



Open Access



Check for updates

REVIEW ARTICLE

Section: *Literature, Linguistics & Criticism*

Modeling of corpus linguistics with semantic-aware deep neural decision forest for enhanced emotion detection from 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**

Emotion detection from textual data plays an important role in understanding the emotional content expressed through written language. Moreover, the natural language processing (NLP) provides useful techniques for analyzing textual content and identifying meaningful linguistic patterns associated with different emotions. It helps to recognize different emotional states from the words and context present in a text. However, traditional machine learning approaches are often relied on manually designed features and may not capture the underlying semantic relationships effectively. Deep learning models can learn representations automatically, but distinguishing between closely related emotions remains challenging. To address these limitations, this paper introduces a Deep Neural Decision Forest-based Framework for Semantic-Driven Emotion Detection (DNDF-SED) using textual data. The proposed model first preprocesses the text through cleaning, contraction expansion, lemmatization, negation handling, and subword tokenization to obtain clean and consistent text for subsequent word representation process. Then, DeBERTa-v3 is utilized to learn contextual and semantic representations from the processed text. These representations are given to a Deep Neural Decision Forest classifier, which integrates deep feature learning with decision-tree-based classification to distinguish different emotion categories. Moreover, NAdam is applied to optimize the model parameters and enhance the learning process. Comprehensive experimental outcomes demonstrate the effectiveness of the proposed DNDF-SED framework, achieving greater accuracy of 95.70% on Emotions dataset for NLP and 98.71% on Emotion Dataset Raw compared with existing methodologies. Additionally, the ablation study confirms the contribution of every component by showing improved classification performance with the proposed methodology. These results demonstrate the effectiveness of integrating contextual language representations with deep decision forest classification for semantic-driven emotion detection.

KEYWORDS: Emotion Detection, Natural Language Processing, DeBERTa-v3, Deep Neural Decision Forest, NAdam Optimizer

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). Modeling of corpus linguistics with semantic-aware deep neural decision forest for enhanced emotion detection from textual data. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/ccdy8p06>



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

Human emotions are highly significant in decision-making, motivation and action, social connection, adaptation, and psychological health. The rapid growth of user-generated content on social media websites and review platforms emphasizes the need for precise and reliable emotion recognition [1]. By interpreting the emotions through textual documents, businesses can improve customer service and the quality of human–computer interaction (HCI). Text-based emotion recognition (TER) attempts to identify emotional condition automatically through text-based expressions, including joy, happiness, anger, sadness, fear, surprise, and disgust [2]. Existing Sentiment analysis (SA) methods categorize emotional expressions as positive, negative, and neutral, but current applications need a finer-grained emotion classification [3]. In recent times, TER has attained significant attention in the research field, which has led to the advancement of computational and natural language processing (NLP) techniques for attaining precise and reliable emotion detection through textual data [4]. Furthermore, TER has been implemented in various real-time applications, including education, business institutions, consumer assessment, healthcare, and HCI. However, NLP-based techniques still face limitations due to computational and linguistic challenges, which make it difficult for models to understand and create HCI using text and audio [5]. NLP-based approaches mainly aim to develop machines that can effectively operate by understanding emotions, sentiment, and thoughts. Textual data often retains and delivers emotion in various forms, which makes it difficult to capture and interpret emotions [6]. Emotions can be effectively expressed by using emotion-based terms and language, which makes it easy to interpret and analyze. However, the major challenges in TER include several factors, such as choice of words, structure, formation of sentences, and the usage of punctuation [7].

In recent times, advancements in Artificial Intelligence (AI) have emerged and provided promising automated solutions, thereby driving significant progress in various domains. In addition, deep learning (DL)-based approaches have gained significant attention across various research domains, by demonstrating their efficacy in a wide-range of applications, particularly in the field of information technology [8]. Over the past few years, DL-based approaches have decreased the dependence on manual feature extraction and achieved promising solutions. Most of the research on deep emotion recognition methods focuses primarily on the visual, audio, and multimodal modalities [9]. The automated detection of emotional signs supports business institutions and investigators in attaining valuable insights about user behavior and emotional states. By applying linguistic patterns in a wide range of text collections enables precise and reliable emotion detection [10]. Therefore, DL-based approaches are essential for learning subtle discrepancies in language, including sarcasm and irony, syntactic ambiguity, and homonyms, which can result in robust and accurate emotion assessment in various contexts [11]. Although DL approaches have enhanced emotion recognition from text, classifying the capture expressed in a sentence can be still complex. The meaning of words could modify based on the sentence and the neighbouring context. Conventional approaches can also miss-subtle emotional signs and semantic interactions during text. Consequently, a semantic-based method is essential to comprehend the meaning of textual content and enhance emotion detection. To overcome these challenges, this paper introduces a Deep Neural Decision Forest-based Framework for Semantic-Driven Emotion Detection (DNDF-SED) using textual data. The proposed framework integrates text preprocessing, DeBERTa-v3-based contextual word representation, DNDF-based emotion classification, and NAdam optimization to improve the detection of different emotional categories from textual data. The core contributions of this research are summarized as follow:

- Utilize DeBERTa-v3 to capture contextual and semantic representations from the input text, which helps to generate meaningful word representations for emotion classification.
- Design a Deep Neural Decision Forest (DNDF) classifier to learn complex decision patterns from the obtained textual representations and accurately distinguish different emotion classes, namely joy, sadness, anger, fear, love, and surprise.
- Integrate the NAdam optimizer to effectively update the model parameters and achieve better convergence during training.
- Perform the experimental analysis of the proposed DNDF-SED model using two datasets, namely the Emotions dataset for NLP and Emotion Dataset Raw, to demonstrate its effectiveness under various evaluation measures.

