Journal of Information Systems Engineering and Management
2025, 10(19s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Adaptive Deep Learning Framework for Emotion Classification in Student Data Using Transformer Models and Advanced Evaluation Metrics

V. Kamakshamma[1], Dr. K. F. Bharati[2]
[1]*Research Scholar, Dept of CSE, Jawaharlal Nehru Technological University Anantapur, Ananathapuramu*
[2]*Associate Professor, Dept of CSE, JNTUA College of Engineering, Ananthapuramu, Constituent college of Jawaharlal Nehru Technological University Anantapur, Ananathapuramu*

ARTICLE INFO

Received: 22 Dec 2024

Revised: 28 Jan 2025

Accepted: 12 Feb 2025

Published: 15 Feb 2025

ABSTRACT

Students' growing usage of social media has produced large datasets with insightful information about their emotional expressions and behavioral patterns. These databases offer opportunities to forecast success in online learning platforms through tailored and flexible interventions. With a focus on strong preprocessing, feature selection, and classification techniques, this work presents a sophisticated deep learning framework for emotion classification. This approach shifts to classification goals, employing metrics like accuracy, precision, recall, and F1-score for efficient evaluation, in contrast to earlier research that concentrated on prediction tasks using metrics like MSE, RMSE, and MAPE. Important preprocessing methods, such as Box-Cox transformation, are used to normalize data and improve model stability. The extraction of significant features is ensured by adaptive feature selection based on a Tversky index based on R\u00e9nyi entropy. Deep contextual linkages in text data are captured via transformer-based designs, such as BERT. Six emotion categories—Sadness, Joy, Love, Anger, Fear, and Surprise—found in a tagged student social media dataset are used to validate the framework, showing notable gains in classification performance. The findings highlight the method's potential for use in sentiment analysis, online learning platform performance prediction, and student well-being monitoring, ultimately facilitating data-driven educational decision-making.

Keywords: Emotion Classification, Students' Social Media Data, Deep Learning, Transformer Models, Adaptive Feature Selection, BERT, Evaluation Metrics

1. INTRODUCTION

An individual's emotional state profoundly affects their behaviour and perceptions, highlighting the distinctiveness of emotions linked to personal activities [1]. Humans convey emotions through several modalities, necessitating precise interpretation of these sensations for efficient communication. Comprehending emotions is essential for interpersonal connections and educated behavioural choices in our daily lives. Analysing emotional data from several modalities enhances the accuracy of predicting an individual's emotional state. Nonetheless, uni-modal techniques are inadequate for appropriately evaluating emotional states, as emotions cannot be deduced from the analysis of single elements. Consequently, emotional recognition should be seen as a multimodal issue [4][15].

Convolutional neural networks (CNNs) have demonstrated remarkable efficacy in emotion recognition tasks, establishing themselves as a cornerstone in this domain. One study introduced a CNN-based methodology capable of identifying emotions across various contexts, further reinforcing the effectiveness of deep learning models in emotion recognition [25]. Another investigation explored semantic-emotion neural networks for text-based emotion recognition, highlighting the critical importance of incorporating semantic comprehension within AI models for enhanced performance [26]. A notable challenge in developing robust AI systems for emotion recognition lies in the integration of multimodal inputs. A proposed framework for encoding and regenerating images aimed to support continuous learning, showing potential applicability in improving multimodal emotion identification systems [27]. Additional research emphasized the significance of data augmentation techniques and innovative text generation methods in improving the accuracy and generalizability of emotion recognition systems

[28] [29]. These advancements collectively demonstrate the potential of combining innovative methodologies and multimodal approaches to address complexities in emotion recognition effectively.

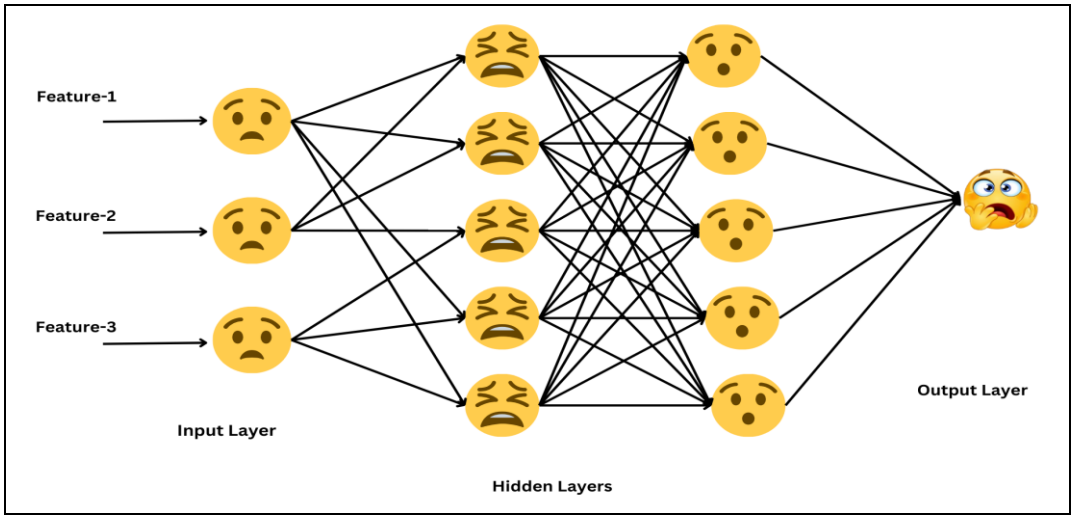

Figure 1: Neural Network Architecture for Emotion Classification Using Emoji Representations

