

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

Context-aware bidirectional recurrent network with transformer encoding for text-based emotion detection

Mohammad Mahzari^{1*}, Noah N. Alsager¹ & Anwer Mustafa Hilal²

¹Department of English Language and Literature, College of Science & Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia

²Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, Saudi Arabia

*Correspondence: m.mahzari@psau.edu.sa

ABSTRACT

Emotion detection is the process of recognizing the emotional state expressed through textual content. It has become an important task in Natural Language Processing (NLP) due to the growing amount of text-based communication. Moreover, the NLP techniques help to analyze textual information and classify it into different emotional categories. Traditional machine learning methods generally rely on manually selected features, while deep learning models can automatically learn features, they may not consistently capture the diverse linguistic patterns present in textual data. This may affect the accurate identification of emotions across different types of textual content. This manuscript introduces a Transformer-based Bidirectional Recurrent Network for Emotion Detection (TBiRN-ED), designed to learn complex textual representations for distinguishing different emotion categories. The proposed model employs Term Frequency-Inverse Document Frequency with Category-Distribution Weight, which integrates both term importance and category-specific information into the feature representation. These representations are further passed to the Transformer-Based Bidirectional Recurrent Network model, where transformer-based contextual encoding and bidirectional long short-term memory are combined to capture both contextual relationships and sequential dependencies among words. In addition, the Adan optimizer is utilized to improve the training process and parameter updates. The experimental analysis of the proposed TBiRN-ED framework is evaluated using the Emotions dataset and the Emotions Dataset for NLP, achieving accuracy of accuracy of 96.42 % and 95.70 % compared with existing methodologies. Additionally, an ablation study is conducted to examine the contribution of each component of TBiRN-ED to the overall classification performance. The results demonstrate its effectiveness in distinguishing different emotional categories from textual input, supporting its suitability for text-based emotion detection.

KEYWORDS: emotion detection, natural language processing, deep learning, adan optimizer, bidirectional long short-term memory

Research Journal in Advanced Humanities

Volume 7, Issue 4, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 29 July 2026

Accepted: 14 September 2026

Published: 18 October 2026

HOW TO CITE

Mahzari, M., Alsager, N. N., & Hilal, A. M. (2026). Context-aware bidirectional recurrent network with transformer encoding for text-based emotion detection. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/s9bme461>



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2026 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. Introduction

In human behaviour, emotions have a significant influence on social connection, perception, understanding, and decision-making processes. Moreover, humans have the capacity to recognize emotions through multiple forms such as voice tone, facial expression, and body language [1]. However, machine-based models still face challenges in learning and classifying emotions due to the complexities of emotional expression and contextuality [2]. Emotion detection and classification are essential in today's data-driven world due to a wide range of textual data generated across multiple applications such as emails, social media, forums, and feedback [3]. Moreover, speech or image-based emotion detection and textual emotion detection rely on acoustic and visual signs and are increasingly in demand across various applications such as mental health assessment, customer experience and support, social media monitoring, product and market research, and chatbots [4]. In recent times, emotion detection has gained considerable attention in the field of Natural Language Processing (NLP) due to the rapid growth of educational, business, and psychology platforms [5]. Emotion detection performs a significant role in content creation and business organizations by enabling them to understand customers and develop products according to user preferences, thereby improving human-computer communication and user experience [6].

Recently, the rapid advancement in Machine Learning (ML) and NLP has emerged and provided promising solutions with accurate detection of emotions [7]. In terms of emotion detection, NLP-based approaches are used to detect emotions through textual context and identify the type of emotion expressed. ML-based models involve existing approaches that can capture patterns from data and generate predictions without explicit programming [8]. Over the past few years, Deep Learning (DL)-based methods are extensively adopted in research, specifically Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANNs) [9]. These DL-based algorithms have the capacity to classify and identify emotional cues in complex textual context, thereby effectively recognizing emotions [10]. Their capabilities in in-context learning and strong generalization without additional training have encouraged research into their potential for emotion identification and further use as an annotation mechanism [11]. By influencing linguistic patterns across a group of texts, emotion detection and classification can be performed more precisely and effectively. Therefore, an automated mechanism is required for the accurate extraction of linguistic signs, including sarcasm, accent, and contextual tones, to provide precise, effective, and robust solutions for emotion assessment among various contexts [12]. Although DL techniques have shown strong performance in text-based emotion detection, they still face challenges in capturing long-range contextual dependencies and subtle emotional cues. Nevertheless, DL models can effectively learn complex semantic representations from textual data. Hence, a robust framework is required to jointly capture contextual and sequential information for accurate emotion classification. To overcome these limitations, this research work presents a Transformer-based Bidirectional Recurrent Network for Emotion Detection (TBiRN-ED) framework, designed to improve the representation and classification of emotion-related text. The key contributions of this study are summarized as follows:

- Utilize a Term Frequency-Document Frequency with Category-Distribution Weight (TF-IDF-CDW) framework to identify the importance terms and capture their category-specific distribution in text data.
- Design a Transformer-Based Bidirectional Recurrent Network (TBiRN) to learn contextual representations and sequential patterns from text to improve the discrimination of different emotion categorizes.
- Integrate the Adam optimizer to enable adaptive parameter updates and support stable convergence of the model.
- Validate the experimental outcomes using Emotions dataset and the Emotions Dataset for NLP, and the comparison analysis demonstrates that the proposed TBiRN-ED method gains promising results compared to existing methodologies.

