



Open Access



Check for updates

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

A corpus linguistics enhanced hybrid deep learning for semantic-driven emotion detection on textual data

Noah N. Alsager¹, Mohammad A. Mahzari¹ & Haroon N. Alsager^{1*}¹Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Saudi Arabia*Correspondence: h.alsager@psau.edu.sa

ABSTRACT

Corpus linguistics offers a structured approach to assess huge set of texts (corpora) to uncover language patterns, structural relationships, and semantic interpretations. Within text-based emotion recognition (TER), corpus linguistics facilitates the identification of linguistic features that show underlying emotional expressions. However, the existing machine learning-based models often struggle to learn subtle semantic relationships, contextual relationships, and differences in emotional expression across textual content. Recent advancements in deep learning have demonstrated excellent capabilities for contextual and relational representation learning, yet accurate and robust multi-class emotional classification remains challenging. Therefore, this study proposes a Hybrid Deep Learning Framework for Semantic-driven Emotion Detection (HDLF-SED) on Textual data, aiming to improve learning representation and the accurate separation of multi-class emotional categories. To achieve this, the model uses RoBERTa-Large as a feature embedding technique to learn subtle linguistic patterns and contextual relationships to produce discriminative feature representations for further analysis. The model employs hybrid learning by integrating a temporal convolutional network with a multi-head attention mechanism to jointly capture sequential and local contextual patterns and selectively focus on emotion-relevant features to perform accurate multi-class emotion classification. Furthermore, the RAdam optimization algorithm is used to facilitate more stable and effective optimization during model training and improve classification accuracy. Experimental evaluation is carried out using two benchmark datasets across various performance measures; these results highlight that the proposed HDLF-SED model demonstrates the potential of combining contextual embedding with hybrid learning for more accurate and reliable motion detection and performs multi-class classification.

KEYWORDS: Semantic-Driven Emotion Detection, Corpus Linguistics, Temporal Convolutional Network, Deep Learning, Multi-Head Attention Mechanism

Research Journal in Advanced Humanities

Volume 7, Issue 4, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 25 July 2026

Accepted: 14 September 2026

Published: 17 October 2026

HOW TO CITE

Alsager, N. N., Mahzari, M. A., & Alsager, H. N. (2026). A corpus linguistics enhanced hybrid deep learning for semantic-driven emotion detection on textual data. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/1vhm8s42>



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 recent times, communication plays a significant role across vast geographical areas due to the rapid growth of the internet and network connectivity. Human communication increasingly takes place through textual context across various social media applications and platforms. Among these, emotions play a significant role in daily life, as they enable humans to express their feelings and thoughts [1]. Moreover, Emotion detection through textual context has recently become a significant feature of sentiment assessment across multiple fields, such as customer experience and support, mental healthcare, conversational agents, and social media monitoring [2]. Current sentiment analysis approaches can categorize texts only as positive, negative, or neutral, whereas emotion classification can effectively recognize a wide range of human emotions, including happiness, sadness, anxiety, surprise, fear, anger, and disgust [3]. These valuable insights into emotions are significant in fields like mental health support, in which emotional context can provide early and accurate solutions. Systems can emphasize instant messages by identifying particular emotions, thereby providing warning alerts or personalized support for human intervention when required [4]. The level of accuracy allows systems to provide contextual meaning and understand the emotional condition of the patients. In recent times, Natural Language Processing (NLP) has been widely adopted for effective detection of emotions through text. Compared to traditional classification methods, NLP-based models provide better solutions for classification tasks, such as sentiment and emotion detection [5]. However, these models still face difficulties due to the limited availability of annotated corpora and NLP mechanisms, which restrict the scalability of emotion detection systems across socially diverse contexts [6]. To address these limitations, automated mechanisms are needed for effective emotion detection in resource-constrained environments.

In recent times, progressions in artificial intelligence (AI) and machine learning (ML) have developed and been widely adopted to extend automated solutions for emotion detection through text using NLP-based models. However, existing text-based emotion recognition approaches mostly depend on supervised (ML)-based approaches, which capture features from pre-processed data and train classifiers using the extracted features [7]. Therefore, the concept of feature extraction is an essential attribute of existing ML-based models. Furthermore, deep learning (DL)-based methods have gained significant attention and require less pre-processing, thereby offering an alternative to conventional approaches [8]. Particularly, DL-based approaches, including Auto-Encoders (AE), Convolutional Neural Networks (CNN), transformer-based learning, and Large Language Models (LLM), have been widely adopted in different fields of computer applications, such as NLP and computer vision (CV) [9]. Furthermore, DL-based algorithms apply deep neural networks (DNNs) to capture intricate features from large-scale data representations, thereby offering scalable and reliable systems for emotion recognition. These AI-based models have the capacity to learn from high-dimensional data and provide valuable insights in developing complex models to accurately interpret human emotions [10]. However, the growing application of text-driven communication across various domains and the difficulty of interpreting multiple emotional classes motivate the development of a robust model that has the capability to capture subtle linguistic patterns and offer more accurate classification of diverse emotion categories. Therefore, this study develops an HDLF-SED framework by incorporating RoBERTa-Large-based feature embedding and hybrid learning with RAdam optimization to support effective multi-class emotional classification. The primary contribution of this research is listed below:

- Applies robustly optimized Bidirectional Encoder representation from transformer (RoBERTa)-Large to effectively learn semantic and contextual information from the pre-processed input, supporting the detection of discriminative emotion-based patterns.
- Employs a hybrid classifier that integrates a temporal convolutional network (TCN) to capture sequential and local contextual patterns, while using a multi-head attention (MHA) mechanism to learn different relationships simultaneously and focus on emotion-based features to perform accurate multi-class emotion classification.
- Incorporates Rectified Adam (RAdam) to provide more stable model learning and improved convergence.
- Performs experimental analysis through two benchmark datasets across multiple evaluation measures; these results demonstrate that the presented framework has the ability to accurately separate multiple emotion categories.