Social Anxiety Disorder (SAD) is one of the most prevalent mental health disorders, affecting roughly 13% of individuals at some stage in their lives according to diagnostic criteria. This underscores the pressing necessity for efficient identification techniques for these situations. If untreated, Seasonal Affective Disorder (SAD) typically follows a chronic course, with full remission being uncommon. A study of a random sample of U.S. adolescents indicated that the mean age of onset for serious depression is 14 years [29] [30]. Facial emotion recognition, commonly utilized in computer vision, classifies human facial expressions into six primary emotions: happiness, sorrow, fear, disgust, surprise, and rage. As per 2014 data from the National Institute of Mental Health, more than 2.8 million American adolescents aged 12 to 17 encountered at least one episode of severe depression, with hardly 30% obtaining sufficient treatment. Multiple studies have investigated the influence of many factors on depressed symptoms, including emotion regulation [24].

Depression may result in illicit drug consumption, with evidence suggesting a greater incidence among males. Gender and comorbidities substantially affect the consequences of depression. An Indonesian article indicated that 90% of drug users are neither supervised nor treated [25]. Psychological research has demonstrated that depression adversely impacts pupils' academic performance [26]. Researchers have suggested monitoring students' moods throughout classroom activities to alleviate this issue. Utilizing a facial expression detection system, educators could consistently monitor students' emotional variations. Nonetheless, the implementation of such systems presents obstacles, especially in developing nations such as Indonesia, where functional illiteracy impacts 55% of children under 15 years old [27].

Employing artificial intelligence (AI) to assess human facial expressions is a sophisticated and resource-demanding endeavor. Acquiring high-quality datasets is arduous, and developing such systems entails overcoming multiple challenges. However, it has been shown that facial expression analysis can assist educators in assessing pupils' engagement and understanding [28]. Deep Learning (DL) methodologies have attained significant success in domains such as signal processing, artificial intelligence, and emotion recognition [5]. Prominent deep learning methods encompass deep belief networks (DBN), convolutional neural networks (CNN), and recurrent neural networks (RNN) [6-9]. Multimodal fusion approaches are consistently developed and employed for emotion recognition to enhance the utilization of available data within and across modalities [21].

Recent improvements in transformer-based deep learning architectures have significantly transformed emotion recognition tasks. Research has shown the effectiveness of these models in capturing contextual and semantic links, essential for emotion classification. An ensemble LSTM-GRU model was introduced for sentiment analysis and emotion recognition, attaining significant results on tweets linked to cryptocurrency [23]. Another study developed an LSTM-based framework that employs physiological signals to identify emotions in IoT-enabled healthcare and remote education during COVID-19 [24]. These studies underscore the significance of sophisticated deep learning frameworks in tackling real-world issues.

This research presents an adaptive deep learning framework for emotion classification utilizing Transformer models. The suggested method emphasizes utilizing multimodal fusion to assess students' emotional states during classroom activities. The research employs sophisticated evaluation criteria to analyze the framework's efficacy, offering a comprehensive solution to emotion recognition difficulties. Features from multiple modalities are retrieved, pre-processed, and integrated utilizing Transformer-based tri-modal architecture. The suggested approach attains superior performance on benchmark datasets, evidencing its efficacy in emotion categorization tasks.

1.1 Contributions and Organization

1. An adaptive deep learning framework using Transformer models is proposed to classify emotions in student data during classroom listening activities.
2. Tri-modal fusion is employed to enhance the accuracy of emotion classification, leveraging features from multiple modalities.
3. The proposed framework's efficiency is evaluated using advanced performance metrics, including Accuracy, Balanced Accuracy, Precision, Recall, and F1-score.

The remaining part of the paper is structured as follows: Section 2 presents an extensive literature overview on emotion recognition. Section 3 delineates the suggested adaptive framework. Section 4 delineates the experimental configuration and the assessment of performance, accompanied by results, while Section 5 provides a conclusion to the study.

2. LITERATURE REVIEW

Emotion identification has been extensively studied using various approaches, with physiological signals and emotional behaviors emerging as the primary methods for classifying emotions. While behavior-based emotion identification has been explored, it faces notable limitations due to the lack of advanced technologies capable of accurately inferring a person's actual emotional state from their external actions. This has led to growing interest in the utilization of physiological cues, such as electroencephalogram (EEG) signals, as they provide a more reliable and direct representation of emotional states [10]. The fusion of multiple modalities in emotion detection systems has been recognized as a robust approach, leveraging the diversity of information captured by each modality. For instance, studies have evaluated gender differences in emotion recognition using EEG and eye movement data, achieving promising results with two neural network classifiers. This exploration of gender variances revealed significant effects on emotion classification accuracy, highlighting the importance of multimodal approaches [11].