II. Related Work

In this section, a concise overview of existing literatures on text-based emotion detection is discussed, along with highlighted research gaps. Brezina [13] analyzed the role of corpus linguistics in addressing AI-based approaches such as large language models (LLMs). Furthermore, the authors analyzed various human-computer interaction approaches. These approaches can analyse the decisions that are made by the machines and can

deliver meaningful insights. These approaches will highlight the difference between form and meaning, with meaning based on social context and human communication. Cheung and Crosthwaite [14] developed a novel AI-based application called CorpusChat, which is mainly designed for university scholars to improve their writing skills. This application provides custom-made corpus-informed chatbots, thereby supporting university-level students to enhance their domain-specific theoretical verbal knowledge. Hamed and Alqurashi [15] designed an innovative Applied Legal Linguistics (ALL) by incorporating various theories, studies, and analyses. This framework is mainly designed to emphasize limitations such as communication, translation, and pedagogic issues of the authorized language employed in multilingual legal contexts. Furthermore, the authors analyzed the limitations faced by Arab students in the law department in the field of English because they translate legal texts among structurally and traditionally different legal systems. Ahmed and Abbas [16] introduced a novel hybrid framework called DeepEmotion+ that collects customized emotional context along with a transformer-based learning approach. This framework is mainly designed to improve the reliability, precision, and speed of the emotion classification task. The authors integrated Lexical-Preprocessing and Transformer-Classification into the framework for practical emotional detection in English text.

Kusal et al. [17] explored the roles and advantages of emoji in automated detection of emotions in text messages, thereby emphasizing on social media interaction. The authors further designed a hybrid approach that integrates DL-based models for automated detection of emotions. In addition, the authors integrated existing neural network approaches to learn the fused methods and emoji representations. Singh et al. [18] introduced an innovative hybrid approach by integrating DL and ML-based models for detecting text-based emotion. Furthermore, the authors integrated an NLP-based approach with a hybrid framework for effective detection of text-based emotion. The efficiency of this framework is determined by training on multiple datasets, which consist of sentences, tweets, and dialogues. Kusal et al. [19] introduced multiple AI-based approaches for scalable and automated detection of text-based emotion. The authors further analyzed various approaches, from ML-based methods to cutting-edge pre-trained approaches and DL-based methods. This framework provides an automated approach for conversational agents in text-based emotion detection, thereby highlighting natural human-machine communication.

A. *Research Gaps*

Existing text-based emotion detection approaches mostly face challenges in capturing both global contextual information and sequential dependencies within textual data. Existing DL-based approaches can miss essential contextual relationships during the processing of prolonged text sequences. Transformer-based approaches can efficiently form contextual representations but they cannot effectively sustain sequential patterns. Among these, recurrent approaches can effectively learn sequential dependencies but have complexity in demonstrating broader contextual relationships. Moreover, traditional methods still face challenges in differentiating similar emotional classes and maintain refined semantic variations in text. In addition, the incorporation of contextual and sequential representation learning has not yet been sufficiently analysed for consistent emotion classification. These challenges and limitations represent the necessity for a unified framework, which can capture complementary contextual and sequential features from the given textual data. To overcome these limitations, this manuscript proposes a TBiRN-ED, which is mainly developed to capture intricate textual representations for differentiating various emotion categories.

III. Materials and Methods

This section presents the datasets, preprocessing steps, and proposed framework used for text-based emotion detection.

A. *Dataset Description*

The empirical assessment of the TBiRN-ED framework is performed using Emotion Dataset (Dataset 1) Emotion dataset [20] and Emotions Dataset for NLP (Dataset 2) [20]. A brief overview of these two datasets is described as follows.

Dataset 1: Emotions describes how people respond to interactions, experiences, and situations in their day-to-day lives. They can impact the way individuals think, communicate, and make decisions. Various emotions

may occasionally be shown through the same words, making them complex to differentiate from text alone. Consequently, interpreting the context and flow of language is significant for identifying the emotion carried in the pieces of texts. It contains 175707 samples, and six emotion categories are represented in Table 1.

TABLE I. DATASET DETAILS FOR EMOTION

Class Labels	Class Count
Sadness	48475
Joy	56427
Love	13822
Anger	22927
Fear	19085
Surprise	14972
Total	175707

Dataset 2: The Emotions dataset is developed to support NLP models comprehend emotions shown in written text. It comprises text instances labeled with diverse emotions, like surprise, joy, love, anger, sadness, and fear. These labeled samples permit a model to capture how diverse emotions are generally expressed in languages. The dataset is utilized for tasks like text-based emotion recognition, sentiment analysis, and emotion classification. It presents a total of 20000 instances, as represented in Table 2.

TABLE II. DATASET DETAILS ON EMOTION DATASET FOR NLP

Class Labels	Class Count
Joy	6761
Sadness	5797
Anger	2709
Fear	2373
Love	1641
Surprise	719
Total	20000

B. Proposed Methodology

The proposed model follows a multi-stage pipeline for identifying text-based emotions using deep learning methodologies. The overall process of the TBiRN-ED model is presented in Fig. 1. Initially, the input text is preprocessed to reduce textual variations and retain relevant linguistic information through lowercasing, tokenization, contraction expansion, lemmatization, and Part-of-Speech tagging. In addition, TF-IDF-CDW is used to extract discriminative term information associated with different emotion categories. Furthermore, the TBiRN learns the contextual relationships and sequential dependencies within textual representations, while the Adan optimizer improves the efficiency of model parameter optimization. The detailed description of each component is described below.