II. Literature Review

This part of work offers a brief note about the previous studies based on text-driven emotion detection and classification. Khan et al. [11] developed Long Short-Term Memory (LSTM) approaches for effective emotion detection and mental health monitoring. Furthermore, the authors integrated a RoBERTa transformer-based approach with this hybrid framework for scalable emotion detection. The framework demonstrates effective detection of complex emotional signs, thereby enhancing the emotion detection states. Avetisyan et al. [12] designed a novel prototype-guided emotion network framework for effective emotion detection through speech. Prototype learning has been widely adopted in emotion recognition and further analysed to evaluate its efficacy in resource-limited datasets. Furthermore, the authors demonstrated the efficiency of the framework in handling various Russian Speech Emotion Recognition (SER) datasets. Peng et al. [13] designed an innovative AI-driven model for multimodal semantic context construction, which attains various expressions of contextual semantics through learning of speech, text, and video clues. Furthermore, the authors integrated existing transformer-based algorithms with a self-attention module to increase the reliability and robustness of the framework. Adamu et al. [14] developed a novel hybrid stacked ensemble approach called Hyb-Stack, which incorporates cross-validation stacking with simple stacking. The authors further integrated various types of transformer-based approaches by employing classifier models to boost the classification accuracy. The authors emphasized the efficacy of integrating multiple transformer models into an optimized decision layer to further develop the emotion classification tasks on various linguistic contexts.

Moreno-Ortiz and García-Gómez [15] explored the representation of emotions in the television series *Severance* using a multi-layered annotation system according to the hierarchical taxonomy of emotions. Furthermore, the authors integrated LLM-based models to capture intricate patterns and features from large-scale datasets. In addition, NLP and LLM-based approaches provide novel opportunities for automated detection of emotion with superior contextual awareness. Kusal et al. [16] analysed the advantages and roles of emojis in detection of emotions in text messages, thereby emphasizing social media interaction. Furthermore, the authors developed a hybrid approach that integrates DL-based models for automated detection of emotions. In addition, the authors combined existing neural network approaches to learn fused methods and emoji representations. The authors emphasize the framework's capacity for elevating emotion interpretation in online social communication. Malik et al. [17] developed an innovative approach by integrating fusion and ensembling methods for effective emotion detection and classification. Furthermore, the authors integrated LSTM and RF-based algorithms for effective audio classification. In addition, Text-based assessment is carried out by applying Logistic Regression and CNN-based methods for accurate sentiment analysis and emotion detection.

Although recent developments in language representation and DL-based models have increased the performance of emotion recognition, several research gaps remain to be addressed in achieving more accurate multi-class classification. The existing model uses traditional embedding approaches that may not fully learn long-range semantic relationships from the textual content, thus limiting its capability of capturing complex emotional cues and attaining accurate classification outcomes. The previous model does not sufficiently use an attention-based mechanism, which creates challenges in identifying the most informative emotional features and also introduces less relevant information, negatively affecting the classification accuracy. The existing research uses standard or single learning architectures that may have limited ability to capture semantic, temporal, and contextual dependencies from text, which reduces the overall model performance. In addition, most existing methods use unstable or inefficient parameter optimization that affects the model convergence and generalization capability. To overcome these challenges, the proposed model develops the HDLF-SED framework by integrating RoBERTa-Large-based feature embedding, TCN-MHA classifier, and RAdam optimization into a unified model to perform accurate text-based multi-class emotion classification.

III. Proposed HDLF-SED Framework

The presented HDLF-SED model undergoes four major steps to identify and categorize multiple emotional classes from textual context. In the initial stage, the model applies an effective preprocessing pipeline to prepare the raw data into a cleaner and more standardized format for model development. In the second stage, the model applies RoBERTa-Large as a feature embedding module to capture rich and semantic representations from the pre-processed inputs for effective emotion-based analysis. In the third stage, the model employs hybrid

TCN-MHA, in which TCN captures sequential patterns, and MHA highlights the significant emotion-based information for effective classification. In the final stage, the model uses RAdam optimization to optimize model parameters, supporting effective convergence and enhanced classification performance. The overall architecture of the presented HDLF-SED approach is displayed in Fig. 1.

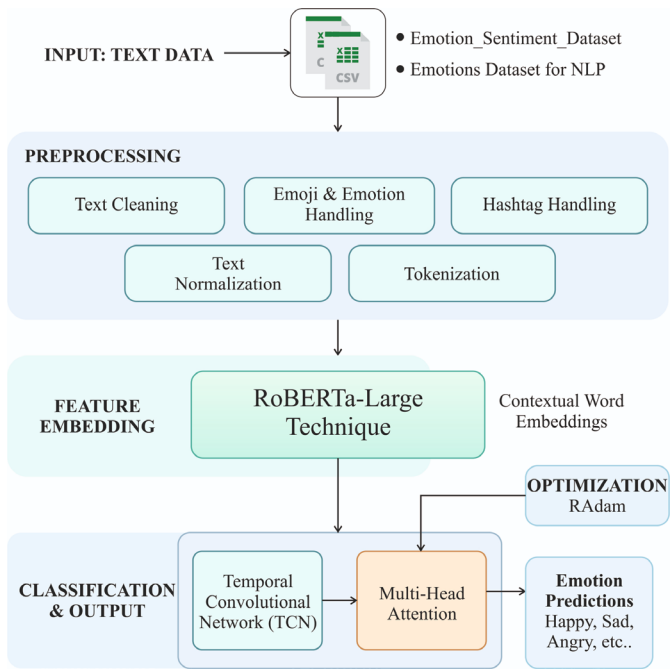


Fig. 1. Overall workflow of the presented HDLF-SED model

A. Dataset Description

The empirical evaluation of the HDLF-SED approach is performed using two datasets including Sentiment, Emotion analysis for mental health based on text (Dataset 1) [18] and Emotions dataset for NLP (Dataset 2) [19]. A summary of the two datasets is given below.