II. Literature Review on Emotion Detection

In this section, a concise overview of existing literature on emotion detection is discussed, along with the identified research gaps. Avetisyan et al. [12] developed an innovative framework for multimodal speech emotion recognition. Furthermore, the authors integrated a prototype-based objective to support acoustic embeddings with the commentators through cosine classification. Furthermore, this framework offers an effective supplementary prototype-based objective for accurate recognition of Speech Emotion. Sun et al. [13] introduced a novel cross-modal transformer-based framework for risk monitoring. The authors further integrated an AI-based approaches to build a combined multimodal financial sensing system for automated sensing of anomaly and risk awareness. Furthermore, a multilingual encoding model is developed to extract inconsistent textual sensor noise and alleviate cross-lingual semantic sensing representations. Peng [14] designed an innovative algorithm-based multimodal semantic context construction approach, which attains expression of contextual semantics using integrated learning of visual, audio, and textual indications. The authors further highlighted the reliability of the framework in dynamic environments and its capacity to provide robust and accurate automated support for interaction enhancement in multimodal systems. Sun et al. [15] explored the vision-language-motion diffusion approach for creating instruction-aware co-speech signs for humanoid robots in the educational environment. Furthermore, the authors examined the mixture-of-experts approaches to predict Valence and Arousal through text, visual, and speech patterns that can be mapped to separate teaching-task classes using an affect-based policy.

Tian et al. [16] developed an innovative cross-modal GNN-GRU approach for extracting animation-video key frames and creating semantic-based special effects. The author integrated this architecture with various modalities for key frame extraction and semantic-driven special-effects generation in animation videos. Those animation frames are generated by CNN-based models, speech semantics are generated by Transformer-based models, and rhythm indications are generated by Fourier analysis. Huo et al. [17] introduced an innovative framework called as SBT-Net for accurate depression identification through textual and speech data. Furthermore, the authors integrated a cross-modal gating module to improve the arrangement among models and used an emotion trend modeling (ETM) mechanism to learn the temporal progression of depressive emotional conditions. Yuan et al. [18] developed a novel collaborative framework that incorporates BERT-based approach with VITS for emotional speech synthesis. The authors mainly designed this framework to improve the sincerity, articulation, and reaction of HIC quality. This framework assures reliability and accuracy, thereby enabling an immersive communication experience in various application environments, including education, client satisfaction, and psychological companies. Hu et al. [19] designed an innovative approach by employing Label Semantic-driven Contrastive Learning (LaSCL) for emotion recognition through speech. Furthermore, the authors integrated speech representation learning for accurate emotion recognition from speech data. This framework utilizes label embeddings as semantic anchors to clearly develop relations among emotions and further used a label divergence loss for better and transparent emotion boundaries.

A. Research Gap

Although existing emotion detection approach have obtained better performance, but accurately understanding emotions from content is still difficult. Numerous conventional methods concentrate only on textual features and cannot learn the meaning of words in diverse contexts. The interaction among words and their emotional meaning could also differ through sentences, increasing classification complexity. Emotions with same meanings are frequently hard to differentiate from one other. Several conventional approaches also provide constrained interpretation of why a specific emotion is allocated to a text. These limitations show the essential for a model that could enhance use semantic data whereas preserving consistent classification. Thus, this research work presents a DNDF-SED structure utilizing textual information.

III. Methodological Framework

The proposed DNDF-SED model follows a multistage pipeline to identify different emotion states from textual data. The overall process of the proposed DNDF-SED is represented in Fig. 1. In this initial stage, the input text is first preprocessed using text cleaning, contraction expansion, lemmatization, negation handling, and subword tokenization to prepare the text for further processing. Besides, the DeBERTa-v3 model is applied to extract

the contextual and semantic representations, while DNDF approach is utilized to classify the different emotion classes based on the obtained textual representations. Moreover, the NAdam is used to optimize the model parameters and improve the overall classification performance. The detailed description of each component is explained below.

