ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Hybrid Emotion Classification of MOOC Reviews Using the NRC Lexicon and a Multi-Channel Deep Learning Model

Raja Ouadad[1] and Hicham Mouncif[2]

ABSTRACT

Text-based emotion recognition has received extensive attention in applied computing research, but its effectiveness in online learning contexts remains limited. In this study, we introduce the TriFusion Attention Network, a hybrid deep learning model that classifies emotions in Massive Open Online Course (MOOC) reviews. Using the NRC Emotion Lexicon, we annotated learner reviews and designed the model to integrate multiple channels capturing both semantic and affective information. Its architecture combines Bidirectional Long Short-Term Memory (BiLSTM), Bidirectional Gated Recurrent Units (BiGRU), Convolutional Neural Networks (CNN), and attention mechanisms to model the complexity of learner feedback effectively. Experiments conducted on Coursera reviews demonstrate that the model effectively identifies both explicit and subtle emotional cues, achieving over 95% accuracy, F1-scores around 0.95, and AUC-ROC values approaching 0.99 on both balanced and imbalanced datasets. These results confirm that the proposed approach achieves superior performance compared to existing methods and facilitates improved learner engagement while offering richer analytical insights into their experiences.

Article information:

Keywords: Emotion Classification, MOOC Reviews, NRC Emotion Lexicon, Multi-Channel Deep Learning, Attention Mechanism

Article history:
Received: June 27, 2025
Revised: August 14, 2025
Accepted: September 18, 2025
Published: October 4, 2025 (Online)

DOI: 10.37936/ecti-cit.2025194.262786

1. INTRODUCTION

The delivery of education has changed with the rise of Massive Open Online Courses (MOOCs), which provide learners around the world with greater flexibility and access [1]. The growth in student enrolment has led to an increase in the feedback they provide, offering a helpful insight into their feelings and learning experiences [2]. Designing better learning experiences, raising engagement, and enhancing course content all depend on an understanding of these emotions. A critical problem in natural language processing (NLP) [3] is emotion analysis, which reveals the emotional tone concealed in text [4]. It helps teachers better understand how students feel about their educational experiences. Beyond education, emotion analysis plays a crucial role in fields such as business, public health, and politics, where awareness of emotions can enhance decisions and service quality [3].

Despite advances, accurately classifying emotions from text remains difficult, though. There are still challenges because human emotions are complex, people express themselves in various ways, and emotional categories are not evenly distributed [5]. Lexicon-based techniques and machine learning models are the two main categories of traditional approaches to emotion detection. Although lexicon-based approaches, including those that use the NRC Emotion Lexicon [6], are simpler to interpret, they frequently fail to capture contextual information Conversely, deep learning models excel at recognizing intricate semantic patterns; however, they usually demand substantial labelled data and function as "black boxes," which complicates the interpretation of their outputs [7].

To address these challenges, we propose the TriFusion Attention Network, a hybrid framework that combines the feature-learning power of deep learning with the transparency of lexicon-based approaches. To capture a wide range of semantic and emotional patterns in MOOC reviews, we first annotate the text data using the NRC Emotion Lexicon. The enriched text is then processed through a multi-channel deep learning architecture that integrates

[1,2]The authors are with the LIMATI Laboratory, Mathematics and Informatics Department, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco, Email: raja.ouadad@usms.ac.ma and h.mouncif@usms.ac.ma
[1]Corresponding author: raja.ouadad@usms.ac.ma

Bidirectional Long Short-Term Memory (BiLSTM) networks, Bidirectional Gated Recurrent Units (BiGRU), Convolutional Neural Network (CNN), and attention mechanisms. This approach aims to improve both the interpretability and accuracy of emotion classification. Experiments on Coursera learner reviews demonstrate that our model effectively detects both strong and subtle emotions, outperforming current methods. By distinguishing nuanced emotional states, the TriFusion Attention Network can help educational platforms better understand learners' experiences and provide more responsive support. The remainder of the paper presents related work, outlines our methodology, discusses experimental results, and highlights conclusions and future research directions.