Dataset 1: This dataset encompasses 160,000 social media text records collected from two different sources for effective emotion recognition and sentiment analysis. Each record is classified into one of ten particular classes, including common emotions, sentiments, and psychological states such as depression. This dataset is extremely appropriate for training ML and DL-based models. Table 1 represents the dataset that comprises numerous emotional classes with a total of 160000 samples.

TABLE I. DETAILS OF DATASET I

Class Labels	Class Count
‘Love’	39553
‘Happiness’	27175
‘Sadness’	17481
‘Normal’	16351
‘Hate’	15267
‘Anger’	12336
‘Depression’	10333
‘Fun’	10075
‘Surprise’	6954
‘Worry’	4475
Total	160000

Dataset 2: The dataset is a text-based dataset developed for effective emotion sorting in NLP. It comprises short English sentences labeled with emotions like sadness, joy, love, anger, fear, and surprise. It includes 20000 text instances that are used for training and evaluating ML and DL approaches for identifying emotions expressed

in texts. The distribution of the six emotional classes is shown in Table 2.

TABLE II. DETAILS OF DATASET 2

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

Text Preprocessing

Data pre-processing is an important step in text-based emotion recognition, as it minimizes noise from the dataset, normalizes textual content, and prepares original data for efficient representation learning [20]. In this work, two databases were utilized, which were processed through multiple pre-processing steps, including data cleaning, emoji & emoticon handling, hashtag removal, Hashtag Handling, Text Normalization, and tokenization. The pre-processing is developed to maintain meaningful semantic and emotion-based information, while eliminating redundant textual differences. Initially, data cleaning is performed, since the original data may consist of non-printable characters, Email addresses, URLs, numerous whitespaces, and other redundant elements. These components are eliminated or suitably replaced in preprocessing. Later, Emojis and emoticons are transformed into descriptive textual embeddings to maintain their emotional importance. Social media information is frequently gathered utilizing hashtags. In preprocessing, these hashtags are treated distinctly from the key text. Because hashtags could comprise numerous words, they are divided into useful specific words. Subsequently, text normalization is applied, in this stage the unreliable text forms, common abbreviations, repeated characters, and spelling variances are normalized while maintaining semantic and emotional signs. Example: “I am sooo happpppy to seeee u” → “I am happy to see you”.

Finally, tokenization is applied, this process splits the text into particular words or tokens for upcoming examination. For instance – “I am so happy today!” → [I, am, so, happy, today, !]. Afterwards, the cleaned and normalized textual information are passed to RoBERTa-Large for contextual feature representation, allowing rich semantic extraction and emotion-relevant representations for the following classification.

B. RoBERTa-Large Feature Embedding

RoBERTa-Large is used as the feature representation method due to its ability to efficiently learn contextual and semantic data from textual inputs, supporting the identification of useful emotion-relevant patterns. RoBERTa-Large [21] is a transformer-based language model derived from the BERT structure with an optimized pre-training method. It creates high contextual representations by treating the interactions among words within their neighboring text, allowing efficient representation of the semantic and linguistic features of the input text. In this method, the input representations are responsible for transforming input pre-processed text tokens into higher-dimensional dense vectors. These vectors are generated by integrating token and positional data, in which token embedding signifies the semantic data of each token, whereas positional embeddings maintain their order in the sequence. The input embeddings can be represented as:

Here, $\mathbf{e}_i$ and $\mathbf{p}_i$ denote the token and the positional embeddings of the i -th token. The embedded tokens are passed through multi-head self-attention to learn contextual relations between words. This module classifies the significance of all tokens with respect to other tokens, facilitating the model to learn semantic and emotion-relevant patterns.

Let $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ denote the query, key, and value matrices of the attention head. d_k indicates the key dimension. The outcomes from numerous attention heads are integrated to create enhanced contextual representations. The attention output is then fed to a feed-forward network containing dual linear layers with a GELU function. It

captures non-linear transformations and high-level semantic patterns from the contextual embeddings:

In this Eq. (3), $W1$ and b represent trainable weight matrices, and b indicate bias terms. This procedure further improves the semantic and emotion-relevant characteristics.