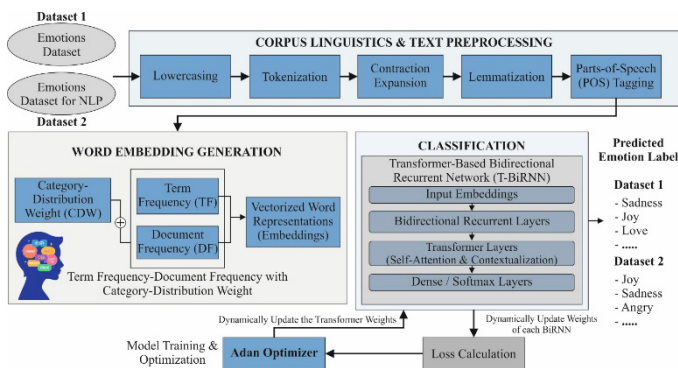


Fig. 1. Overall process of the proposed TBiRN-ED model

C. Text Preprocessing

The preprocessing stage prepares the input text for further analysis by reducing textual inconsistencies and preserving relevant linguistic information through lowercasing, tokenization, contraction expansion, lemmatization, and Part-of-Speech tagging [22].

Lowercasing: For maintaining consistency, every text is transformed into lowercase while avoiding treating similar words with distinct capitalization as separate terms. For example, “Happy”, “HAPPY”, and “happy” are transformed to “happy”. This minimize unwanted changes in the vocabulary and delivers a unique representation in text.

Tokenization: All samples of text are categorized into specific tokens or words during feature extraction; this separates the sentence into essential communicated units that can be processed. For example, “I feel very happy today” is tokenized into “I”, “feel”, “very”, “happy”, and “today”. For feature representation, the respective tokens serve as the basic units for then linguistic investigation.

Contraction Expansion: Common contractions are extended into the entire forms while preserving the deliberate purpose of the text. For example, “can’t” is transformed into “cannot”, and “didn’t” to “did not”, and “I’m” to “I am”. This offers a set of words and helps to sustain meaningful information. It also prevents to compressed forms prior to treated as individual vocabulary items.

Lemmatization: Words are reduced to their proper base or dictionary form, which considers their linguistic meaning. For example, “running” denotes “ran”, and “runs” represents to “run”. This decreases changes in similar sentences or words and generates a more reliable vocabulary. Accordingly, related word forms can be presented through a widespread linguistic root.

Parts-of-Speech Tagging: All token is assigned its grammatical class, namely verb, noun, adverb, adjective, conjunction or pronoun. For example, in the sentence “She feels extremely happy”, “feels” is represented as a verb, “extremely” denotes an adverb, and “happy” depicts an adjective. In the representation and classification process, the grammatical data helps to retrain the linguistic role of words. This is especially suitable for identifying words that hold significant emotional features in a sentence.

D. Text Representation using TF-IDF-CDW

To represent the emotional content of textual information efficiently, TF-IDF-CDW is used to allocate high weights to terms that are both significant in each document and distinctive across emotion classes [23]. While standard TF-IDF learns the document-level significance of a term, it does not explicitly consider how well a term distinguishes an individual class. Integrating category distribution weighting (CDW) addresses this by providing category-aware term weighting, enabling emotionally distinctive words to receive higher significance throughout classification. The Term Frequency (TF) computes how frequently a term happens in a document and is expressed as:

Here, tf represent the count of occurrences of t in d , and the denominator shows the overall count of t in D . The Inverse Document Frequency (IDF) determines the distinctive significance of a term through the corpus and can be represented as:

Let N imply the overall count of documents in the corpus, and n_d represent the number of documents covering t . The standard TF-IDF score is further attained by:

Even though TF-IDF efficiently learns the significance of terms at the document level, it does not explicitly consider their distribution across all emotion classes. To tackle these challenges, CDW is proposed to highlight terms that show robust associations with individual classes.

Let C represent the overall count of classes, and p_d implies the probability of documents that associated with category c . The category distribution entropy of term t is represented as:

The standardized category distribution is:

Hence, the CDW can be represented as:

This weighting allocates a maximal value to terms focused in individual classes and a minimal value to terms distributed comparatively uniformly through classes. Consequently, category-specific terms receive further attention, while terms with restricted discriminative ability are reduced. Finally, the proposed TF-IDF-CDW representation are computed as:

The final TF-IDF-CDW representation obtains category-aware textual characteristics that highlight emotion-discriminative terms, thus providing relevant inputs for the following emotion classification procedure.

E. TBiRN-based Emotion Classification

The emotional meaning of a word frequently relies on its surrounding context and its position within the sentence during text-based emotion recognition. Therefore, it is important to capture both the contextual dependencies among words and the sequential relationship within the text to differentiate emotions with same linguistic patterns [24]. To overcome these challenges, the TBiRN incorporates a Transformer encoder with a Bidirectional Long Short-Term Memory (BiLSTM) network. BiLSTM model the sequential data from both forward and backward directions, whereas the Transformer extracts contextual relationships using self-attention mechanisms.

An embedding matrix E represents the first input sequence; here, d refers to the embedding dimension, and n indicates the sequence length. The Transformer encoder accepts these representations and produces context-enriched characteristics. For every input representation, the query, key, and value matrices are computed by

here W_q , W_k , and W_v depicts learnable projection matrices. The scaled dot-product attention is subsequently determined by

Here, d_k signifies the key vector dimension. Based on their contextual relevance, this operation allocates diverse weights to the input tokens. Multiple attention functions are operated in parallel during multi-head attention, and their outputs are combined by

here W_o signifies the output projection matrix, and h specifies the number of attention heads. The resulting representation is then forwarded using a feed-forward network:

here, b_1 and b_2 denotes bias vectors, σ implies the activation function, and W_1 and W_2 means weight matrices. The Transformer then generates a sequence of contextual representations H . The contextual representations attained from the Transformer are further passed into the BiLSTM layer for capturing sequential relationships in both directions. The BiLSTM consists of dual LSTM networks, while the forward LSTM performs the sequence from left to right, and the backward LSTM perform it from right to left. This ensures every representation to combines data from both preceding and succeeding words. Consider the given input x_t , the input, forget, and output gates are evaluated at time period t ,