2. LITERATURE REVIEW

The efficacy of natural language processing has improved considerably due to recent advances in machine learning and deep learning. (NLP) tasks like machine translation [8], entity resolution [8][9][10], sentiment analysis [11], question answering [12], and emotion classification [13]. Emotion classification based on student reviews from online learning platforms and MOOCs has made significant progress in recent years [14]. Still, several challenges remain [15]. One of the biggest hurdles is the complexity of human emotions: students express their experiences and opinions through a wide range of styles and subtle cues [7]. Informal language, shifting meanings depending on context, and the often-messy nature of online feedback all make it harder for NLP to interpret emotions [16] accurately.

Despite recent progress, emotion classification in MOOCs still faces challenges in extracting meaningful features and training effective models. A key research gap lies in developing models capable of handling the informal and diverse nature of student feedback. The following literature review explores the current approaches and innovations focused on enhancing feature representation, supervised learning, and deep learning techniques for emotion detection in educational settings.

Recent research on recognizing emotions in online learners has predominantly relied on text-based analysis, with a primary focus on applying sentiment analysis techniques to determine the polarity (positive or negative) of student reviews through various methodologies[17]. For instance, studies such as [18] [19] have employed lexicon-based approaches, while others [20] [21] have utilized machine learning (ML)-based methods. However, relatively few studies have gone beyond polarity detection to specifically identify and classify the nuanced emotional states underlying these positive and negative sentiments. Patrick *et al.* [22] categorized children's emotional responses during learning activities into several types: positive emotions (such as interest, happiness, and relaxation), boredom, pain, and anger. This classification, however, only distinguishes between positive and negative emotions, which may not fully address the complexities of emotional experiences in educational settings. To refine this approach, Pekrun *et al.* [23] expanded on the concept of academic emotions by incorporating the intensity of emotional activation. They proposed a four-dimensional framework, distinguishing between positive activating, positive deactivating, negative activating, and negative deactivating emotions, based on the level of student engagement and emotional arousal during the learning process.

To the best of our knowledge, the earliest initiative aimed at developing a text-based AI system for the automatically recognizing of online learners' emotions was presented by Tian *et al.* [24]. This study, published in 2014, conducted a series of experiments focused on detecting five negative emotional states—anger, anxiety, hopelessness, disgust, and sadness—expressed by learners. The researchers tested 16 classification algorithms using a text corpus from a chat system in an intelligent e-learning platform. Of all the models, the Random Forest CSC algorithm achieved the highest accuracy.

Oramas Bustillos *et al.* [25] created an opinion-mining module to predict online learners' sentiments and emotions. They evaluated six conventional machine learning classifiers: Bernoulli Naive Bayes, Multinomial Naive Bayes, Support Vector Machine (SVM), Linear SVC, Stochastic Gradient Descent (SGD), and K-Nearest Neighbors (KNN). They also tested a deep learning (DL) model that integrated Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures. The Linear SVC classifier performed best for emotion recognition with an accuracy of 60.00%, while the CNN-LSTM model had the highest sentiment classification accuracy at 88.26%.

Research on students' feelings and emotions in online learning has increased since 2020. To examine student feedback on online instruction, Kumar *et al.* [26] employed a lexicon-based sentiment analysis technique, namely the NRC Word-Emotion Association Lexicon. They divided the feedback into three categories: positive, negative, and neutral. They discovered that, although their opinions differed depending on the time of the classes, students generally had positive feelings about online learning.

SentiText, which includes opinions about programming languages, and EduSere, which focuses on emotions related to learning, such as frustration, boredom, excitement, and engagement, were two new corpora presented by Barrón Estrada *et al.* [27] in 2020. Three approaches were used to train the models: Deep Learning (DL), conventional Machine Learning (ML), and Evolutionary Algorithms. When paired with a BERT-based model, the EvoMSA algorithm produced the best results (84%) on EduSere and the highest accuracy (93%) on SentiText Similarly, Feng *et al.* [28] developed deep learning models, A-CNN and LSTMATT, to classify academic emotions based on comment features. Their models' 89% and 71% accuracy rates demonstrate how we can use deep learning to pick up on subtle emotional cues in learning environments.