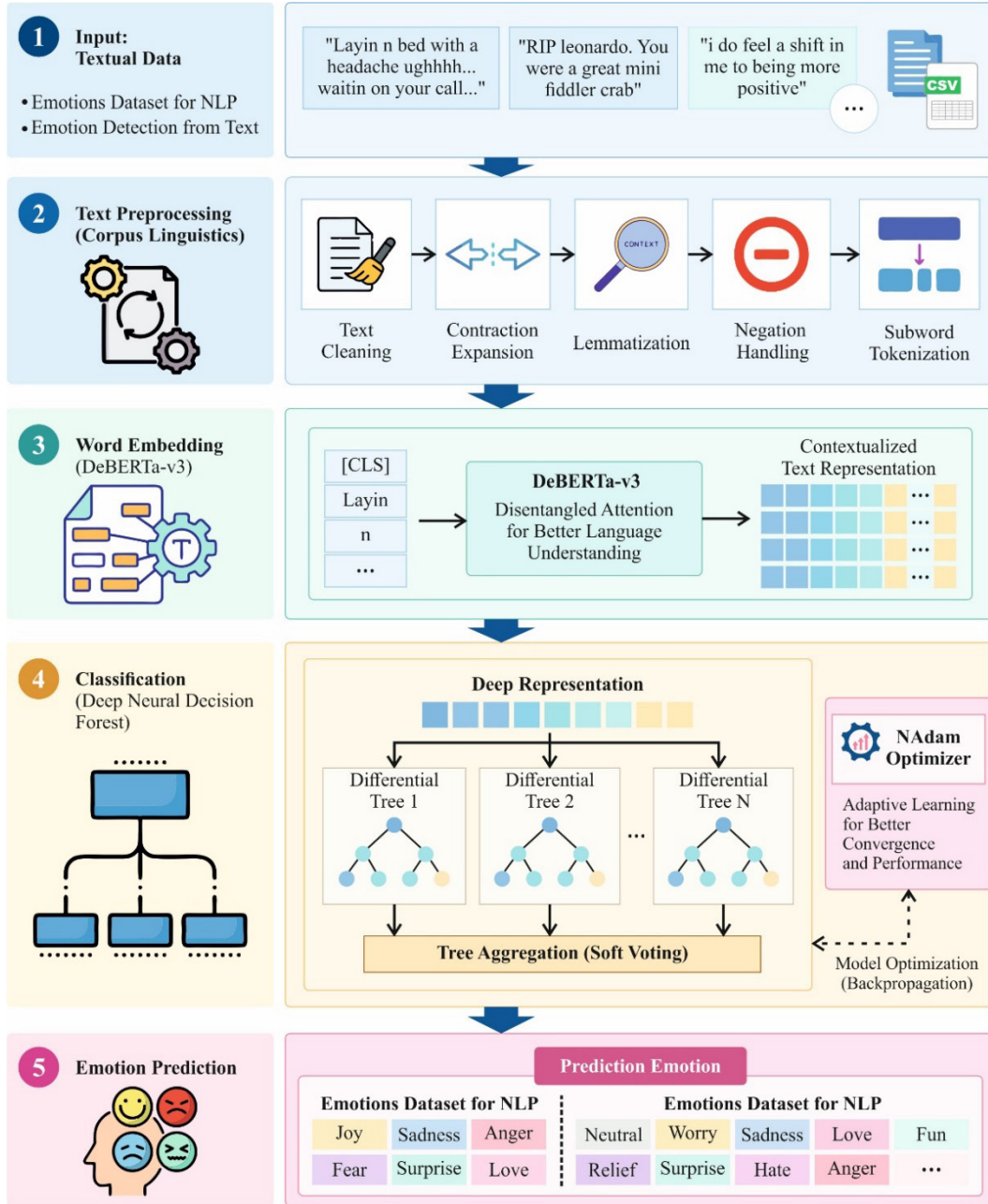


Fig. 1. Complete workflow of the proposed model

A. Text Preprocessing

The preprocessing stage is performed to clean and prepare the input text for further analysis. The text is processed using text cleaning, contraction expansion, lemmatization, negation handling, and subword tokenization [20]. **Text Cleaning:** Text cleaning is used to eliminate unwanted elements that might impact the examination. Unnecessary symbols, punctuation marks, extra spaces, URLs, and other redundant content are eliminated from the text. It retains the meaningful data while minimizing noise in the input features. The cleaned text is further forwarded to the subsequent pre-processing phases.

Contraction Expansion: Contractions such as “isn’t”, “can’t”, and “don’t” are transformed into their full forms, such as “is not”, “cannot”, and “do not”. It supports preserve the actual meaning of the words within text processing. Extending contractions also obtains a constant representation of same expressions. The final text is more appropriate for successive linguistic examination.

Lemmatization: Lemmatization is utilized to transform words into their suitable base or dictionary

forms whereas maintaining their actual meaning. Different simple stemming, it treats the grammatical format and context of a word before minimizing it. For instance, words such as “writing “, “wrote”, and “writes” are converted into “write”. This decreases the difference between relevant words and produces a more consistent representation of the textual characteristics, making it easier to process in classification.

Negation Handling: Negation handling detects phrase or words that variation the meaning of a statement, includes “no”, “not”, and never. These terms were thoroughly preserved so that the original sentence meaning is not lost. For instances, “not happy” must be considered differently from “happy”. This stage is especially meaningful for accurately learning the emotion or sentiment expressed in the text.

Subword-Level Tokenization: Subword-level tokenization splits words into small valuable units before they are provided to the language method. It can manage lengthy, uncommon, and previously unseen words while maintaining beneficial parts of the actual word. This also provides a balance between character-level and word-level tokenization, making it appropriate for processing diverse textual information.

B. Contextual Word Representation using DeBERTaV3

Emotion recognition involves understanding the semantic meaning and contextual interactions during text, as emotional expressions could differ depending on the neighboring words and sentence context. Thus, DeBERTa-V3 are used to capture high contextual representation for semantic-based emotion detection [21]. Unlike standard BERT-based encoders, DeBERTa divides the content data of a token from its positional data within the attention calculation. It enables the model to learn both the meaning of all words and its location in the sequence. Given a sequential input text , the hidden vector can be obtained by the embedding layer and numerous Transformer layers. The disentangled attention module considers text-to-text and text-to-position relationships. The attention value is represented as here, and imply the content- and position-based queries, while and denote the content- and position-based keys. The denotes the key dimensions. For dividing these components, DeBERTa-V3 can model the interaction between words without considering positional data as part of the word representation itself. Fig. 2 signifies the architecture of DeBERTaV3 model.

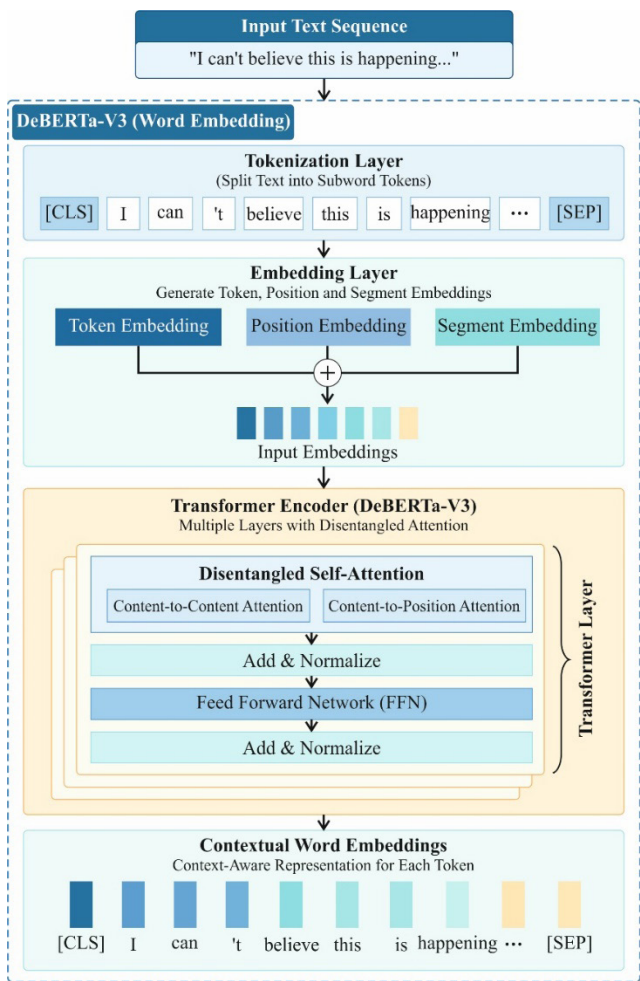


Fig. 2. Architecture of DeBERTa-V3 model

DeBERTa-V3 also utilizes an enhanced pre-training objective that depends on replaced token detection. In pre-training, the method learns to classify whether a token is substituted with another token instead of only rebuilding a masked word. It offers more robust contextual representations and supports the encoder in differentiating among the original and replaced tokens.