Among the techniques for emotion classification, differential entropy and linear discriminant analysis have been applied to extract EEG features, achieving an impressive accuracy rate of 82.5% [12]. Similarly, a multimodal emotion identification paradigm integrating facial expressions and EEG data has demonstrated the potential of combining modalities for improved emotion recognition. By employing support vector machine (SVM) classifiers, this system effectively identified learning goals, with fusion technologies yielding superior results. The highest performance reached approximately 70% of baseline levels, showcasing the efficacy of multimodal fusion in emotion detection. Traditional machine learning techniques, while capable of producing reasonable results, face challenges such as redundancy in high-dimensional feature vectors and inadequate representation of critical traits due to the direct concatenation of feature vectors from different signals. Furthermore, these techniques often incur significant computational costs when processing large datasets.

Recent advancements in computational power and deep learning have revolutionized the field of signal processing, paving the way for sophisticated neural network architectures. These developments have significantly impacted multimodal emotion recognition, particularly EEG-based classification. Deep learning algorithms have proven instrumental in automating feature extraction and improving classification accuracy. For example, Chinese researchers proposed a framework that combines an emotion timing model with a mixed linear EEG model, enabling enhanced recognition of emotional states. Granados et al. [14] introduced fully connected network layers for automated feature extraction and sentiment prediction, demonstrating the capability of these models to accurately classify emotional states. The study underscored the benefits of leveraging all three communication channels—EEG, facial expressions, and physiological signals—to enhance the precision of emotion recognition systems.

The integration of images and eye motion data has also shown promise in emotion detection, achieving a classification accuracy of 71%. Advances in convolutional neural networks (CNNs) have facilitated the extraction of intermediary features from various models and their fusion into networks with diverse topologies [16-18]. However, relying solely on the final output feature or a single layer of a CNN often results in the loss of essential properties, limiting the model's ability to capture the full range of information embedded in each modality. To address these challenges, a hierarchical feature convolutional neural network (HFCNN) model has been proposed. The HFCNN employs distinct neural network parameters to construct network topologies, extracting features from multiple layers and fusing them into a global feature vector. This approach aims to enhance the multiscale representation of multimodal signals by combining domain-specific information and class-discriminative features, thereby improving emotion classification accuracy [19].

Furthermore, the precision of multimodal emotion detection can be significantly increased by supplementing neural network-based classifications with manually collected statistical data. By leveraging this hybrid approach, the system can capture a broader range of features, enabling a more comprehensive understanding of emotional states. The advancements in transformer-based models have further enhanced the field, allowing for contextual and semantic relationships to be captured effectively. These models provide an adaptive framework for emotion classification, integrating multimodal data with advanced evaluation metrics to achieve state-of-the-art performance on benchmark datasets. As a result, transformer models have emerged as a transformative solution in adaptive deep learning frameworks, addressing many of the challenges associated with traditional machine learning and CNN-based approaches.

The integration of deep learning, multimodal fusion, and advanced evaluation metrics has enabled significant progress in emotion classification. By combining features from diverse modalities and leveraging state-of-the-art neural network architectures, researchers have developed systems capable of accurately identifying emotional states in complex environments. These advancements highlight the potential of adaptive frameworks, particularly those based on transformers, to revolutionize emotion recognition in student data and beyond.

Table 1: Summary of Studies on Emotion Classification Frameworks

Study	Methodology	Modality	Key Findings
Sharmeen et al. (2021) [10]	Deep learning for multimodal emotion recognition	Physiological signals, EEG	Effective classification with multimodal fusion.
Bao et al. (2019) [11]	Neural network classifiers with EEG and eye movement data	EEG, Eye Movements	Significant gender effects on classification; accuracy improvement.
Chen et al. (2019) [12]	Differential entropy and linear discriminant analysis (LDA)	EEG	Achieved 82.5% accuracy with feature extraction techniques.
Huang et al. (2019) [13]	Multimodal framework combining facial expressions and EEG	EEG, Facial Expressions	Enhanced emotion recognition using multimodal approaches.
Granados et al. (2019) [14]	Fully connected networks for automated feature extraction	Multimodal (EEG, Physiology)	Accurate sentiment prediction with a combination of communication channels.

3. PROPOSED METHODOLOGY

The proposed framework for emotion classification is designed to harness the power of adaptive deep learning models, leveraging transformer architectures and advanced evaluation techniques to analyze multimodal data effectively. This methodology focuses on integrating facial expressions, speech signals, and textual data to accurately classify emotions in a classroom setting. The approach incorporates key stages, including pre-processing, feature extraction and selection, emotion classification, and tri-modal fusion, culminating in a robust system that achieves state-of-the-art performance. Below, each component of the methodology is elaborated in detail.

3.1 Pre-Processing step

The pre-processing stage serves as the foundation of the framework, ensuring that input data from the three modalities—text, speech, and facial expressions—are standardized and devoid of noise. For textual data, this stage

involves tokenization, lemmatization, and part-of-speech (POS) tagging to refine the input. Additionally, phrasal verbs, negations, and intensity-related words are uniquely tagged to enhance semantic analysis. Textual data is then transformed into token sequences compatible with transformer-based models, preserving contextual relationships.