Finally, Residual connections are used around the transformer sub-layers to maintain meaningful data and enable gradient flow, whereas layer normalization steadies the hidden representations. The function is represented as:

Now, z denotes the self-attention output, and h implies the feed-forward embeddings. These functions help RoBERTa-Large create consistent and enhanced contextual feature embeddings from the pre-processed textual information. The final features are subsequently applied by the TCN with MHA for efficient emotion classification. Fig.2. represents the architecture of RoBERTa-Large model

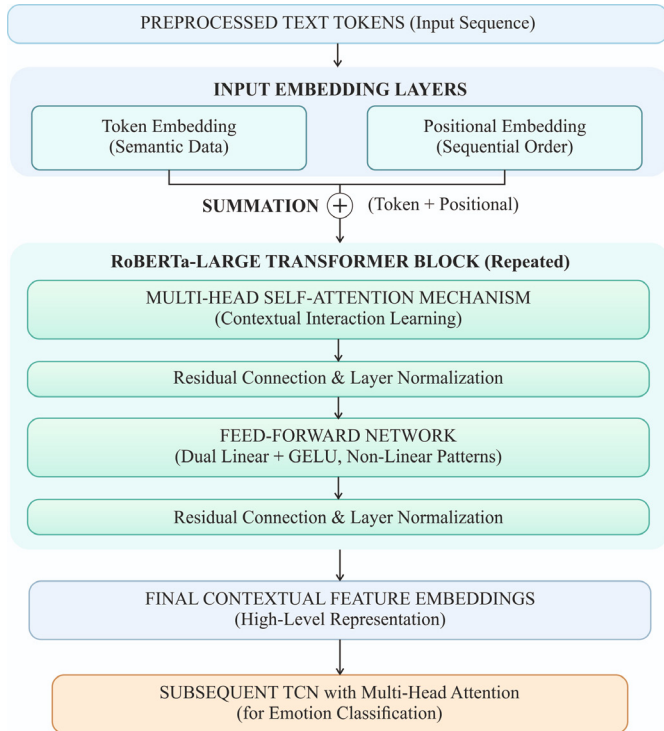


Fig. 2. Architecture of the RoBERTa-Large model

C. Hybrid TCN-MHA classifier

To capture emotional expressions based on both sequential interactions and contextually significant semantic features, a hybrid TCN-MHA classifier can jointly learn sequential patterns and isolate emotion-related characteristics from the RoBERTa-Large representations [22].

1) Temporal Convolutional Network

Traditional CNNs can't efficiently learn longer-range interactions in sequential textual representations. Therefore, a TCN is utilized to efficiently model the sequential contextual patterns from the RoBERTa-Large characteristics. TCN is categorized by layered stacks of dilated causal convolution (DCC) with residual connections (RC). Causal convolution in TCN maintains the sequential order of textual embedding by enabling the output at location t based on the recent and previous elements. This allows the method to preserve significant contextual data from previous textual locations. To learn a wider contextual range without extremely increasing network depth, dilated convolution can be proposed. Dilated convolution instantiates the input representations in pre-defined intervals, thus extending the receptive field and facilitating the TCN to learn contextual data from diverse positions of the text. The DCC function on element x_t of the sequence can be expressed as follow:

Here, d implies the dilation factor, k denotes the filter size, f indicates the convolution filter, and X and Y represent the input and output. A residual block is a fundamental structure of the TCN is given by:

Let $\mathbf{x}$ denote the input representation and $\mathbf{z}$ represent the transformed feature embeddings.

2) Multi-Head Attention (MHA)

Even though TCN obtains sequential relationships, different words and contextual representations might contribute unevenly to emotion detection. Therefore, MHA is combined to highlight the most relevant semantic and contextual characteristics. The self-attention module allocates different importance weights to the input representations as per their contextual significance. The scaled dot-product attention can be represented as:

Now, $\mathbf{z}$ represents the TCN-refined feature representation, $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ imply the query, key, and value matrices. In addition, $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ indicate the weight matrices of $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$, correspondingly, and d_k is the dimensionality of the three matrices. MHA integrates numerous self-attention modules, enabling the method to capture diverse contextual interactions concurrently and concentrate on emotionally relevant textual characteristics. The multi-head attention function can be represented as:

Let $\mathbf{Q}_h$, $\mathbf{K}_h$, and $\mathbf{V}_h$ denote the projection matrices of the h th attention head, H implies the total count of attention heads, and $\mathbf{W}_O$ indicates the projection matrix of the output.

3) Multi-Class Emotion Classification

The resulting contextual representation extracted from the TCN–MHA mechanism is fed to a fully connected (FC) layer followed by a Softmax activation to achieve multi-class emotion classification. The Softmax function converts the output logits into probability values for the C emotion categories:

Here, $\mathbf{z}_c$ implies the logit corresponding to the emotion category, C denotes the overall count of emotion classes, and $\mathbf{p}_c$ represents the forecasted likelihood of category c . The class with a higher probability can be chosen as the resulting predicted emotion. This hybrid TCN–MHA classifier integrates a sequential dependency model with attention-driven contextual weighting, whereas the resulting Softmax layer allows multi-class emotion prediction from the captured textual embeddings.

D. RAdam Optimization Algorithm

RAdam is a modified version of the Adam optimizer developed to solve instability in the early phase of the learning process. In the presented system, RAdam is appropriate for optimizing the RoBERTa-Large-based representations and the successive TCN–MHA classification layers, where efficient and consistent parameter updates are vital for learning intricate emotional patterns and textual information [23]. The main advance in RAdam is its variance rectification module that adapts the learning rate dynamically in early training. Traditional Adam adjusts the learning rate using the 1st and 2nd gradient moments, RAdam, rectifies unreliable variance estimates to provide more stable parameter updates. The RAdam can be expressed as:

Here, $\hat{v}_t$ represents the variance rectification computed at training step t , $\hat{v}_{max}$ implies the theoretical higher value of the variance measure, τ denotes the recent training step, and β_2 is the exponential decay rate applied for approximating the 2nd gradient moment. Within this rectification measure, the learning rate can be adapted as:

where η denotes the base learning rate and $\hat{\eta}_t$ represents the rectified learning rate. Therefore, RAdam offers stable and adaptive optimization by rectifying untrustworthy variance estimates in the early training stage, thus allowing smoother parameter updates and more reliable convergence of the presented model for text-based emotion recognition.