In input sequence, the resulting hidden embeddings can be represented by h_i , denotes the contextual feature of the i -th token. The representation of the special [CLS] token is utilized as the overall representation of the input content:

Let S denote the global semantic data of the content, while the whole hidden sequence maintains token-level data for subsequent processing.

C. DNDF-based Emotion Detection

Emotion classification is challenging because contextual and semantic information can have complex relationships, making it difficult to distinguish between emotions that share similar characteristics. Then, DNDF is employed to capture these risk patterns using a deep neural network and identify the resulting characteristics through a decision forest [22]. The learned semantic representations effectively differentiate among various emotion types, and their integration helps the proposed model. At first, the NN converts the input data into an informative feature representation. Consider an input vector x , the fully connected layer output can be determined by $h = Wx + b$, specifies its parameters, σ denotes the activation function, and T denotes the transformation performed by the neural network. To define the routing of every sample, the representation is applied through the decision nodes of the forest.

Let n be a decision node, the probability of routing can be stated by p_n , and θ_n represent the n -th decision node parameters, and ϕ refers to the probability of directing the sample through one branch of the tree, and f denotes the learned feature representation. The complementary branch is predetermined by $1 - \phi$.

As the instance is fed into successive nodes of the decision tree, the probabilities of routing are incorporated to attain the probability of reaching a specific leaf. Given a leaf l , its routing probability can be quantified by $P_l = \prod_{n \in \text{path}(l)} \phi_n$, where $\text{path}(l)$ represents the set of decision nodes along the path to leaf l , and ϕ_n indicates the branch selected at node n . Every leaf includes a distribution of class probabilities. The prediction of an individual tree is further accomplished by incorporating the routing probabilities along the tree with the class distribution correlated with leaves:

here P_l specifies the probability set to class c at leaf l . Let a forest be T trees, the final output can be gained by averaging every tree prediction.

The final probability with the highest class is chosen as the predicted category:

Accordingly, the NN extracts the feature representation, whereas the decision forest utilizes the learned representation to discretize the classes. The integration permits DNDF to manage nonlinear feature correlations though helping the ensemble structure of different decision trees.

D. Hyperparameter Tuning using NAdam Optimizer

The emotion classification performance is based not only on the learned text representations but also on the efficient optimization parameters. Hence, NAdam is used to facilitate parameter optimization with adaptive learning rates, utilizing Nesterov momentum to enhance the training process convergence [23]. Nesterov-accelerated Adaptive Moment Estimation (NAdam) integrates the Adam with adaptive learning-rate mechanism for look-ahead property of Nesterov momentum. To control the parameter updates, NAdam applies the first- and second-order moments of the gradients. Let the gradient g_t at iteration t be the first- and second-moment estimates are evaluated by m_t and v_t specify their decay rates, and β_1 and β_2 imply the first- and second-moment estimates, respectively. Initially, these estimates are set at 0, bias correction is utilized by $\hat{m}_t$ in the first-moment estimate, NAdam combines the Nesterov look-ahead term. The resulting momentum term can be stated by

The parameters of the model are subsequently updated by

here denotes the learning rate, depicts a small constant added for numerical stability, and indicates the updated parameters. Using Nesterov momentum and adaptive gradient scaling, NAdam constrains updates to parameters while enhancing the convergence stability of the emotion classification model during training.

IV. Result Analysis

This section discusses the experimental results of the proposed DNDF-SED framework and evaluates its performance on the Emotions Dataset for NLP and Emotion Dataset Raw.

A. Dataset Description

The performance evaluation of the DNDF-SED technique was carried out on the Emotions dataset for NLP (Dataset 1) [24] and the Emotion Dataset Raw (Dataset 2) [25]. A short explanation of the two datasets is summarized below.

Dataset 1: The Emotions Dataset is a set of text records with diverse emotions including joy, sadness, anger, fear, love, and surprise. It is frequently applied in NLP to train models to recognize the emotional meaning beyond the sentences to be written. Every text is combined with an emotional class label, which makes it suitable for training and testing emotion-detection results. This dataset can help to construct applications like chatbots, sentiment analysis, and systems that identify emotions from text. Table 1 shows the distribution of 20,000 text instances over six emotions, with joy having the highest-class count and surprise the lowest class count.

TABLE I. DATASET DETAILS FOR THE EMOTIONS DATASET FOR NLP

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

Dataset 2: The Emotion Dataset Raw is a labeled dataset that is designed for sentiment analysis and emotion detection tasks in NLP. The dataset consists of two main categories: Text and Emotion. This dataset offers a base for training and assessing models for emotion classification and can be further applied in various applications, including chatbots, recommendation systems, and psychological health monitoring. The dataset is suitable for supervised learning techniques and fine-tuning NLP approaches for emotion understanding. This dataset consists of 34,792 samples with eight different emotion classes, as illustrated in Table 2.

TABLE II. DATASET DETAILS FOR EMOTION DATASET RAW

Class Labels	Class Count
Joy	11045
Sadness	6722
fear	5410
Anger	4297
Surprise	4062
Neutral	2254
Disgust	856
Shame	146
Total	34792

B. Overall Classification Performance of Dataset 1

Fig. 3 exhibits the confusion matrices for (a-b) 70:30 on Dataset 1. The confusion matrices denote that the model performs well in both training and testing phases. Joy (4625) and sadness (3958) samples are accurately classified in training, whereas in the testing phase, Joy (1968) and sadness (1711) samples have maximum values focused along the diagonal. In contrast, few instances are confused with the same emotions. Surprise has relatively less precise predictions, particularly in the testing phase. The outcomes show good and reliable emotion classification with small misclassification.