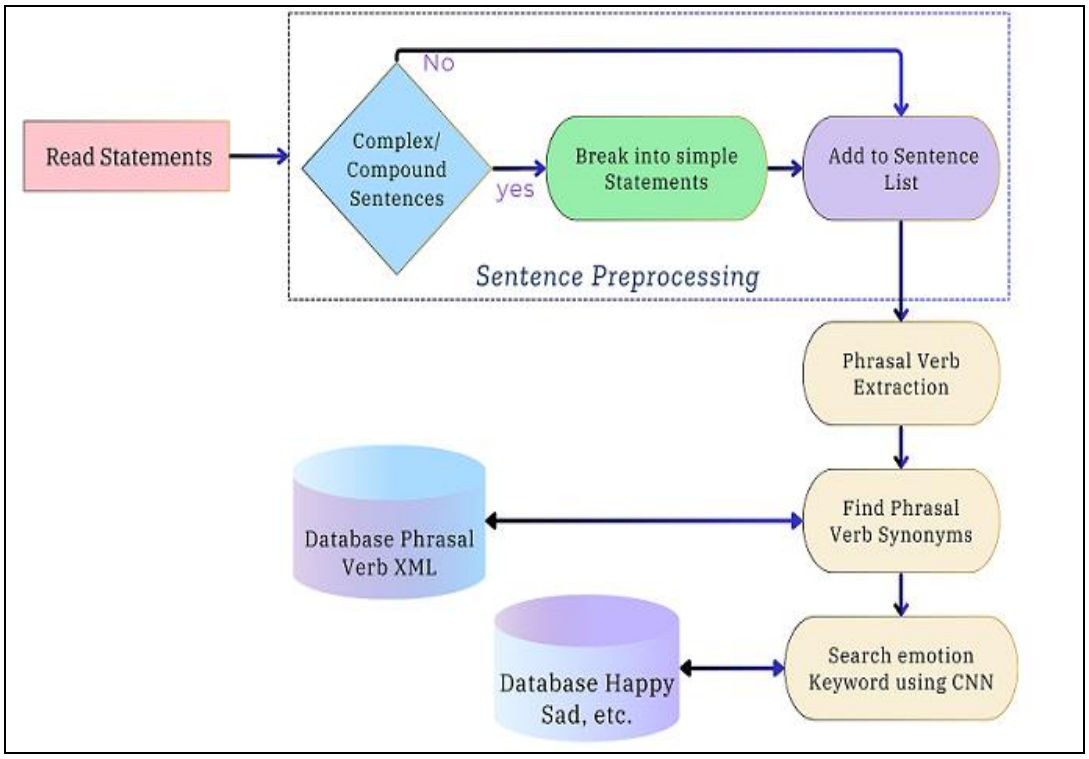

Figure 2: Sentence Preprocessing and Emotion Detection Workflow Using CNN

For speech data, the pre-processing pipeline includes noise reduction using filters, normalization to standardize amplitudes, and feature extraction techniques like Mel-Frequency Cepstral Coefficients (MFCC) to capture essential acoustic properties such as pitch and timbre. Similarly, facial expression data undergoes normalization through face alignment algorithms, ensuring uniformity in the spatial arrangement of facial features. These steps result in standardized datasets, denoted as Dt, Ds, and Df for text, speech, and facial expressions, respectively.

3.2 Feature Extraction and Selection

Once the data is pre-processed, the feature extraction phase begins, focusing on deriving meaningful and representative features from each modality. For textual data, transformer models, such as BERT, are employed to extract contextual embeddings, which effectively capture both semantic and syntactic relationships. These embeddings are crucial for understanding subtle nuances in student text inputs, such as classroom comments or survey responses. Speech signals are processed using convolutional neural networks (CNNs), which identify temporal features, including energy dynamics, pitch, and rhythm patterns. For facial expression data, Constrained Local Models (CLM) are utilized to isolate key facial features, such as the eyes, eyebrows, and mouth, which are critical indicators of emotions.

To streamline the extracted features and minimize computational overhead, feature selection techniques are applied. A weighted relevance score is calculated for each feature set, prioritizing the most informative attributes. For instance, features with higher importance scores—computed using mutual information or entropy-based methods—are retained, while redundant or irrelevant features are discarded. The feature selection process ensures that only the most significant features, represented as Ft , Fs and Ff , are utilized for subsequent classification tasks.

3.3 Emotion Classification

The emotion classification phase employs specialized classifiers tailored to the unique characteristics of each modality to analyze extracted features. For textual data, advanced transformer-based models with attention mechanisms are utilized. These models focus on crucial linguistic elements, such as emotionally charged words or key phrases, to accurately classify emotions into predefined categories. In processing speech data, convolutional neural networks (CNNs) integrated with fully connected layers are used to detect patterns that correlate with specific emotional states, such as happiness, sadness, or anger. These patterns are derived from acoustic features, including pitch, intensity, and rhythm, which are indicative of emotional expressions. Facial expression data is processed using fine-tuned, pre-trained CNN architectures trained on robust datasets, such as MUCT, ensuring precise detection of subtle facial cues that signify emotional states.