IV. Experimental Result and Discussion

This section provides the experimental analysis of the HDLF-SED model using two bench mark datasets along with detailed discussion

A. Classification out of the proposed model using Dataset 1

Figs. 3a and 3b describe the confusion matrices of the HDLF-SED framework with a 70:30 training/testing split for Dataset 1. The matrices indicate that higher diagonal values are properly categorized across all classes, whereas off-diagonal values denote misclassified samples, demonstrating the effective classification of the proposed text-based emotion detection.

Table 3 portrays the textual emotion classification outcomes of the HDLF-SED approach with 70:30 training/testing for Dataset 1. The proposed HDLF-SED system achieved overall accuracy of 97.03% and 96.95% for the training and testing phases. This performance indicates the reliability and capability of the proposed system for text-based emotion detection.

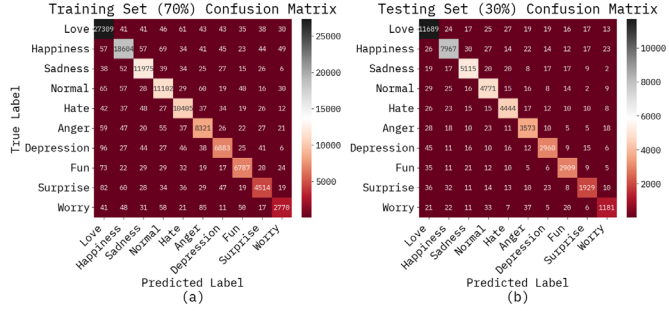


Fig. 3. Confusion matrices of the HDLF-SED approach with (a-b) 70:30 training/testing split for Dataset 1

TABLE III. TEXTUAL EMOTION CLASSIFICATION RESULTS OF THE HDLF-SED METHOD WITH A 70:30 SPLIT FOR DATASET I

Classes	Accu- racy	Preci- sion	Recall	F1- Score	MCC
Training Set (70%)					
Love	99.17	98.02	98.63	98.32	97.77
Happi- ness	99.28	97.94	97.80	97.87	97.43
Sadness	99.48	97.35	97.86	97.60	97.31
Normal	99.35	96.66	96.99	96.83	96.46
Hate	99.45	96.93	97.36	97.14	96.84
Anger	99.38	95.67	96.36	96.01	95.68
Depres- sion	99.45	96.24	95.16	95.70	95.40
Fun	99.54	96.47	96.24	96.36	96.11
Sur- prise	99.46	94.65	92.73	93.68	93.40
Worry	99.50	93.36	88.44	90.83	90.61
Average	99.41	96.33	95.76	96.03	95.70
± SD	± 0.11	± 1.38	± 2.91	± 2.14	± 2.08
Overall Accuracy: 97.03					
Testing Set (30%)					
Love	99.08	97.78	98.51	98.14	97.53
Happi- ness	99.23	97.75	97.73	97.74	97.28
Sadness	99.42	97.21	97.54	97.37	97.05
Normal	99.35	96.38	97.27	96.82	96.46
Hate	99.44	97.09	97.03	97.06	96.75
Anger	99.43	96.07	96.54	96.31	96.00

Depression	99.47	96.29	95.48	95.89	95.60
Fun	99.53	96.32	96.23	96.28	96.03
Surprise	99.49	95.54	92.47	93.98	93.73
Worry	99.47	92.63	87.94	90.22	89.98
Average	99.39	96.31	95.67	95.98	95.64
$\pm$ SD	± 0.13	± 1.41	± 3.02	± 2.21	± 2.15
Overall Accuracy: 96.95					

Figs. 4a and 4b represent the PR curves of the HDLF-SEDSsystem for 70:30 training/testing with Dataset 1. The consistently higher AP scores within the range of 0.940–0.993 denote that the model efficiently identifies emotions from textual input with higher precision and recall. The same training and testing performance gives better generalization, with Love performing best and Worry exhibiting comparatively minimal performance.

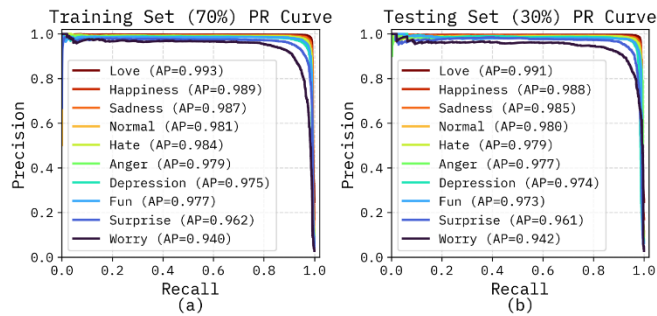


Fig. 4. PR curve of the HDLF-SED method with (a-b) 70:30 training/testing split for Dataset1

B. Classification out of the proposed model using Dataset 2

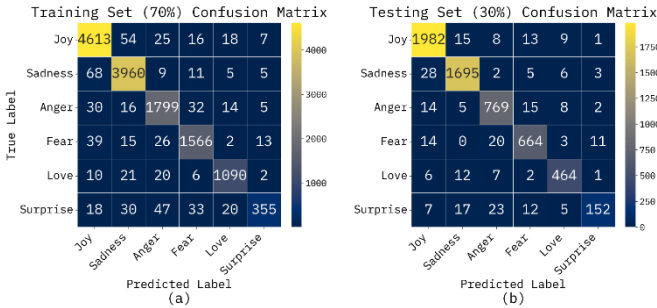


Fig. 5. Confusion matrices of the HDLF-SED approach with (a-b) 70:30 training/testing split for Dataset 2