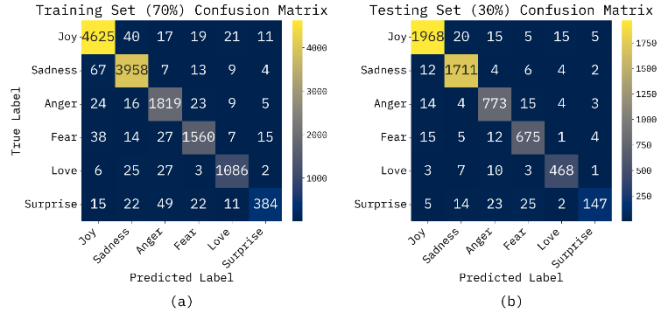


Fig. 3. Confusion matrices for the classification performance of the model on (a-b) 70:30 training-testing split for Dataset 1.

Table 3 depicts the emotion detection performance of the DNDF-SED method for 70:30 on Dataset 1. The results imply that the model obtains an overall accuracy of 95.94% in the training set, where accuracy across the six emotion types was $98.65\% \pm 0.32$, along with an average MCC of $92.83\% \pm 4.54$, respectively. For the testing set, it obtains an overall accuracy of 95.70%, and the average accuracy and MCC were $98.57\% \pm 0.32$ and $91.79\% \pm 6.38$, demonstrating consistent classification outcomes across the six emotion categories.

TABLE III. EMOTION CLASSIFICATION RESULTS OF THE DNDF-SED ALGORITHM USING 70:30 ON DATASET 1

Classes	Accu- racy	Preci- sion	Recall	F1- Score	MCC
Training Set (70%)					
Joy	98.16	96.86	97.72	97.29	95.89
Sadness	98.45	97.13	97.54	97.33	96.24
Anger	98.54	93.47	95.94	94.69	93.86
Fear	98.71	95.12	93.92	94.52	93.79
Love	99.14	95.01	94.52	94.76	94.30
Sur- prise	98.89	91.21	76.34	83.12	82.90
Average	98.65	94.80	92.66	93.62	92.83
$\pm$ SD	± 0.32	± 2.02	± 7.43	± 4.84	± 4.54
Overall Accuracy: 95.94					
Testing Set (30%)					
Joy	98.18	97.57	97.04	97.31	95.94
Sadness	98.70	97.16	98.39	97.77	96.86
Anger	98.27	92.35	95.08	93.70	92.71
Fear	98.48	92.59	94.80	93.68	92.83
Love	99.17	94.74	95.12	94.93	94.48
Sur- prise	98.60	90.74	68.06	77.78	77.92
Average	98.57	94.19	91.42	92.53	91.79
$\pm$ SD	± 0.32	± 2.53	± 10.52	± 6.79	± 6.38

Fig. 4 denotes the ROC and PR curves for the DNDF-SED method on Dataset 1. Fig. 4a exhibits the training ROC curves stay close to the top-left corner, representing remarkable class-wise discrimination with AUC values 0.99 approximately. Fig. 4b implies the testing ROC curves that also remain close to the top-left corner, showing that the model sustains robust classification performance on hidden data. Fig. 4c portrays that the training PR curves indicate reliably high precision across maximum recall levels, while surprise shows a relatively greater decline at high recall. Fig. 4d exhibit that the testing PR curves sustain high precision for most categories, whereas surprise shows the strongest decrease in precision, with AP values of 0.881.

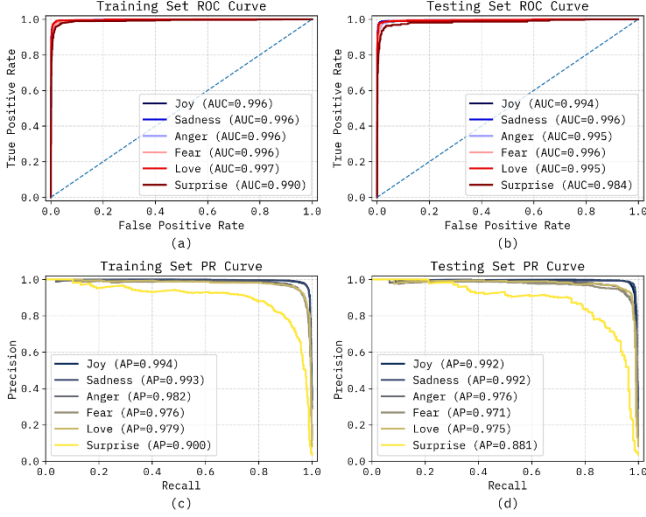


Fig. 4. (a-b) ROC and (c-d) PR curve for six emotion classes of the DNDF-SED system on Dataset 1

C. Overall Classification Performance of Dataset 2

Fig. 5 exhibits the confusion matrices for the 70:30 training and testing phase for eight emotion categories for Dataset 2. The matrices denote that the majority of the instances are properly categorized, as shown by the robust diagonal values. Joy and Sadness attain predominantly better recognition. Some confusion happens among closely associated emotions. The testing outcomes follow the same pattern as the training outcomes, signifying that the model generalizes relatively well.

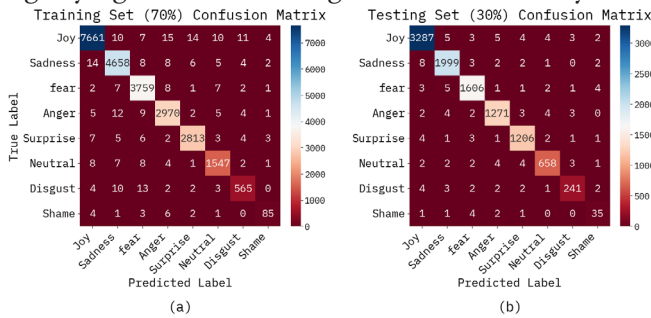


Fig. 5. Confusion matrices for (a-b) 70:30 training and testing set for eight emotion categories on Dataset 2

Table 4 represent the emotion detection performance for the DNDF-SED approach for 70:30 on Dataset 2. The outcomes show that the model attain the general accuracy of 98.78% in training phase, while accuracy across the eight different emotion types was $99.70\% \pm 0.10$ along with average MCC of $96.44\% \pm 4.37$, correspondingly. In the testing phase, it attains an overall accuracy of 98.71%, and the average accuracy and MCC were $99.68\% \pm 0.08$ and $95.29\% \pm 7.09$, representing reliable classification results among eight emotion classes.