Tzacheva *et al.* [29] expanded this research in 2021 by examining end-of-semester student responses to identify eight core emotions (joy, trust, anger, sadness, etc.) and to detect sentiment polarity as positive or negative. According to their study, Active Learning strategies—including Light Weight Teams and Flipped Classrooms—enhanced students' emotional experiences, boosting trust and positive sentiment and lowering negative emotions. A sentiment analysis model based on BiLSTM was created by Ji et al. [30] to assess emotional orientations in MOOC course reviews. By contrasting straightforward data replication methods with Easy Data Augmentation (EDA), they addressed data imbalance and suggested an improved version that uses synonym replacement. Their model's accuracy of over 90% shows how well BiLSTM works when paired with data augmentation techniques.

A deep neural network-based system for analyzing student feedback and emotional expressions was presented by Asghar *et al.* [31] in 2022. To categorize feelings like love, happiness, rage, and contempt, they employed a Bi-Directional LSTM (BiLSTM) model. According to their findings, the BiLSTM model outperformed previous research and conventional machine learning classifiers.

A machine learning system automatically recognizing emotions from online learner feedback was proposed by Kasliwal *et al.* in 2023 [32]. Their method was trained using a sizable dataset of 80,940 interactions, comments, and reviews, employed both BERT-encoded features and linguistic features (such as word count and sentence length) to train classification models. Using the Random Forest algorithm, the system reached its peak accuracy of 94%.

In 2025, Kamakshamma *et al.* [33] proposed a deep learning framework to categorize student emotions from social media data, integrating transformer models such as BERT with advanced preprocessing techniques. By classifying emotions into six categories—sadness, joy, love, anger, fear, and surprise—their model significantly enhanced classification performance and offered valuable insights into students' well-being and the effectiveness of online learning.

As emphasized in this section, previous research on emotion recognition from online learner text has largely concentrated on assigning emotional labels to student feedback through deep learning or lexicon-based approaches. Notwithstanding recent progress, students continue to exhibit diverse emotions, and correctly classifying particularly the minority or less frequent emotions remains a major challenge. We suggest a hybrid model that combines a Multi-Channel Deep Learning architecture with the NRC Emotion Lexicon to get around these restrictions. This method combines the deep contextual understanding provided by neural networks with the benefits of lexicon-based emotional feature extraction. To show the model's efficacy in both increasing the detection of minority class emotions and overall classification accuracy, we test it on both balanced and imbalanced datasets. The results of our experiments demonstrate that our approach offers a robust and fair framework for emotion classification in online learning environments.

3. MATERIAL AND METHODS

This section describes the approach used to categorize learner reviews according to their emotions. The data source and its properties are introduced first, after which we describe the preprocessing procedures applied to prepare the text data. Next, we outline the techniques used for dataset balancing and the initial emotion labelling process, which relies on the NRC lexicon. We then describe the experimental setup for training and evaluation, along with the architecture and hyperparameters of the proposed multi-channel deep learning model. Finally, we present the evaluation metrics used to assess the model's performance.

3.1 Data Source and Description

In this study, we used a dataset of 107,018 English-language reviews from learners enrolled in Coursera courses, one of the leading online education platforms. The original dataset contained three attributes: ID (a unique identifier for each review), Review (the textual feedback provided by learners), and Label (the course rating). For this analysis, we retained only the ID and Review columns, focusing on the textual content to examine learners' sentiments and emotions. This dataset forms the basis for the emotion classification assignments undertaken in this study. This dataset forms the basis for the emotion classification tasks conducted in this study and is publicly available on Kaggle: `https://www.kaggle.com/datasets/septa97/100k-courseras-course-reviews-dataset`

3.2 Data Preprocessing

The preprocessing phase was crucial for ensuring the dataset remained clean, consistent, and ready for modelling [34]. Several transformations were applied to the raw textual feedback to improve data quality and extract meaningful features, ultimately enhancing model performance. The preprocessing steps involved removing duplicate entries, cleaning the text

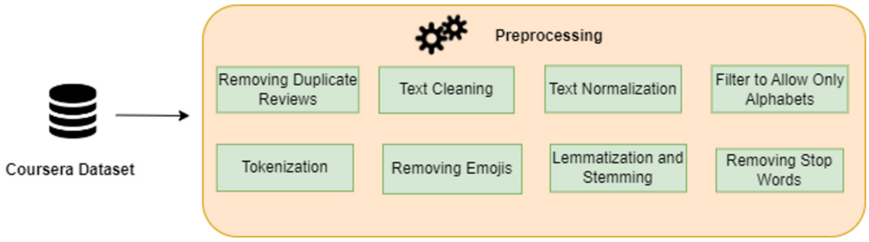


Fig.1: *Data Preprocessing Pipeline.*

(by stripping out HTML tags, hyperlinks, memorable characters, and punctuation), normalizing the text (through lowercasing and removing Unicode and non-alphabetic characters), filtering out stop words, and performing both lemmatization and stemming. We also removed emojis and performed tokenization to prepare the text for further analysis. A summary of the preprocessing workflow is shown in Figure 1.