Here, i_t and f_t demonstrate the input, forget, and output gates, correspondingly. The candidate cell state has been defined as

The cell state is then updated by

here $\odot$ represents element-wise multiplication. The hidden state has been calculated by

For obtaining the forward and backward hidden representations, represented by H_f and H_b . The similar functions are operated in the reverse direction. These representations are combined to form the final bidirectional representation:

here implies concatenation. The resulting representation compresses contextual data attained from both directions of the sequence within the text. It provides the classification layer for designating the respective emotion class. The resulting bidirectional representation is passed to a fully connected layer to classify the text into five emotion classes: joy, sadness, anger, fear, and surprise. The Softmax function is then used to obtain the probability of each class as follows:

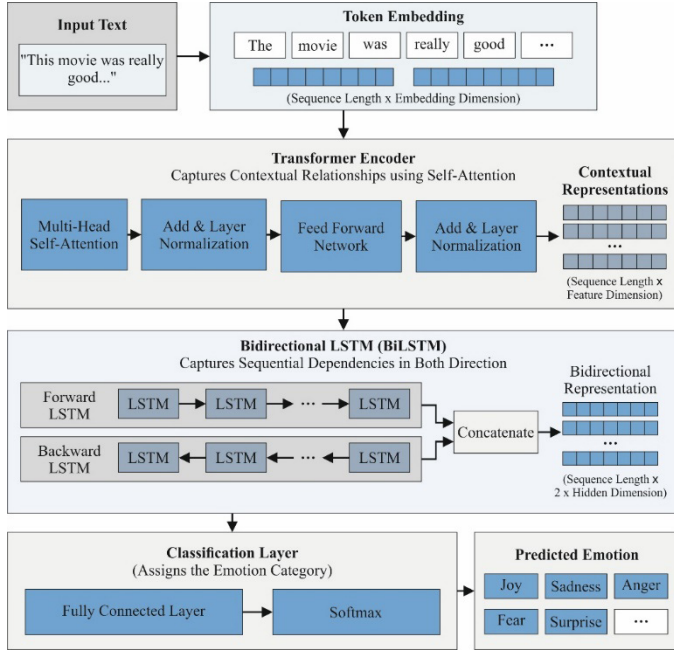


Fig. 2. Structure of TBiRN module

Here, represents the final bidirectional representation, while and denote the weight and bias of the corresponding output class. Fig. 2 presents the structure of TBiRN module.

F. Adan-based Optimization Strategy

Text-based emotion detection requires effective parameter optimization to learn contextual and sequential patterns from textual representations. Therefore, the Adan optimizer is employed to optimize the proposed network during training [25]. Adan adaptively updates the model parameters by considering the current gradient, the gradient difference between consecutive iterations, and the second-order moment of the gradient. This helps provide stable parameter updates while improving the learning process of the deep network. Let denote the stochastic gradient at iteration , and denote the model parameters. The first momentum estimate is calculated as

The gradient different is estimated using

The second-order moment is then obtained in

The adaptive learning rate is calculated by

The network parameters are updated as

Where , , and are the momentum coefficients, implies the base learning rate, represents a smaller stability constant, and represents the weight-decay coefficients.

IV. Experimental Analysis

This section presents the experimental results and analysis of the proposed TBiRN-ED framework using the Emotions dataset and the Emotions Dataset for NLP to evaluate its performance.

A. Result Analysis on Dataset 1

Fig. 3 illustrates the confusion matrix of the TBiRN-ED framework with 70:30 for Dataset 1. In both phases, the majority of the higher values are highlighted on the main diagonal, which means that most of the instances are appropriately recognized. The model attains especially optimistic outcomes for Joy and Sadness, whereas Love and Surprise demonstrate moderately more confusion with alternative emotions. The testing matrix follows the same pattern as the training matrix, indicating that the model retains reliable performance once applied to unobserved data.

Table 3 portray the emotion classification results of the TBiRN-ED system on 70:30 for Dataset 1. The outcome implied that the proposed TBiRN-ED method achieves the overall accuracy 96.51% with average MCC of $94.69\% \pm 2.12$ in training set while in testing set it obtains overall accuracy 96.42% with average MCC of $94.54\% \pm 2.28$, signifying the efficiency of the proposed model and indicates its strong generalization capability.

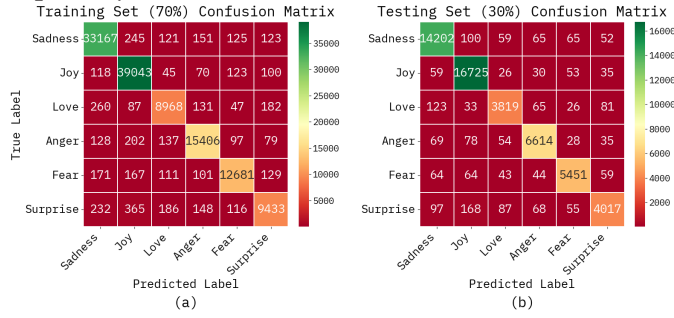


Fig. 3. Confusion matrix (a-b) of the TBiRN-ED approaches with 70:30 for Dataset 1