TABLE IV. EMOTION DETECTION RESULTS OF THE DNDF-SED MODEL WITH 70:30 IN DATASET 2

Classes	Accu- racy	Preci- sion	Recall	F1- Score	MCC
Training Set (70%)					
Joy	99.53	99.43	99.08	99.26	98.91
Sadness	99.59	98.90	99.00	98.95	98.70
fear	99.66	98.58	99.26	98.92	98.72
Anger	99.66	98.51	98.74	98.62	98.43
Sur- prise	99.76	99.01	98.94	98.98	98.84
Neutral	99.73	97.85	98.04	97.94	97.80
Disgust	99.75	95.44	94.32	94.88	94.75
Shame	99.88	87.63	83.33	85.43	85.39
Average	99.70	96.92	96.34	96.62	96.44
± SD	± 0.10	± 3.70	± 5.15	± 4.43	± 4.37
Overall Accuracy: 98.78					
Testing Set (30%)					
Joy	99.52	99.28	99.22	99.25	98.89
Sadness	99.63	98.96	99.11	99.03	98.80
fear	99.66	98.83	98.95	98.89	98.69
Anger	99.66	98.68	98.60	98.64	98.45
Sur- prise	99.71	98.61	98.93	98.77	98.61
Neutral	99.69	97.92	97.34	97.63	97.46
Disgust	99.74	95.63	93.77	94.70	94.57
Shame	99.80	74.47	79.55	76.92	76.86
Average	99.68	95.30	95.68	95.48	95.29
± SD	± 0.08	± 7.94	± 6.33	± 7.15	± 7.09
Overall Accuracy: 98.71					

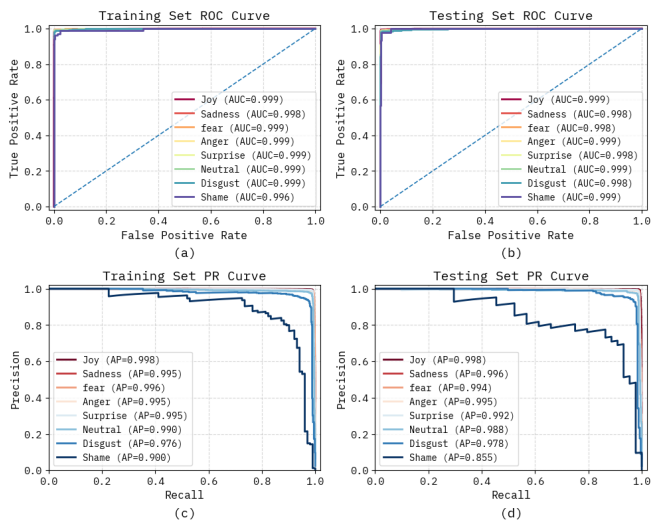


Fig. 6. (a-b) ROC and (c-d) PR curves for eight emotion classes of the DNDF-SED system on Dataset 2

Fig. 6a signifies that the training ROC curves remain near the top-left corner, with AUC values of approximately 0.996–0.999, indicating remarkable class discrimination. Fig. 6b displays the testing curves represent the same robust performance, with AUC values above 0.998 for most of the emotions, representing good generalization

to unseen data. Fig. 6c describes the training precision–recall curves stay close to the top, with high AP values, indicating robust precision under most of the recall stages. Fig. 6d indicate the testing outcomes sustain high precision for the majority of the emotions. In contrast, Shame and Disgust indicate a visible decrease at higher recall, reflected in their lower AP values.

D. Comparative Analysis and Ablation Study

Table 5 and Fig. 7 portray the comparison analysis of the DNDF-SED technique for both Dataset 1 and 2 [26, 27]. On Dataset 1, the findings suggest that the DNDF-SED system gained stable results with 95.70% accuracy, 94.19% precision, 91.42% recall, and 92.53% F1-score. Meanwhile, Bi-LSTM-CNN [27] method outperforms least outcome whereas ERNIE-CNN [27], ERNIE-BiLSTM [27], and DistilBERT [26] system achieves somewhat inefficient performance. Moreover, BERT-base [26] and XLNet-base [26] algorithms accomplish moderate results. RoBERTa-base [26] methodology performs slightly below the proposed framework with an accuracy of 92.45%. Similarly, in Dataset 2, the results reveal that the DNDF-SED technique attains superior results with 98.71% accuracy, 95.30% precision, 95.68% recall, and 95.48% F1-score. Bi-LSTM-CNN [27] approach performs the worst, while the ERNIE-CNN [27], ERNIE-BiLSTM [27], and DistilBERT [26] approaches achieve somewhat higher results. Besides, BERT-base [26] and XLNet-base [26] methodologies attain moderate performance. RoBERTa-base [26] technique achieved slightly lower performance than the proposed system, with an accuracy of 93.84%, signifying the effectiveness and robustness of the proposed system when compared to other existing methodologies.

TABLE V. COMPARISON STUDY OUTCOME OF THE DNDF-SED METHODOLOGIES FOR BOTH DATASET 1 AND 2

Methods	Emotions dataset for NLP				Emotion Dataset Raw			
	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
XLNet-base [26]	91.00	90.90	91.00	90.80	92.84	92.37	92.16	92.25
RoBERTa-base [26]	92.45	92.50	92.40	92.50	94.16	93.84	93.57	93.70
DistilBERT [26]	88.40	88.30	88.40	88.10	90.72	90.18	90.35	90.21
BERT-base [26]	90.10	90.00	90.10	89.90	92.03	91.58	91.74	91.61
Bi-LSTM-CNN [27]	68.15	68.15	67.61	68.08	70.46	69.73	69.91	69.62
ERNIE-BiLSTM [27]	86.41	86.41	85.97	86.24	88.65	88.12	88.31	88.19
ERNIE-CNN [27]	78.64	78.64	77.91	78.47	81.27	80.64	80.91	80.72
DNDF-SED	95.70	94.19	91.42	92.53	98.71	95.30	95.68	95.48