Classes	Accu- racy	Preci- sion	Recall	F1- Score	MCC
Training Set (70%)					
Joy	97.96	96.55	97.46	97.00	95.46
Sadness	98.33	96.68	97.59	97.13	95.95
Anger	98.40	93.41	94.88	94.14	93.22
Fear	98.62	94.11	94.28	94.20	93.41
Love	99.16	94.87	94.87	94.87	94.41
Sur- prise	98.71	91.73	70.58	79.78	79.85
Average	98.53	94.56	91.61	92.85	92.05
$\pm$ SD	± 0.37	± 1.74	± 9.49	± 5.97	± 5.55

Overall Accuracy: 95.59					
Testing Set (30%)					
Joy	98.08	96.64	97.73	97.18	95.73
Sadness	98.45	97.19	97.47	97.33	96.24
Anger	98.27	92.76	94.59	93.67	92.67
Fear	98.42	93.39	93.26	93.32	92.43
Love	99.02	93.74	94.31	94.02	93.49
Sur- prise	98.63	89.41	70.37	78.76	78.66
Average	98.48	93.85	91.29	92.38	91.54
± SD	± 0.29	± 2.58	± 9.50	± 6.30	± 5.94
Overall Accuracy: 95.43					

Fig. 5 demonstrates the confusion matrix of the HDLF-SED system with a 70:30 training/testing split for Dataset 2. Fig. 5a and 5b represent that the model classifies most emotions appropriately, with excellent performance seen for Joy, Sadness, and Anger. A few confusions arise between similar emotional classes, especially Surprise, Fear, and Love. Likewise, the testing outcomes significantly follow the training patterns, representing consistent model performance.

Table 4 exhibits the textual emotion detection results of the HDLF-SED methodology with 70:30 in Dataset 2. The results depict that the model gained strong and consistent performance, with overall accuracies of 95.59% and 95.43% for both training/testing sets. The close training and testing outcome reveal that the model has good generalization without sign of overfitting. Therefore, the average MCC values with SD of $92.05\% \pm 5.55$ in training while it attains in testing $91.54\% \pm 5.94$.

Fig. 6 displays the PR curve of the HDLF-SED framework for a 70:30 training/testing split on Dataset 2. Fig. 6a presents the training PR curves, which remain closer to the top, demonstrating higher precision and recall for numerous emotions. The Surprise class has a moderately decreased AP (0.891), indicating more complexity in its classification. Fig. 6b shows the testing curves also retain greater precision throughout most recall levels, providing steady performance on unseen instances. Surprise again has the lowest AP (0.909), whereas Joy and Sadness achieve the strongest outcomes, with AP values in the range of 0.992 and 0.991.

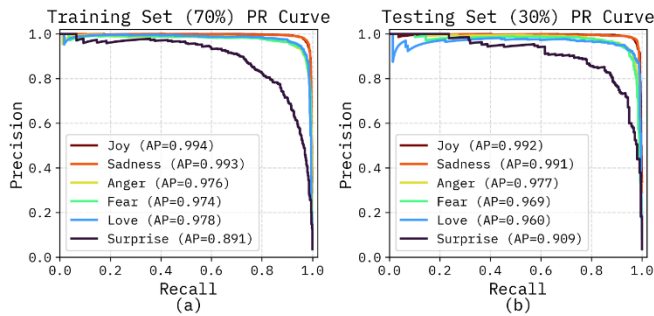


Fig. 6. PR curve of the HDLF-SED method with (a-b) 70:30 of training/testing split for Dataset 2

C. Comparative analysis and ablation study of the proposed model

Table 5 and Fig. 7 portray the comparison outcomes of the HDLF-SED model with existing methodologies utilizing several measures for both datasets [24, 25]. In Dataset 1, the result implied that the proposed HDLF-SED method achieves maximal solutions with an accuracy of 96.95%, a precision of 96.31%, and a recall of 95.67% compared with conventional approaches. Moreover, for Dataset 2, the outcome indicated that the RNN, GRU, and Bi-GRU methods attain fewer solutions, while the LSTM, CNN, BERT, and RoBERT algorithms slightly improved performance. However, the proposed HDLF-SED system obtains higher performance with accuracy of 95.43%, precision of 93.86%, and recall of 91.29%, thereby confirming its stable and effective performance for the presented text-based emotion classification.

Table 6 exhibit the ablation study of the HDLF-SED method with several measures for Dataset 1. The

result implied that the baseline RoBERTa-Large model obtains lower performance, while integrating TCN and MHA improves the model's capability to capture complementary temporal and contextual characteristics. The complete HDLF-SED (RoBERTa-Large + TCN + MHA+ RAdam) method achieves better performance with an accuracy of 96.95%, precision of 96.31%, recall of 95.67%, and an F1-score of 95.98%, indicating that gradually integrating each component efficiently enhances the model's overall performance in text-based emotion detection.