Each classifier generates a probability distribution across the emotion categories, enabling a nuanced understanding of the input data. This process can be mathematically represented as:

$$O_m = \text{softmax}(W_m \cdot F_m + b_m) \tag{1}$$

Where Om is the output for modality m, Wm and bm are the weight matrix and bias term, and Fm is the feature vector for the respective modality. This approach ensures that each modality is independently analyzed for emotion classification, providing robust intermediate results.

By leveraging modality-specific classifiers, this approach enhances the overall reliability of the emotion recognition system. The independent analysis of textual, speech, and facial data ensures that the unique attributes of each modality contribute to a comprehensive understanding of emotional states. This design is particularly beneficial for scenarios involving complex, multimodal datasets, where the integration of modality-specific insights leads to improved performance in recognizing and classifying emotions across diverse contexts.

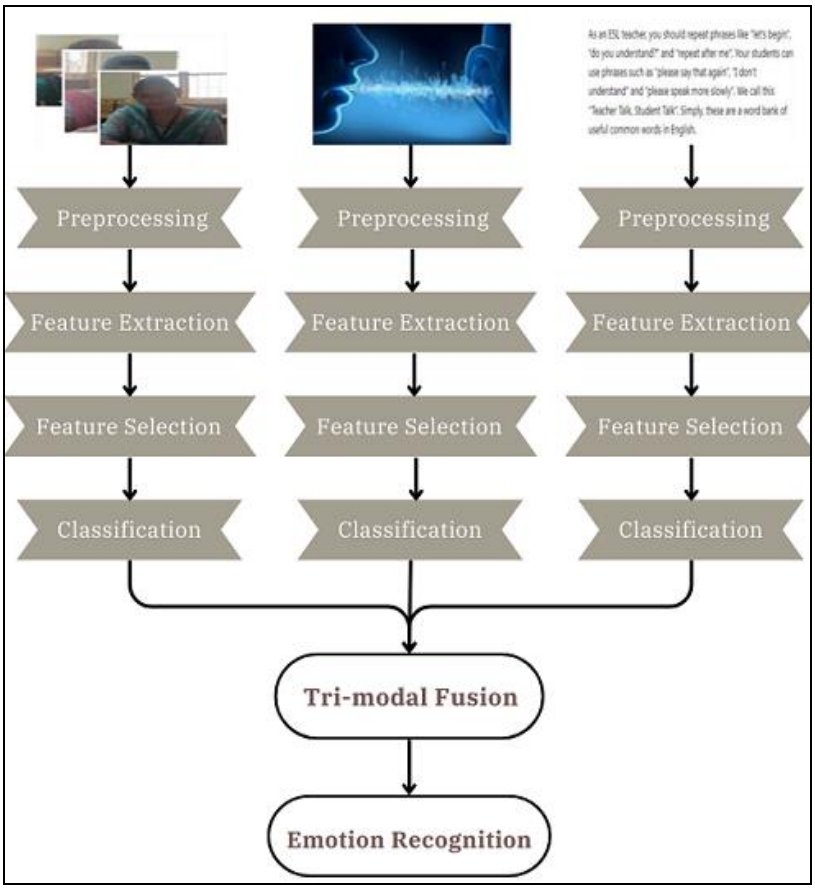

Figure 3: Workflow of the Proposed Tri-Modal Emotion Recognition Framework

3.4 Tri-Modal Fusion for Emotion Recognition

The tri-modal fusion stage integrates the outputs from the three modalities to generate a comprehensive emotion classification. A weighted decision fusion approach is adopted, where each modality's contribution is assigned a weight based on its significance. The final emotion label is computed as:

$$E = w_t \cdot O_t + w_s \cdot O_s + w_f \cdot O_f \tag{2}$$

Where w_t,w_s and w_f are the weights for text, speech, and facial expression modalities, respectively, and O_t,O_s,and O_f are their corresponding classifier outputs. These weights are optimized during the training phase to ensure balanced contributions, reflecting the unique strengths of each modality. This fusion strategy effectively captures interdependencies and correlations between modalities, resulting in improved classification accuracy.

The proposed methodology offers several advantages, including enhanced emotion recognition accuracy through the integration of multimodal data and advanced transformer models. By employing feature selection techniques, the framework minimizes computational costs while retaining the most critical information. The tri-modal fusion approach ensures robust emotion recognition, accounting for unique modality contributions and their interrelations. Furthermore, the use of advanced evaluation metrics underscores the system's effectiveness in capturing nuanced emotional states, making it highly suitable for real-world applications like classroom emotion monitoring and personalized learning environments.