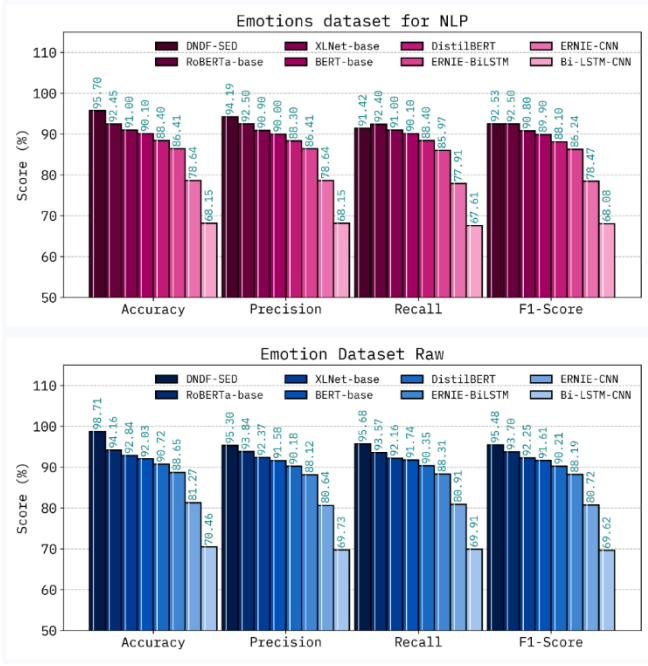


Fig. 7. Comparison performance of the DNDF-SED approach for both Dataset 1 and 2

TABLE VI. INFERENCE TIME OF THE DNDF-SED METHOD WITH VARIOUS MEASURES

Methods	Inference Time (ms)	FLOPs (G)	Params (M)	GPU (GB)
XLNet-base	28.60	22.41	117.31	3.18
RoBERTa-base	26.40	21.75	125.24	3.05
DistilBERT	17.80	11.29	66.36	2.14
BERT-base	24.70	21.46	109.48	2.92
Bi-LSTM-CNN	14.90	8.37	18.62	1.76
ERNIE-BiLSTM	21.80	15.64	34.27	2.31
ERNIE-CNN	16.70	9.82	22.45	1.89
DNDF-SED	11.60	6.18	16.84	1.42

Table 6 and Fig. 8 signify the inference time for the DNDF-SED method with multiple metrics. According to the inference time, the presented DNDF-SED approach attains the minimum outcome with inference time (11.60 ms), FLOPs (6.18 G), parameter count (16.84 M), and GPU usage (1.42 GB). Among transformer models, DistilBERT is more lightweight and quicker than XLNet, RoBERTa, and BERT, as it requires less computation and memory. Bi-LSTM-CNN and ERNIE-CNN also require moderately less computation when compared to other larger transformer-based models. The outcomes demonstrate that the proposed method offers a better balance, which makes it well-suited for resource-constrained applications.

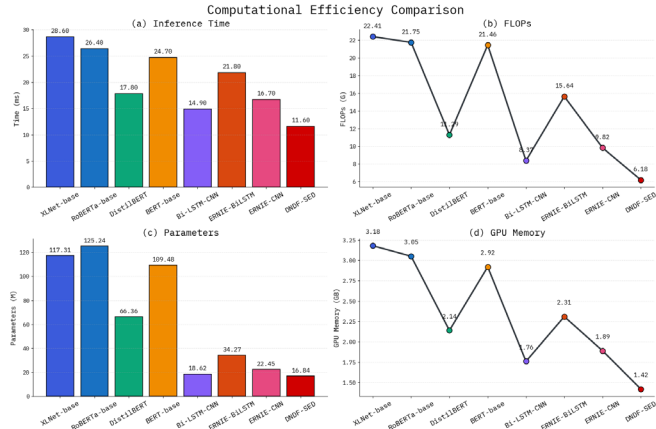


Fig. 8. Computational Efficiency comparison of the DNDF-SED system

Table 7 demonstrates the ablation study analysis of the DNDF-SED approach for Dataset 1. The baseline DNDF model obtains a lower accuracy of 89.46%, with an F1-score of 86.26%. When integrating DeBERTa-v3 with the DNDF algorithm, the accuracy rises to 92.18%, while the F1-score improves to 89.67%, indicating that the robust contextual representation provided by DeBERTa-v3 helps the model differentiate among the relevant patterns more efficiently. Finally, the proposed DNDF-SED (DeBERTa-v3 + DNDF + Nadam) algorithm has achieved stable results with 95.70% accuracy, 94.19% precision, 91.42% recall, and 92.53% F1-score, providing the most effective configuration among the evaluated methodologies.

TABLE VII. ABLATION STUDY PERFORMANCE OF THE DNDF-SED SYSTEM FOR DATASET 1

Methods	Accuracy	Precision	Recall	F1-Score
DNDF	89.46	87.82	84.76	86.26
DeBERTa-v3 + DNDF	92.18	90.74	88.63	89.67
DNDF-SED (DeBERTa-v3 + DNDF + Nadam)	95.70	94.19	91.42	92.53

ABLATION STUDY OUTCOME OF THE DNDF-SED APPROACH ON DATASET 2

Methods	Accuracy	Precision	Recall	F1-Score
DNDF	93.12	89.46	90.18	89.81
DeBERTa-v3 + DNDF	96.42	92.84	93.16	92.99
DNDF-SED (DeBERTa-v3 + DNDF + Nadam)	98.71	95.30	95.68	95.48

Table 8 presents the ablation study analysis of the DNDF-SED algorithm on Dataset 2. The baseline DNDF method performs worse, with an accuracy of 93.12%, while incorporating DNDF with the DeBERTa-v3 method increases accuracy to 96.42%. Finally, the proposed DNDF-SED (DeBERTa-v3 + DNDF + Nadam) framework obtains better results with 98.71% accuracy, 95.30% precision, 95.68% recall, and 95.48% F1-score.

E. Discussion

The findings demonstrate that the proposed DNDF-SED model can efficiently differentiate among various emotions from textual information. The contextual representations attained from DeBERTa-v3 facilitate the model learns the word meaning associate with their corresponding context. By separating closely related emotion classes, the DNDF classifier helps the classification method to learn intricate patterns. During training, the use of NAdam enhance the model learning procedure to optimize the parameters. On both datasets, the reliable outcome is attain indicating that the framework can manage changes in textual emotion expressions. Hence, the results validate that incorporating deep decision forest classification and contextual language representation is well-suited for semantic-based emotion recognition.