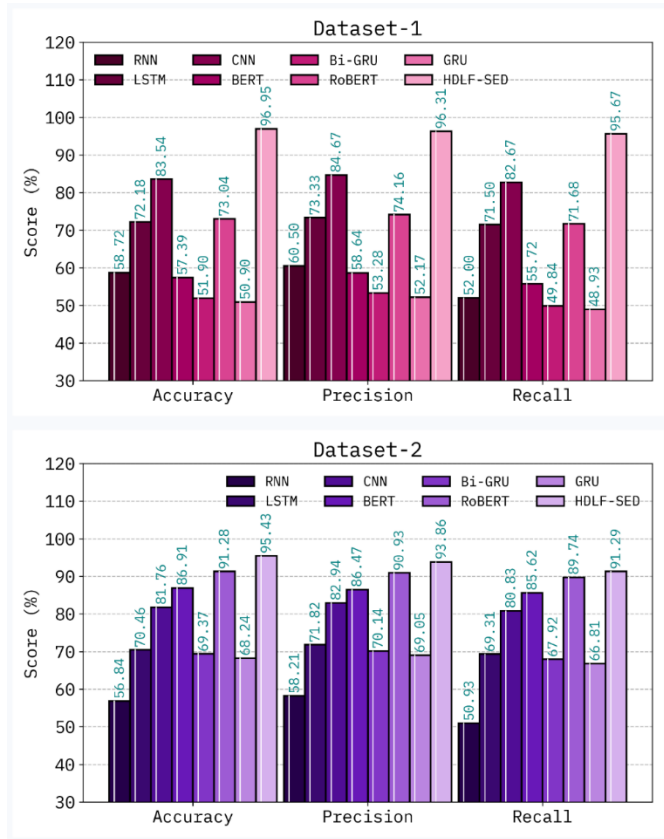


Fig. 7. Comparison result of the HDLF-SED method with metrics using both datasets

TABLE V. COMPARATIVE OUTCOMES OF THE HDLF-SED APPROACH WITH EXISTING SYSTEM USING BOTH DATASETS

Methods	Dataset-1			Dataset-2		
	Accuracy	Precision	Recall	Accu- racy	Precision	Recall
RNN [24]	58.72	60.50	52.00	56.84	58.21	50.93
LSTM [24]	72.18	73.33	71.50	70.46	71.82	69.31
CNN [24]	83.54	84.67	82.67	81.76	82.94	80.83
BERT [25]	57.39	58.64	55.72	86.91	86.47	85.62
Bi-GRU [25]	51.90	53.28	49.84	69.37	70.14	67.92
RoBERT [25]	73.04	74.16	71.68	91.28	90.93	89.74
GRU [25]	50.90	52.17	48.93	68.24	69.05	66.81
HDLF-SED	96.95	96.31	95.67	95.43	93.86	91.29

TABLE VI. ABLATION STUDY OF THE HDLF-SED SYSTEM WITH SEVERAL METRICS FOR DATASET 1

Methods	Accuracy	Precision	Recall	F1-Score
RoBERTa-Large + TCN	91.62	90.48	88.76	89.59
RoBERTa-Large + MHA	93.47	92.16	90.82	91.48
RoBERTa-Large + TCN + MHA	95.28	94.67	93.42	94.02
HDLF-SED (RoBERTa-Large + TCN + MHA+ RAdam)	96.95	96.31	95.67	95.98

Table 7 shows the ablation analysis of the HDLF-SED approach with numerous measures for Dataset 2. The empirical outcomes indicate that the baseline RoBERTa-Large model demonstrates reasonably lower performance. At the same time, the integration of TCN and MHA offers significant improvements by successfully learning temporal dependencies and contextual information. Additionally, the wide-ranging HDLF-SED (RoBERTa-Large + TCN + MHA+ RAdam) models achieve higher outcomes with an accuracy of 95.43%, precision of 93.86%, recall of 91.29%, and an F1-score of 92.38%, demonstrating that progressively combining every component efficiently enhances the overall performance of the model in text-based emotion detection.

TABLE VII. ABLATION STUDY OF THE HDLF-SED APPROACH WITH VARIOUS MEASURES FOR DATASET 2

Methods	Accuracy	Precision	Recall	F1-Score
RoBERTa-Large + TCN	88.74	86.92	83.68	85.26
RoBERTa-Large + MHA	90.86	89.14	86.42	87.74
RoBERTa-Large + TCN + MHA	93.18	91.72	88.94	90.28
HDLF-SED (RoBERTa-Large + TCN + MHA+ RAdam)	95.43	93.86	91.29	92.38

D. Discussion

The outcomes show that the presented HDLF-SED model can efficiently separate multiple emotion classes using textual data. Initially, the effective preprocessing pipeline improves the quality of the data, while RoBERTa-Large generates rich and semantic representations by learning contextual meaning and relationships among words. The TCN effectively captures sequential and local patterns within the RoBERTa-Large-based representations, while MHA highlights the most informative emotion-based features to perform robust multi-class classification. Finally, RAdam supports stable and effective model learning and facilitates faster convergence. Therefore, these combined components of the proposed model capture and complementary semantic, sequential, and temporal dependencies, contributing to more accurate and reliable multi-class emotion classification across two benchmark datasets.

V. Conclusion and Future Work

In this manuscript, the presented HDLF-SED model is designed for semantic-driven emotion detection employing textual data. The model initially preprocesses the raw text data using multiple steps such as text cleaning, emoji and emoticon handling, Hashtag handling, text normalization, and tokenization to provide a cleaner and standardized textual representation for model development. Then, RoBERTa-Large is applied to extract rich and semantic features by learning the interaction among the words within their neighboring context. Then these representations are processed through a hybrid TCN-MHA model, in which TCN captures sequential and local contextual patterns, while MHA emphasizes motion-related information for effective multi-class classification. In addition, the RAdam optimization algorithm is employed for more stable model learning and faster convergence. Through the experimental analysis, the presented system is assessed through two benchmark datasets across various evaluation metrics; these outcomes validate the robustness of the HDLF-SED framework in identifying and categorizing multiple emotional states from textual data. To further strengthen the proposed model and improve its applicability, various opportunities remain in future studies. The proposed model can be extended by incorporating a multi-modal emotional dataset by combining textual information with speech, facial expressions, or visual cues to improve its practical adaptability. In addition, the proposed model can be further improved through incorporating explainable artificial intelligence to identify the textual elements that influence the model prediction, thus enhancing the model transparency and trustworthiness.