TABLE III. EMOTION DETECTION PERFORMANCE OF THE TBiRN-ED ALGORITHM FOR 70:30 ON DATASET 1

Classes	Accuracy	Precision	Recall	F1-Score	MCC
Training Set (70%)					
Sadness	98.64	97.33	97.75	97.54	96.60
Joy	98.76	97.34	98.85	98.09	97.18
Love	98.94	93.73	92.69	93.21	92.63
Anger	98.99	96.25	95.99	96.12	95.54
Fear	99.03	96.15	94.92	95.53	94.99
Surprise	98.65	93.90	90.01	91.91	91.20
Average $\pm$ SD	98.84 ± 0.16	95.78 ± 1.47	95.03 ± 2.99	95.40 ± 2.21	94.69 ± 2.12
Overall Accuracy: 96.51					
Testing Set (30%)					
Sadness	98.57	97.18	97.66	97.42	96.43
Joy	98.77	97.42	98.80	98.11	97.21
Love	98.87	93.42	92.09	92.75	92.14
Anger	98.98	96.05	96.16	96.11	95.52
Fear	99.05	96.00	95.21	95.61	95.07
Surprise	98.60	93.88	89.43	91.60	90.87
Average $\pm$ SD	98.81 ± 0.18	95.66 ± 1.52	94.89 ± 3.22	95.26 ± 2.36	94.54 ± 2.28
Overall Accuracy: 96.42					

Fig. 4 shows the ROC and PR curve of the TBiRN-ED system on Dataset 1. Fig.4a represents the training ROC curves, which remain close to the upper-left corner, representing a robust capability of the model to differentiate among the six emotion classes, with AUC values within the range of 0.993 to 0.995. Fig. 4b demonstrates the testing ROC curves, which also show a near-optimal pattern, with the curves almost following the upper-left

region, with AUC values ranging from 0.994 to 0.996. Fig. 4c shows that the training PR curves retain higher precision over an increased recall range, reflecting accurate identification of the diverse emotional categories, with AP values ranging from 0.957 to 0.992, with Joy and Sadness exhibiting especially strong PR performance. Fig. 4d displays the testing PR remains higher across a wide-ranging range of recall values for all six emotions. With AP values between 0.962 and 0.993, the model shows consistent precision–recall performance and better predictive reliability on the test data.

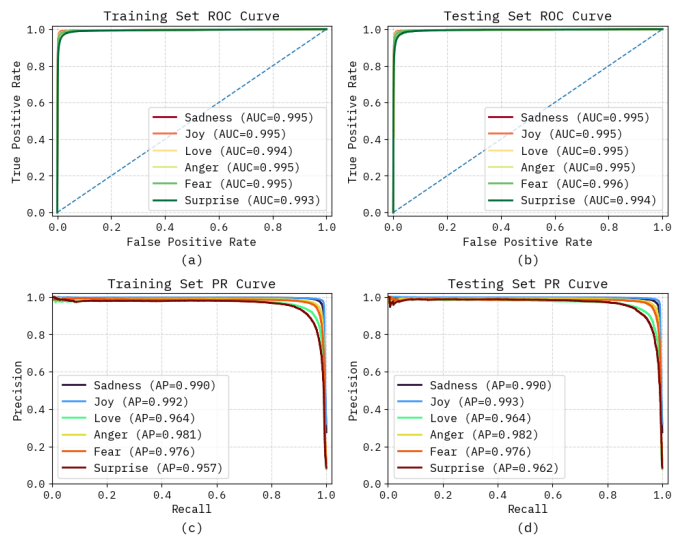


Fig. 4. ROC and PR curves of the TBiRN-ED technique for emotion detection on Dataset 1 (a-b) training and testing ROC and (c-d) training and testing PR curve

B. Result Analysis on Dataset 2

Fig. 5 describes the confusion matrix of the TBiRN-ED method with 70:30 on Dataset 1. The confusion matrices show that the proposed model classifies most emotion instances properly, as indicated by the higher values along the diagonal for these phases. Even though some confusion arises among Anger, Fear, Love, and Surprise, the misclassifications remain relatively low. The identical pattern between the training and testing matrices denotes consistent classification performance and higher generalization of the model.

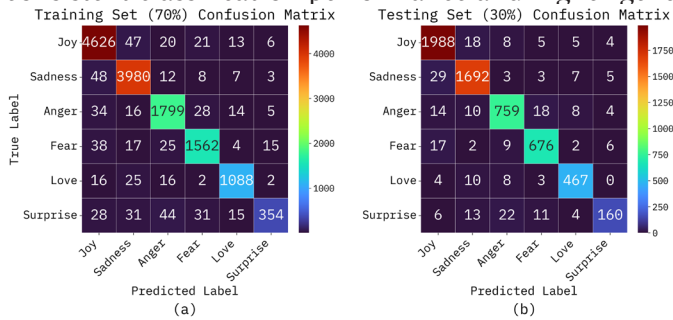


Fig. 5. Confusion matrix of the TBiRN-ED method using 70:30 for Dataset 2

Table 4 exhibit the emotion recognition performance of the TBiRN-ED technique with 70:30 on Dataset 2. The results suggest that the proposed TBiRN-ED framework obtains an overall accuracy of 95.78% with an average MCC of $92.24\% \pm 5.64$ in training. At the same time, it attains overall accuracy of 95.70% with MCC of 92.26 ± 5.27 in testing, demonstrating good classification effectiveness and generalization across the six emotion types.

TABLE IV. EMOTION RECOGNITION RESULT OF THE TBiRN-ED SYSTEM WITH 70:30 ON DATASET 2

Classes	Accuracy	Precision	Recall	F1-Score	MCC
Training Set (70%)					
Joy	98.06	96.58	97.74	97.15	95.69
Sadness	98.47	96.7	98.08	97.38	96.31
Anger	98.47	93.89	94.88	94.39	93.5