V. Conclusion and Future Work

In this work, we proposed a DNDF-SED framework utilizing textual data. The proposed framework initially pre-processes the text through cleaning, contraction expansion, lemmatization, negation handling, and sub-word tokenization to achieve clean and consistent text for the subsequent word embedding process. After that, DeBERTa-v3 is used to capture contextual and semantic representations from the processed text. These representations are then passed through DNDF classification that combines deep feature learning with decision-tree-based classification to differentiate diverse emotion classes. Furthermore, the Nadam optimizer is used to enhance the model parameters and improve model training. Extensive performance validation indicates the presented DNDF-SED model obtains higher accuracy of 95.70% and 98.71% on Emotions dataset for NLP and Emotion Dataset raw compared with conventional approaches. These results confirm the performance of incorporating contextual language representations with deep decision forest classification for semantic-based emotion detection system. Future research will concentrate on using the DNDF-SED model to massive datasets involving a broad range of emotional expressions. The method will then adapt to multilingual and code-mixed text to enhance its applicability to diverse language configuration. Additionally, alternative language and

optimization methods could be analyzed to enhance the classification performance of the model. The technique can also be expanded with more interpretability approaches to offer a clearer understanding of the decisions made for all emotion classes.

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/19690370> reference number [24].

Funding note and acknowledgement: 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

- 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, R., Ma, Y. and Li, H., 2026. Hybrid-Scale Heterogeneous Graph Fusion for Multimodal Intent Recognition. *IEEE Transactions on Affective Computing*.
- Yu, D., Tang, Y., Zhang, C., Wang, W., Yang, G., Zheng, X. and Zhao, Y., 2026. IA 2 GNN: Imbalance-Aware Adaptive Graph Construction for Multi-Modal Image Fusion. *IEEE Transactions on Multimedia*.
- Rai, A., Yadav, D., Kumar, D. and Tiwari, S., 2026, February. Transcripto: An Integrated Study on Inclusive, Multimodal, and Semantic-Aware Speech-to-Text (STT) Systems. In 2026 International Conference on Sustainable Engineering and Technology Innovations (ICSETI) (pp. 606-613). IEEE.
- Li, L., 2025. Semantic-driven textual speech rhythm assessment with contextual augmentation and multimodal synchronisation. *Journal of Computational Methods in Sciences and Engineering*, p.14727978251348625.
- 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.
- Sapkota, R., Raza, S., Shoman, M., Paudel, A. and Karkee, M., 2025. Multimodal large language models for image, text, and speech data augmentation: A survey. *arXiv preprint arXiv:2501.18648*.
- Liu, R., Ma, Y. and Li, H., 2026. Hybrid-Scale Heterogeneous Graph Fusion for Multimodal Intent Recognition. *IEEE Transactions on Affective Computing*.
- Zhang, Y. and Zhu, Z., 2026, April. Application of a label-semantic-guided symmetrical attention mechanism in short-text multilabel classification. In *Second International Conference on Artificial Intelligence, Deep Learning, and Neural Networks (AIDLNN 2025)* (Vol. 14182, pp. 64-71). SPIE.
- Venkatesh, R., Nandhakumar, A., Marimuthu, S., Yashashwini, S., Vishnu, S.K. and Sathiyamurthy, S., 2025, November. Transformer-diffusion hybrid for enhancing digital twin precision in aerospace structural maintenance. In 2025 International Conference on Electrical Engineering and Informatics (ICEEI) (pp. 1-8). IEEE.
- 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.
- Sun, H., Zhang, J., Hong, W., Fang, Y., Ma, M., Shi, K. and Li, M., 2026. A Cross-Modal Temporal Alignment Framework for Artificial Intelligence-Driven Sensing in Multilingual Risk Monitoring. *Sensors*, 26(8), p.2319.
- 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).
- Sun, F., Li, L., Dai, L. and Fan, X., 2026. Empathetic Motion Generation for Humanoid Educational Robots via Reasoning-Guided Vision--Language--Motion Diffusion Architecture. *arXiv preprint arXiv:2603.18771*.
- Tian, T., Wu, B. and Chen, W., 2026. Cross-modal GNN--GRU framework for keyframe extraction and semantic-driven special-effects generation in animation videos. *Discover Applied Sciences*.
- Huo, Y., Chan, W.H., Amerhaider Nuar, A.N.B. and Gao, H., 2025. SBT-Net: a tri-cue guided multimodal fusion framework for depression recognition. *BioData Mining*, 18(1), p.86.
- Yuan, Y., Duo, S., Tong, X. and Wang, Y., 2025. A multimodal affective interaction architecture integrating BERT-based semantic understanding and VITS-based emotional speech synthesis. *Algorithms*, 18(8), p.513.
- Hu, J., Qu, L., Li, H. and Li, T., 2025. Label semantic-driven contrastive learning for speech emotion recognition. In *Proc. Interspeech 2025* (pp. 4348-4352).
- Mete, A.M. and Dhore, M.L., 2026. Analysis of Tokenization Techniques in the Context of Morphology for Multilingual Inputs. *Engineered Science*, 40, p.2094.
- Wen, X., Wen, X. and Dai, X., 2026, February. Enhanced cross-modal representation method for multi-modal

- rumor detection: A fusion framework based on DeBERTa-V3 and ViT-Large. In *Proceedings of the 2026 International Conference on Artificial Intelligence and Control* (pp. 171-176).
- Alrayes, F.S., Zakariah, M., Driss, M. and Boulila, W., 2023. Deep neural decision forest (DNDF): A novel approach for enhancing intrusion detection systems in network traffic analysis. *Sensors*, 23(20), p.8362.
- Liu, X., Qi, H., Jia, S., Guo, Y. and Liu, Y., 2025. Recent advances in optimization methods for machine learning: a systematic review. *Mathematics*, 13(13), p.2210.
<https://doi.org/10.34740/kaggle/dsv/19690370>
- 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.
- 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.