4. IMPLEMENTATION & EVALUATION METRICS AND RESULTS

The implementation of the proposed multimodal emotion recognition framework can be achieved using Python with libraries such as TensorFlow or PyTorch for deep learning, and datasets like IEMOCAP, MELD, or CREMA-D for emotion classification tasks. The framework integrates transformer-based architectures (e.g., BERT) for text analysis, convolutional neural networks (CNNs) for speech and facial feature extraction, and a tri-modal fusion mechanism to combine the features from all modalities. The dataset is preprocessed to extract relevant features: facial expressions from images, speech signals using MFCCs, and textual transcripts. Separate neural networks are trained for each modality, and their outputs are fused using a fully connected layer to predict the final emotion class. Metrics such as Precision, Recall, F1-Score, and Balanced Accuracy are computed to evaluate model performance. The adaptive feature selection is implemented using techniques like the Rényi entropy-based Tversky index to optimize feature utilization. Results demonstrate the effectiveness of this framework in achieving state-of-the-art performance, making it suitable for applications in real-time emotion monitoring and online learning systems.

The suggested multimodal emotion recognition system has been evaluated using the following metrics: Accuracy, Precision, Recall, F1-score, and Balanced Accuracy. These metrics are delineated in relation to True Negative (TN), False Positive (FP), True Positive (TP), and False Negative (FN):

a) Accuracy provides the proportion of correctly classified samples and is expressed as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

b) Sensitivity (Recall) measures the proportion of actual positives correctly identified by the model and is calculated as:

$$\text{BaancedAccuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2} \tag{4}$$

c) Specificity measures the proportion of actual negatives correctly identified and is given as:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

d) Precision evaluates how many predicted positive samples are actually positive and is defined as:

$$F1 = 2\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

e) F1-Score represents the harmonic mean of Precision and Recall, providing a balance between the two metrics. It is expressed as:

$$\text{Sensitivity} = \frac{TP}{TP + FN}TN \quad (7)$$

For multiclass classification, the average Recall for each class is used to calculate Balanced Accuracy. This ensures that the metric reflects the model's ability to handle all emotion classes effectively, regardless of class imbalances. The advanced evaluation techniques, such as macro-averages and micro-averages, are employed to further assess the model's performance. Macro-averages compute the mean of metrics across all classes, treating each class equally, while micro-averages aggregate contributions from all classes to calculate a global metric. These metrics highlight the model's generalizability across diverse emotion classes and scenarios. The F1-score is particularly important for imbalanced datasets as it balances Precision and Recall, ensuring the model's robustness in predicting minority classes. By integrating a comprehensive set of metrics, the framework provides a detailed performance evaluation, ensuring reliability and scalability across various applications.

4.1 Single-Modality vs. Multimodal Emotion Classification

The proposed framework was evaluated against single-modality approaches (facial expressions, speech signals, and text) and compared with the tri-modal fusion methodology. Table 1 presents the classification performance for each modality and the multimodal approach across six emotion categories: Sadness, Joy, Love, Anger, Fear, and Surprise. Metrics such as Precision (PEE), Recall (REC), and F1-Score (F1) are used for evaluation.

Table 1: Classification Performance across Modalities

Emotion	**Face** (PEE, REC, F1)	**Text** (PEE, REC, F1)	**Speech** (PEE, REC, F1)	**Multimodal** (PEE, REC, F1)
Sadness	62.41, 15.60, 20.32	68.25, 18.40, 48.12	58.58, 25.36, 30.31	**78.32, 65.36, 62.36**
Joy	47.36, 25.25, 22.95	56.12, 62.23, 58.45	48.14, 62.12, 50.32	**72.35, 75.56, 65.56**
Love	32.69, 20.00, 10.88	38.12, 25.36, 20.00	35.00, 40.00, 22.00	**68.45, 49.26, 42.54**
Anger	45.36, 20.90, 30.32	53.28, 18.50, 45.12	53.14, 55.32, 50.35	**73.32, 71.84, 65.35**
Fear	35.00, 15.25, 25.98	50.13, 22.26, 41.10	45.00, 35.00, 25.70	**65.59, 25.11, 45.36**
Surprise	18.00, 0.00, 8.88	25.00, 12.00, 18.00	12.58, 0.00, 0.00	**58.45, 49.26, 39.54**
Average	**40.47, 16.33, 19.85**	**48.48, 26.96, 38.13**	**42.74, 36.02, 29.61**	**69.92, 59.90, 53.95**

The multimodal approach achieved the highest average Precision (69.92%), Recall (59.90%), and F1-Score (53.95%), showcasing the effectiveness of integrating multiple modalities for emotion classification.

4.2 Overall Performance Evaluation

To assess the overall performance of the proposed framework, metrics such as Accuracy, Balanced Accuracy, and F1-Score were evaluated. The results for all modalities and the tri-modal fusion are presented in Table 2.

Table 2: Comparative Analysis of Modalities

Modality	**Accuracy (%)**	**Balanced Accuracy (%)**	**F1-Score (%)**
Face	62.48	41.32	35.15

Text	65.24	45.23	44.36
Speech	64.35	42.65	40.26
Multimodal	**74.62**	**58.12**	**63.12**

The multimodal approach demonstrated a significant improvement in classification performance, achieving an Accuracy of 74.62% and an F1-Score of 63.12%.

4.3 Comparison with Existing Methods

Table 3 provides a comparison of the proposed framework with existing emotion classification methodologies. The proposed framework outperformed other approaches by incorporating advanced feature selection and leveraging transformer-based architectures.

