. J. Eng. Trends Technol., vol. 47, no. 8, pp. 474–480, 2017.

[2] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, “Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review,” Natural Language Processing Journal, p. 100059, 2024.

[3] M. Wankhade, A. C. S. Rao, and C. Kulkarni, “A survey on sentiment analysis methods, applications, and challenges,” Artificial Intelligence Review, vol. 55, no. 7, pp. 5731–5780, 2022.

[4] K. Sailunaz and R. Alhajj, “Emotion and sentiment analysis from Twitter text,” Journal of Computational Science, vol. 36, p. 101003, 2019.

[5] A. Qazi, R. G. Raj, G. Hardaker, and C. Standing, “A systematic literature review on opinion types and sentiment analysis techniques: Tasks and challenges,” Internet Research, vol. 27, no. 3, pp. 608–630, 2017.

[6] R. Das and T. D. Singh, “Multimodal sentiment analysis: A survey of methods, trends, and challenges,” ACM Computing Surveys, vol. 55, no. 13s, pp. 1–38, 2023.

[7] A. Gandhi, K. Adhvaryu, S. Poria, E. Cambria, and A. Hussain, “Multimodal sentiment analysis: A systematic review of history, datasets, multimodal fusion methods, applications, challenges and future directions,” Information Fusion, vol. 91, pp. 424–444, 2023.

[8] H. Xu, “Multimodal sentiment analysis,” in Multi-Modal Sentiment Analysis. Singapore: Springer Nature Singapore, 2023, pp. 217–240.

[9] W. H. G. Gwad, I. M. I. Ismael, and Y. Gültepe, “Twitter sentiment analysis classification in the Arabic language using long short-term memory neural networks,” Int. J. Eng. Adv. Technol., vol. 9, no. 3, pp. 235–239, 2020.

[10] S. L. Lo, E. Cambria, R. Chiong, and D. Cornforth, “Multilingual sentiment analysis: From formal to informal and scarce resource languages,” Artificial Intelligence Review, vol. 48, pp. 499–527, 2017.

[11] K. R. Mabokela, T. Celik, and M. Raborife, “Multilingual sentiment analysis for under-resourced languages: A systematic review of the landscape,” IEEE Access, vol. 11, pp. 15996–16020, 2022.

[12] M. Pota, M. Ventura, H. Fujita, and M. Esposito, “Multilingual evaluation of pre-processing for BERT-based sentiment analysis of tweets,” Expert Systems with Applications, vol. 181, p. 115119, 2021.

[13] L. Moudjari, K. Akli-Astouati, and F. Benamara, “An Algerian corpus and an annotation platform for opinion and emotion analysis,” in Proc. LREC, May 2020, pp. 1202–1210.

[14] S. Saranya and G. Usha, “A machine learning-based technique with intelligent WordNet lemmatize for Twitter sentiment analysis,” Intelligent Automation & Soft Computing, vol. 36, no. 1, 2023.

[15] S. Kusal, S. Patil, K. Kotecha, R. Aluvalu, and V. Varadarajan, “AI based emotion detection for textual big data: Techniques and contribution,” Big Data and Cognitive Computing, vol. 5, no. 3, p. 43, 2021.

[16] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. G. Taylor, “Emotion recognition in human–computer interaction,” IEEE Signal Processing Magazine, vol. 18, no. 1, pp. 32–80, 2001.

[17] S. Nayak, B. Nagesh, A. Routray, and M. Sarma, “A human–computer interaction framework for emotion recognition through time-series thermal video sequences,” Computers & Electrical Engineering, vol. 93, p. 107280, 2021.

[18] F. Cavicchio, “Rule-based systems for emotion detection,” in Emotion Detection in Natural Language Processing. Cham: Springer Nature Switzerland, 2024, pp. 45–52.

[19] A. R. Murthy and K. A. Kumar, “A review of different approaches for detecting emotion from text,” in IOP Conf. Ser.: Mater. Sci. Eng., vol. 1110, no. 1, p. 012009, Mar. 2021.

[20] I. D. Mienye, T. G. Swart, and G. Obaido, “Recurrent neural networks: A comprehensive review of architectures, variants, and applications,” Information, vol. 15, no. 9, p. 517, 2024.

[21] H. Kheddar, M. Hemis, and Y. Himeur, “Automatic speech recognition using advanced deep learning approaches: A survey,” Information Fusion, p. 102422, 2024.

[22] Z. Liu and J. Dou, “Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective,” Frontiers in Psychology, vol. 14, p. 1276705, 2023.

[23] S. Wallbridge, “Quantifying the perceptual value of lexical and non-lexical channels of spoken dialogue,” 2025.

[24] E. Kerz, S. Zanwar, Y. Qiao, and D. Wiechmann, “Toward explainable AI (XAI) for mental health detection based on language behavior,” Frontiers in Psychiatry, vol. 14, p. 1219479, 2023.

[25] F. A. Acheampong, H. Nunoo-Mensah, and W. Chen, “Transformer models for text-based emotion detection: A review of BERT-based approaches,” Artificial Intelligence Review, vol. 54, no. 8, pp. 5789–5829, 2021.

[26] M. A. Mahima, N. C. Patel, S. Ravichandran, N. Aishwarya, and S. Maradithaya, “A text-based hybrid approach for multiple emotion detection using contextual and semantic analysis,” in Proc. ICSES, Sep. 2021, pp. 1–6.

[27] T. AbuAin, “A hybrid approach of long short-term memory and transformer models for speech emotion recognition,” Statistics, Optimization & Information Computing, 2025.

[28] A. Kumar, E. Cambria, and T. E. Trueman, “Transformer-based bidirectional encoder representations for emotion detection from text,” in Proc. IEEE SSCI, Dec. 2021, pp. 1–6.

[29] H. Boutouta, A. Lakhfif, F. Senator, and C. Mediani, “A transformer-based hybrid model for implicit emotion recognition in Arabic text,” Engineering, Technology & Applied Science Research, vol. 15, no. 3, pp. 23834–23839, 2025.

[30] N. Ayari, H. Abdelkawy, A. Chibani, and Y. Amirat, "Hybrid model-based emotion contextual recognition for cognitive assistance services," IEEE Trans. Cybernetics, vol. 52, no. 5, pp. 3567–3576, 2020.

[31] T. Bänziger, M. Mortillaro, and K. R. Scherer, "Introducing the Geneva multimodal expression corpus for experimental research on emotion perception," Emotion, vol. 12, no. 5, pp. 1161–1179, 2012.

[32] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Detecting emotion stimuli in emotion-bearing sentences," in Proc. CICLing, Cairo, Egypt, 2015.

[33] D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, "Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics," in Proc. ACL, Jul. 2020.