Data Availability Statement: The data that support the findings of this study are openly available in the Kaggle repository at <https://doi.org/10.34740/kaggle/dsv/19741270> , reference number [18].

Funding note and Acknowledgment: The authors extend their appreciation to Prince Sattam bin Abdulaziz University for funding this research work through the project number (PSAU/2025/01/ 34648)

REFERENCES

- Arzu, G.E., Umar, M., Khan, A., Ali, U., Dang, L.M. and Moon, H., 2026. Adaptive multimodal emotion detection for mental health monitoring using deep learning. *Information Sciences*, p.123385.
- Arumugam, L., Arumugam, S., Chidambaram, P. and Govindasamy, K., 2025. A multi-model deep learning approach for human emotion recognition. *Cognitive Neurodynamics*, 19(1), p.123.
- Bhatia, J., Sodhi, G.K. and Bhatia, S., 2026. Multimodal sentiment analysis: Advances in machine learning and deep learning approaches. *Recent Advances in Computational Methods in Science and Technology*, pp.247-253.
- Huang, J., Chen, M., Zhu, X., Zhang, J. and Duan, Z., 2026, April. SEMSA: A Semantic-Enhanced Framework for Multimodal Sentiment Analysis. In *2026 11th International Conference on Computer and Communication Systems (ICCCS)* (pp. 1189-1196). IEEE.
- Liu, Z. and Xie, L., 2026. MAC-DAG: a modality-adaptive contextual directed acyclic graph network for multimodal emotion recognition in conversations. *Multimedia Systems*, 32(7), p.452.
- Kumar, A. and Agrawal, R., 2026. Real-time emotion detection system using a hybrid machine learning model for enhanced sentiment analysis in online learning environments. *International Journal of Information Technology*, 18(1), pp.233-247.
- Chittepu, S., Martha, S. and Banik, D., 2026. Comprehensive survey on multimodal emotion detection with advances, challenges, and future directions in pairwise conversations. *Cognitive Processing*, 27(2), pp.275-290.
- Rishu and Kukreja, V., 2026. Leveraging multimodal fusion for emotion detection in comics: integrating text and visual cues with RoBERTa and vision transformers. *International Journal on Document Analysis and Recognition (IJ DAR)*, 29(2), pp.267-288.
- Almogren, A.S., Elhessewi, G.M.S., Ghaleb, M., Mahgoub, H., Alhashmi, A.A., Alzubaidi, U., Alghamdi, A.M. and Aljehane, N.O., 2026. Deep learning approach for social IoT and mental health emotion detection from videos. *IEEE Transactions on Computational Social Systems*.
- Shou, Y., Meng, T., Ai, W., Fu, F., Yin, N. and Li, K., 2026. A comprehensive survey on multi-modal conversational emotion recognition with deep learning. *ACM Transactions on Information Systems*, 44(2), pp.1-48.
- 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.
- Avetisyan, L.A., Tigranyan, S.T. and Avetisyan, A.A., 2026. ProtoGEN: A Prototype-Guided Emotion Network for Multimodal Speech Emotion Recognition. *IEEE Access*, 14, pp.93027-93041.
- Peng, Y., 2026, April. Immersive Multimodal Perception-Driven Context Construction for Network New Media. In *Proceedings of the 3rd International Conference on Machine Intelligence and Digital Applications* (pp. 476-481).
- Adamu, H., Azmi Murad, M.A. and Nasharuddin, N.A., 2026. A hybrid stacked ensemble learning framework for multilabel text emotion detection. *Scientific Reports*, 16(1), p.7714.
- Moreno-Ortiz, A. and García-Gámez, M., 2026. Emotion in context: human and AI perspectives on the affective landscape of severance. *Corpus Pragmatics*, 10(1), p.15.
- 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.
- Malik, S.S., Ilyas, M., Haq, Y.U., Sana, R., Razzaq, M.S., Maqbool, F. and Pathan, M.S., 2025. Multi-modal emotion detection and sentiment analysis. *IEEE Access*, 13, pp.59790-59810.
- <https://doi.org/10.34740/kaggle/dsv/19741270>
- Shrivash, B.K., Verma, D.K. and Pandey, P., 2025. A Novel Framework for Text Preprocessing using NLP Approaches and Classification using Random Forest Grid Search Technique for Sentiment Analysis. *Economic Computation & Economic Cybernetics Studies & Research*, 59(2), p.91.
- Khan, H.U., Naz, A., Alarfaj, F.K. and Almusallam, N., 2025. Analyzing student mental health with RoBERTa-Large: a sentiment analysis and data analytics approach. *Frontiers in big Data*, 8, p.1615788.
- Nie, L., Xu, S., Zhang, L., Yin, Y., Dong, Z. and Zhou, X., 2022. Remaining useful life prediction of aeroengines based on multi-head attention mechanism. *Machines*, 10(7), p.552.

- Dhingra, S., Gruosso, G. and Storti Gajani, G., 2025. A benchmark study of optimizers for short-term solar PV power forecasting using neural networks under real-world constraints. *Neural Computing and Applications*, 37(29), pp.23909-23939.
- Al-Milli, N., Elsoud, E.A., Al-Shaikh, A.A., Omar, H. and AlNajjar, A., 2026. Emotion detection and sentiment analysis of textual data. *Egyptian Informatics Journal*, 34, p.100976..
- 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.