Fear	98.65	94.55	94.04	94.3	93.53
Love	99.19	95.35	94.69	95.02	94.58
Sur- prise	98.71	91.95	70.38	79.73	79.83
Average	98.59 $\pm$	94.84 $\pm$	91.63	92.99 $\pm$	92.24
$\pm$ SD	0.34	1.64	$\pm$ 9.63	6.06	$\pm$ 5.64
Overall Accuracy: 95.78					
Testing Set (30%)					
Joy	98.17	96.6	98.03	97.31	95.92
Sadness	98.33	96.96	97.3	97.13	95.96
Anger	98.27	93.82	93.36	93.59	92.59
Fear	98.73	94.41	94.94	94.68	93.96
Love	99.15	94.73	94.92	94.82	94.36
Sur- prise	98.75	89.39	74.07	81.01	80.75
Average	98.57 $\pm$	94.32 $\pm$	92.10	93.09 $\pm$	92.26
$\pm$ SD	0.34	2.48	$\pm$ 8.21	5.56	$\pm$ 5.27
Overall Accuracy: 95.70					

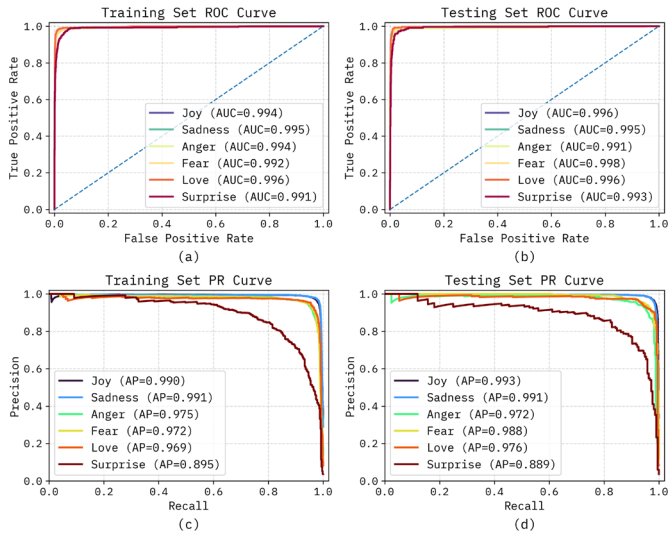


Fig. 6. ROC and PR curves of the TBiRN-ED technique for emotion detection on Dataset 2 (a-b) training and testing ROC and (c-d) training and testing PR curve

Fig. 6 represents the ROC and PR curve of the TBiRN-ED framework for Dataset 2. Fig. 6a demonstrates that the training ROC curves remain closer to the top-left corner for all six emotions, representing strong differentiation between the classes. Furthermore, the AUC values range from 0.991 to 0.996, ensuring reliably higher classification ability. Fig. 6b shows the testing ROC curves retains identical near-optimum pattern, with AUC values between 0.991 and 0.998, demonstrating that the model preserves strong class-separation capability. Fig. 6c illustrates the training PR curves retain higher precision across most of the recall range, representing consistent emotion prediction. The AP values are maximum for Joy and Sadness, whereas Surprise indicates somewhat lesser performance with an AP of 0.895. Fig. 6d exhibits the testing PR curves also retain higher precision for most emotion classes, demonstrating constant predictive performance. Surprise again records the lower AP (0.889), but Joy and Sadness achieve the enriched PR outcomes.

C. Comparison Analysis and Ablation Study

Table 5 and Fig. 7 portray the comparison performance of the TBiRN-ED technique for both Dataset 1 and 2 [26, 27]. On Dataset 1, the findings reveal that the TBiRN-ED methodology gained stable and consistent

outcomes with 96.42% accuracy, 95.66% precision, and 94.89% recall. Moreover, the Bi-LSTM-CNN [27] system performs the lowest results, while the ERNIE-CNN [27], ERNIE-BiLSTM [27], and DistilBERT [26] models obtain moderate performance. Meanwhile, BERT-base [26], XLNet-base [26], and RoBERTa-base [26] methodologies achieve reasonably higher results when compared to other existing algorithms. Similarly, in Dataset 2, the results suggest that the TBiRN-ED approach gained superior outcomes with 95.70% accuracy, 94.32% precision, and 92.10% recall. Bi-LSTM-CNN [27] approach performs least outcome whereas ERNIE-CNN [27], ERNIE-BiLSTM [27], and DistilBERT [26] system gains somewhat higher results. Furthermore, BERT-base [26] and XLNet-base [26] methodologies obtain moderate performance. RoBERTa-base [26] technique achieved slightly lower performance than the proposed system, with an accuracy of 91.42%, demonstrating the efficacy and robustness of the proposed methodology when compared with existing approaches.

Table 6 shows the ablation study analysis of the TBiRN-ED system for Dataset 1. The baseline TBiRN method obtains the lowest 90.84% accuracy, which rises to 92.73% when integrating with TF-IDF model and further increases to 94.58% after adding TF-IDFCDW algorithm. Finally, the proposed TBiRN-ED (TF-IDFCDW + TBiRN + Adan) framework achieves superior performance with 96.42% accuracy, 92.73% precision, 94.58% recall, and a 95.26% F1-score, suggesting that each component contributes positively to the overall efficacy of the proposed architecture.

TABLE V. COMPARATIVE ANALYSIS OF THE TBiRN-ED METHODOLOGY FOR BOTH DATASET 1 AND 2.