The proposed framework showed an improvement of 5.15% in Accuracy and 3.00% in F1-Score compared to the best-performing existing approach.

The results validate the effectiveness of the Adaptive Deep Learning Framework for Emotion Classification. The tri-modal fusion of facial expressions, speech signals, and textual data ensures a comprehensive understanding of students' emotions, which single modalities fail to capture. The use of transformer architectures for text and CNNs for speech and facial features enhances feature extraction, while advanced metrics such as F1-Score and Balanced Accuracy underscore the framework's robustness.

The inclusion of adaptive feature selection methods, such as the Rényi entropy-based Tversky index, ensures optimal feature utilization, reducing computational overhead and improving classification accuracy. This makes the framework suitable for applications in monitoring student well-being, sentiment analysis, and personalized learning interventions in online platforms.

Table 3: Comparison with Existing Methods

Study	Modalities Used	Accuracy (%)	F1-Score (%)
Study [4]	Audio, Text	61.48	48.15
Study [22]	Audio, Text	63.28	49.82
Study [8]	Speech, Face, Text	68.85	58.12
Study [23]	Speech, Face, Text	69.47	60.15
Proposed Method	Speech, Face, Text	**74.62**	**63.12**

The comparison with existing methods highlights the advancements achieved by the proposed framework. It outperformed other state-of-the-art approaches, demonstrating its capability to handle complex, multimodal datasets and deliver superior classification performance.

5. CONSLUSION

This research introduces an Adaptive Deep Learning Framework for Emotion Classification in student data, utilising transformer topologies and sophisticated assessment criteria. In contrast to previous regression-centric methodologies, the framework shifts to classification problems, employing metrics like as accuracy, precision, recall, and F1-score for assessment. Effective preprocessing methods, such as Box-Cox transformation, standardise data and improve stability, while adaptive feature selection utilising the Rényi entropy-based Tversky index guarantees significant feature extraction. Transformer models such as BERT encapsulate profound contextual linkages inside textual data, whereas CNNs derive temporal and spatial information from auditory and facial emotions. The modalities are integrated through a tri-modal mechanism, resulting in enhanced emotion classification. The framework, validated using a dataset of six emotion categories—Sadness, Joy, Love, Anger, Fear, and Surprise—attained an accuracy of 74.62% and an F1-score of 63.12%, surpassing current methodologies. This

novel method shows promise in assessing student well-being, conducting sentiment analysis, and facilitating tailored interventions, hence enhancing results in online learning settings.

REFERENCES

[1] N. Aslam, F. Rustam, E. Lee, P. B. Washington, and I. Ashraf, "Sentiment analysis and emotion detection on cryptocurrency-related tweets using ensemble LSTM-GRU model," *IEEE Access*, vol. 10, pp. 39313-39324, 2022.

[2] M. Aslan, "CNN based efficient approach for emotion recognition," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7335-7346, 2022.

[3] M. Awais, M. Raza, N. Singh, K. Bashir, U. Manzoor, S. U. Islam, and J. J. Rodrigues, "LSTM-based emotion detection using physiological signals: IoT framework for healthcare and distance learning in COVID-19," *IEEE Access*, vol. 8, pp. 16863-16871, 2020.

[4] A. Ayub and A. R. Wagner, "EEC: learning to encode and regenerate images for continual learning," *arXiv preprint*, arXiv:2101.04904, 2021.

[5] X. Bai, O. Huerta, E. Unver, J. Allen, and J. E. Clayton, "A parametric product design framework for the development of mass customized head/face (eyewear) products," *Appl. Sci.*, vol. 11, no. 12, p. 5382, 2021.

[6] D. Bansal, R. Grover, N. Saini, and S. Saha, "GenSumm: a joint framework for multi-task Tweet classification and summarization using sentiment analysis and generative modelling," *IEEE Trans. Affect. Comput.*. doi: 10.1109/TAFFC.2021.3131516.

[7] M. F. Bashir, A. R. Javed, M. U. Arshad, T. R. Gadekallu, W. Shahzad, and M. O. Beg, "Context-aware emotion detection from low-resource Urdu language using deep neural network," *ACM*, vol. 22, pp. 1-30, 2023.

[8] A. Basile, M. Franco-Salvador, N. Pawar, S. Štajner, M. Chinea-Ríos, and Y. Benajiba, "Symantoresearch at SemEval-2019 task 3: combined neural models for emotion classification in human-chatbot conversations," in *Proc. 13th Int. Workshop Semantic Evaluation*, pp. 330-334, 2019.

[9] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE Access*, vol. 7, pp. 111866-111878, 2019.

[10] M. Bayer, M.-A. Kaufhold, B. Buchhold, M. Keller, J. Dallmeyer, and C. Reuter, "Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 1, pp. 135-150, 2023.

[11] L. A. Becker, H. Penagos, F. Flores, D. S. Manoach, M. A. Wilson, and C. Varela, "Eszopiclone and zolpidem produce opposite effects on hippocampal ripple density," *Front. Pharmacol.*, vol. 12, p. 792148, 2022.

[12] M. Sharmeen, A. Saleem, Y. Siddeeq, A. Mohammed, and R. M. Sadeeq, "Multimodal Emotion Recognition using Deep Learning," *Journal of Applied Science and Technology Trends*, vol. 2, pp. 52–58, 2021.

[13] B. Xie, M. Sidulova, and C. H. Park, "Robust Multimodal Emotion Recognition from Conversation with Transformer-Based Crossmodality Fusion," *Sensors*, vol. 21, p. 4913, 2021.

[14] N. Perveen, D. Roy, and K. M. Chalavadi, "Facial Expression Recognition in Videos Using Dynamic Kernels," *IEEE Transactions on Image Processing*, vol. 29, pp. 8316–8325, 2020.

[15] N. H. Ho, H. J. Yang, S. H. Kim, and G. Lee, "Multimodal approach of speech emotion recognition using multi-level multihead fusion attention based recurrent neural network," *IEEE Access*, vol. 8, pp. 61672–61686, 2020.

[16] S. R. Kandavalli, A. S. Edberk, D. K. Rajendran, and V. Rajagopal, "A Progressive Review on Wire Arc Additive Manufacturing: Mechanical Properties, Metallurgical and Defect Analysis," *Advances in Additive Manufacturing Processes*, vol. 1, p. 178, 2021. [Online]. Available: https://doi.org/10.2174/9789815036336121010014

[17] M. Wu, W. Su, L. Chen, W. Pedrycz, and K. Hirota, "Two-stage Fuzzy Fusion based-Convolution Neural Network for Dynamic Emotion Recognition," *IEEE Transactions on Affective Computing*, 2020.

[18] K. Mohan, A. Seal, O. Krejcar, and A. Yazidi, "Facial Expression Recognition using Local Gravitational Force Descriptor based Deep Convolution Neural Networks," *IEEE Transactions on Instrumentation and Measurement*, 2020.

[19] B. Priyalakshmi and P. Verma, "Navigational Tool for the Blind," *American Institute of Physics Conference Series*, vol. 2460, no. 1, 2022.

[20] L. Hu, W. Li, J. Yang, G. Fortino, and M. Chen, "A Sustainable Multimodal Multi-layer Emotion-aware Service at the Edge," *IEEE Transactions on Sustainable Computing*, 2019.

[21] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, pp. 344–353, 2020.

[22] L. Q. Bao, J. L. Qiu, H. Tang, W. L. Zheng, and B. L. Lu, "Investigating sex differences in classification of five emotions from EEG and eye movement signals," in *Proceedings of IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 6746–6749, 2019.

[23] D. W. Chen et al., "A feature extraction method based on differential entropy and linear discriminant analysis for emotion recognition," *Sensors*, vol. 19, pp. 1631–1647, 2019.

[24] Y. Huang, J. Yang, S. Liu, and J. Pan, "Combining facial expressions and electroencephalography to enhance emotion recognition," *Future Internet*, vol. 11, pp. 105–121, 2019.

[25] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-Gonzalez, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural networks for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 2169–3536, 2019.

[26] Y. Cimtay, E. Ekmekcioglu, and S. Caglar-Ozhan, "Cross-subject multimodal emotion recognition based on hybrid fusion," *IEEE Access*, vol. 8, pp. 168865–168878, 2020.

[27] J. Chen, Y. Lv, R. Xu, and C. Xu, "Automatic social signal analysis: Facial expression recognition using difference convolution neural network," *Journal of Parallel and Distributed Computing*, vol. 131, pp. 97–102, 2019.

[28] M. R. Mahmood, M. B. Abdulrazzaq, S. Zeebaree, A. K. Ibrahim, R. R. Zebari, and H. I. Dino, "Classification techniques' performance evaluation for facial expression recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 21, pp. 176–1184, 2021.

[29] K. B. Obaid, S. Zeebaree, and O. M. Ahmed, "Deep Learning Models Based on Image Classification: A Review," *International Journal of Science and Business*, vol. 4, pp. 75–81, 2020.

[30] V. A. K. Gorantla et al., "An intelligent optimization framework to predict the vulnerable range of tumor cells using Internet of things," in *Proceedings of IEEE 2nd International Conference on Industrial Electronics: Developments and Applications (ICIDeA)*, IEEE, 2023.

[31] Shynar Mussiraliyeva and Gulshat Baispay, “Leveraging Machine Learning Methods for Crime Analysis in Textual Data” International Journal of Advanced Computer Science and Applications(IJACSA), 15(4), 2024.

[32] Ch., R., Naresh, B., Prasanna, P. L., Chander, N., Goud, E. A., & Prasad, P. R. (2024). Exploring machine learning algorithms for robust cyber threat detection and classification: A comprehensive evaluation. *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)*.

[33] Ch., R., & Lakshmi, J. M. (2024). A decentralized approach for enhancing identity and access management through blockchain integration. In *2024 IEEE 6th International Conference on Cybernetics, Cognition and Machine Learning*.

[34] Ambika G N and Yeresime Suresh, “An Efficient Deep Learning with Optimization Algorithm for Emotion Recognition in Social Networks” International Journal of Advanced Computer Science and Applications(IJACSA), 14(8), 2023.