Methods	Emotions			Emotions dataset for NLP		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
DistilBERT [26]	88.40	88.30	83.30	87.62	87.41	86.95
BERT-base [26]	90.10	90.00	90.10	89.34	89.12	88.76
XLNet-base [26]	91.00	90.90	91.00	90.48	90.21	89.93
RoBERTa-base [26]	92.45	92.50	92.40	91.67	91.42	91.18
ERNIE-BiLSTM [27]	86.41	86.41	85.97	85.73	85.31	84.96
ERNIE-CNN [27]	78.64	78.64	77.91	77.92	77.36	76.58
Bi-LSTM-CNN [27]	66.29	66.29	67.97	64.87	66.12	63.94
TBiRN-ED	96.42	95.66	94.89	95.70	94.32	92.10

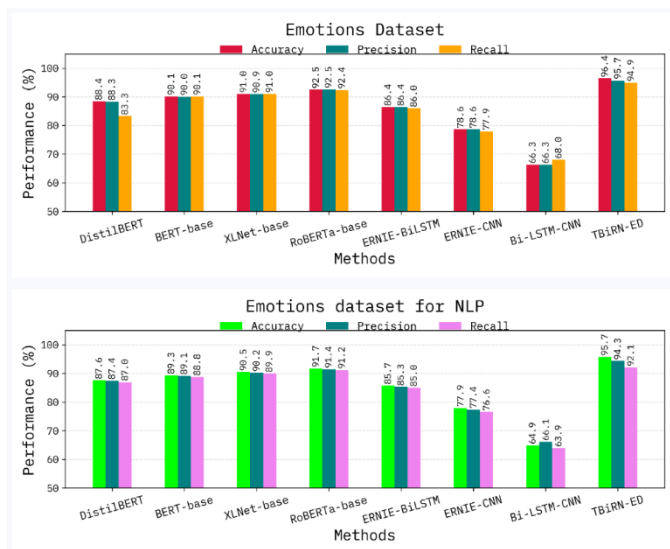


Fig. 7. Comparison performance of the TBiRN-ED algorithm for both Dataset 1 and 2

Table 7 signifies the ablation study of the TBiRN-ED model on Dataset 2. The baseline TBiRN system gained the lowest performance with 89.16% accuracy, which improves to 91.28% by incorporating the TF-IDF algorithm and further increases to 93.47% after combining the TF-IDFCDW methodology. The proposed TBiRN-ED (TF-IDFCDW + TBiRN + Adan) architecture obtains the maximum results, accomplishing 95.70% accuracy, 94.32% precision, 92.10% recall, and a 93.09% F1-score, representing that each component contributes to

better feature representation across emotion classification results.

TABLE VI. ABLATION STUDY RESULT OF THE TBiRN-ED SYSTEM ON DATASET 1

Methods	Accuracy	Precision	Recall	F1-Score
TBiRN	90.84	89.26	87.72	88.46
TF-IDF + TBiRN	92.73	91.48	89.86	90.63
TF-IDFCDW + TBiRN	94.58	93.42	92.17	92.76
TBiRN-ED (TF-IDFCDW + TBiRN + Adan)	96.42	95.66	94.89	95.26

TABLE VII. ABLATION STUDY OUTCOME OF THE TBiRN-ED ALGORITHM ON DATASET 2

Methods	Accuracy	Precision	Recall	F1-Score
TBiRN	89.16	86.72	83.84	85.19
TF-IDF + TBiRN	91.28	88.64	86.17	87.34
TF-IDFCDW + TBiRN	93.47	91.36	89.42	90.31
TBiRN-ED (TF-IDFCDW + TBiRN + Adan)	95.70	94.32	92.10	93.09

D. Discussion

The proposed TBiRN-ED framework integrates contextual with sequential feature learning for effective detection of emotion through text. TF-IDF with Category-Distribution Weight indicates the significance of terms by considering their distribution among emotion classes. The transformer captures contextual relationships between the words, while BiLSTM learns their sequential dependencies. Adan supports to enhance the parameter updates during training, whereas the ablation study demonstrates the contribution of all component to the final performance. The collective features deliver valuable insights for differentiating closely associated emotions in textual data. By capturing both contextual and sequential patterns can also neglects relying on a particular type of feature representation. Generally, the proposed TBiRN-ED attains accuracy of 96.42% on the Emotions dataset and 95.70% on the Emotions Dataset for NLP, thereby demonstrating its capacity to classify multiple emotion classes.

V. Conclusion and Future Work

In this study, we proposed a TBiRN-ED to identify and distinguish diverse emotional categories from textual data. Initially, the proposed model undergoes preprocessing to decrease textual variations and retain relevant linguistic information. Moreover, the TF-IDF-CDW framework is applied to generate category-aware representations of emotion-related text. In addition, the TBiRN utilized to learn informative representations from the generated textual features for distinguishing different emotion categories, while the Adan optimizer facilitates efficient parameter optimization during training by adaptively updating the model parameters to improve learning stability and convergence. Comprehensive experimental outcomes demonstrate the superiority of the proposed TBiRN-ED approach, with greater accuracy of 96.42 % and 95.70 % under the Emotions dataset and the Emotions Dataset for NLP, across various evaluation metrics, demonstrating its effectiveness in accurately distinguishing different emotion categories from textual data. Future work will consider larger and multilingual datasets to test the model across different languages and text sources. The framework can also be extended to multimodal emotion detection by combining text with speech and facial expressions. A lightweight version of the model can be developed for applications with limited computational resources. Further studies can examine the performance of TBiRN-ED on real-time and cross-domain emotion detection tasks.

Data Availability Statement: The data that support the findings of this study are openly available in Kaggle repository at <https://doi.org/10.34740/kaggle/dsv/20875185>, reference number [20].

Acknowledgment: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/ 01/34640).

REFERENCES

- [1] Khan, B., Usman, M. and Binsawad, M., 2026. A novel hybrid model for emotion detection in text through sequential and transformer-based approaches: LSTM enhanced RoBERTa (LER). *Scientific Reports*, 16(1), p.2224.
- [2] Muhammad, S.H., Ousidhoum, N., Abdulmumin, I., Yimam, S.M., Wahle, J.P., Ruas, T., Beloucif, M., De Kock, C., Belay, T.D., Ahmad, I.S. and Surange, N., 2025, July. SemEval-2025 task 11: Bridging the gap in text-based emotion detection. In *Proceedings of the 19th international workshop on semantic evaluation (SemEval-2025)* (pp. 2558-2569).
- [3] Lecourt, F., Croitoru, M. and Todorov, K., 2025, May. 'Only ChatGPT gets me': An Empirical Analysis of GPT versus other Large Language Models for Emotion Detection in Text. In *Companion proceedings of the ACM on web conference 2025* (pp. 2603-2611).
- [4] Junianto, E., Puspitasari, M., Zakaria, S.I., Arifin, T. and Agung, I.W.P., 2025. Emotion Detection in Indonesian Text Using the Logistic Regression Method. *Media Jurnal Informatika*, 17(2), pp.305-316.
- [5] Deshpande, U.U., Shanbhag, S., Sukhasare, A., Dixit, M.M., Patil, R., Sangani, S., Srinivasaiah, S.H., Goudar, S. and Managuli, M., 2025. Multimodal sentiment analysis using image and text fusion for emotion detection. *Discover Computing*, 28(1), p.230.
- [6] Setyoningrum, N.G., SM, N.N.F., Alam, A., Nurdin, A.M., Nursamsi, D.R. and Lodana, M.B., 2025. The Impact of Linguistic Features on Emotion Detection in Social Media Texts. *Innovation in Research of Informatics (Innovatics)*, 7(1).
- [7] Wu, Y., Zhang, S. and Li, P., 2025. Multi-modal emotion recognition in conversation based on prompt learning with text-audio fusion features. *Scientific Reports*, 15(1), p.8855.
- [8] Ngo, A. and Kocoń, J., 2025. Integrating personalized and contextual information in fine-grained emotion recognition in text: A multi-source fusion approach with explainability. *Information Fusion*, 118, p.102966.
- [9] Pugh, R., Wing, C., Huerta, M.X.J., Hernández, Á.M. and Tyers, F., 2025, April. Ihquin tlahtouah in Tetelahtzincocah: An annotated, multi-purpose audio and text corpus of Western Sierra Puebla Nahuatl. In *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)* (pp. 3549-3562).
- [10] Barbara, S.W.Y., Afzaal, M. and Aldayel, H.S., 2024. A corpus-based comparison of linguistic markers of stance and genre in the academic writing of novice and advanced engineering learners. *Humanities and Social Sciences Communications*, 11(1), p.284.
- [11] Perkins, M. and Roe, J., 2024. Decoding academic integrity policies: A corpus linguistics investigation of AI and other technological threats. *Higher Education Policy*, 37(3), pp.633-653.
- [12] Incelli, E., 2025. Exploring the future of corpus linguistics: Innovations in AI and social impact. *International Journal of Mass Communication*, 3, pp.1-10.
- [13] Brezina, V., 2025. Corpus linguistics and AI: #LancsBox X in the context of emerging technologies. *International Journal of Language Studies*, 19(2), p.75.
- [14] Cheung, L. and Crosthwaite, P., 2025. CorpusChat: Integrating corpus linguistics and generative AI for academic writing development. *Computer Assisted Language Learning*, pp.1-27.
- [15] Hamed, D.M. and Alqurashi, N., 2025. A corpus-driven study on legal translation challenges among Arab law students: examining gender and academic level in applied legal linguistics. *Humanities and Social Sciences Communications*, 12(1), p.1065.
- [16] Ahmed, I.M. and Abbas, M.K., 2025. A Hybrid Lexico-Transformer Model for Real-Time Emotion Detection in English Text. *interaction (HCI)*, 16, p.17.
- [17] Kusal, S., Patil, S. and Kotecha, K., 2025. Multimodal text-emoji fusion using deep neural networks for text-based emotion detection in online communication. *Journal of Big Data*, 12(1), p.32.
- [18] Singh, G., Singh, D., Sharma, R. and Bhardwaj, K., 2024, June. Potential approach for text-based emotion detection using NLP coupled with deep learning of sentiment analysis. In *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-5). IEEE.
- [19] Kusal, S.D., Patil, S.G., Choudrie, J. and Kotecha, K.V., 2024. Understanding the performance of AI algorithms in text-based emotion detection for conversational agents. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(8), pp.1-26.

[20] <https://doi.org/10.34740/kaggle/dsv/20875185>

- [21] Bonthu, S., Muddam, H.S.S.G.B., Mudunuri, K.V., Dayal, A., Rao, V.M. and Bolla, B.K., 2023, August. TextRank–Based Keyword Extraction for Constructing a Domain-Specific Dictionary. In International Conference on Cognitive Computing and Cyber Physical Systems (pp. 340-349). Cham: Springer Nature Switzerland.
- [22] Li, C., Xie, Z. and Wang, H., 2025. Short text classification based on enhanced word embedding and hybrid neural networks. *Applied Sciences*, 15(9), p.5102.
- [23] Khan, S., Fazil, M., Sejwal, V.K., Alshara, M.A., Alotaibi, R.M., Kamal, A. and Baig, A.R., 2022. BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection. *Journal of King Saud University Computer and Information Sciences*, 34(7), pp.4335-4344.
- [24] Li, C., Wu, H., Zhang, T., Lu, J. and Li, J., 2024. Lightweight network of multi-stage strawberry detection based on improved YOLOv7-Tiny. *Agriculture*, 14(7), p.1132.
- [25] Azhar, M., Riaz, N., Azeem, W., Dewi, D.A., Amjad, A. and Arman, M., 2026. Explainable Transformer Models for Human Emotion Recognition: A Multi-Method Explainability Study in the Context of Mental Health. *Information*, 17(5), p.496.
- [26] Yang, S., Xing, J., Liu, Z. and Sun, Y., 2025. Short-text sentiment classification model based on bert and dual-stream transformer gated attention mechanism. *Electronics*, 14(19), p.3904.