3.3 Emotion Classification using NRC Lexicon-Based Approach

The NRC (National Research Council) Emotion Lexicon is a widely adopted resource for lexicon-based emotion classification. It maps words to eight fundamental emotions—Anger, Fear, Anticipation, Trust, Surprise, Sadness, Joy, and Disgust—as well as to two sentiment categories: positive and negative. In the NRC Lexicon, each word is assigned to a binary value (1 or 0), indicating the presence or absence of an association with a particular emotion or sentiment. This structure enables the classification of text based on the presence and frequency of emotionally charged words [35]. The emotion score for a given review in relation to a specific category c (e.g., ‘joy’ or ‘fear’) is computed as follows:

$$S_c = \sum_{w \in R} Score(w, c) \quad (1)$$

Where:

—S_c is the raw emotion score for category c.
—R is the set of words in the review.
—$Score(w, c)$ is defined as:

$$Score(w, c) = \begin{cases} 1 & \textit{if word } w \textit{ is associated with category } c \\ 0 & \textit{otherwise} \end{cases} \quad (2)$$

To better illustrate the application of the NRC lexicon to learners’ reviews, Table 1 presents sample reviews along with their assigned emotion categories. Each review carries a label corresponding to one of the eight basic emotions.

While the NRC Lexicon provides a valuable foundation for emotion annotation, it is inherently context independent. This limitation can lead to misclassifications, particularly when the meaning of a word shifts depending on its surrounding context. For example, a word such as “*surprised*” might express excitement in one review but disappointment in another, yet the lexicon assigns it a fixed label. Consequently, the initial emotion labels generated in this study may contain noise or bias. To address this, the subsequent deep learning model is specifically captures semantic and affective information. these static lexicon associations, capturing semantic and contextual patterns in learner reviews. In this way, the

Table 1: *Example Reviews with Preprocessing and Emotion Classification.*

Original Review	Cleaned Review	Emotion Classification
“I loved it very much. Easy to understand, visual and enjoyable explanations.”	“love much easi understand visual enjoy explan”	Joy
“The subjects are very well covered. Great explanation. Very recommendable”	“subject well cover great explan recommend”	Trust
“Just finished the first week and I have hope that I will see some improvments at the end of it, it seems to be very promising.”	“finish first week hope see improv end seem promis”	Anticipation
“Great course with clear instruction and a final peer-review project with clear expectations and explanations.”	“great cours clear instruct final peerreview project clear expect explan”	Positive
“Bad teaching quality.”	“bad teach quality”	Anger
“A bit difficult in terms or using very complicated words....”	“bit difficult term use complic word”	Fear
“Boring and can’t make sense of the presentation / language and idiom use.”	“bore cant make sens present languag idiom use”	Negative
“It was okay.”	“okay”	No emotion

model refines and corrects the initial annotations, leading to more robust and context-aware emotion classification.

3.4 Dataset Balancing

After assigning emotions to the reviews using the NRC lexicon, we observed a strong imbalance in the dataset. The distribution revealed that some emotions—such as positive (48,675 samples), no emotion (19,826 samples), and anticipation (19,601 samples)—were highly overrepresented, while others—such as Anger (480 samples), fear (901 samples), and negative (2,730 samples)—were significantly underrepresented. To address this imbalance, a combination of techniques was employed: we first applied undersampling to reduce the number of samples in the majority classes. Then we applied SMOTE (Synthetic Minority Over-sampling Technique) to augment the minority classes synthetically. This dual strategy ensured that each emotion class contained exactly 16,311 samples, as shown in Figure 2. Maintaining both the original imbalanced dataset and the newly balanced version allowed us to evaluate the model's performance under both realistic and balanced conditions, providing a comprehensive assessment of its robustness across all emotional categories.

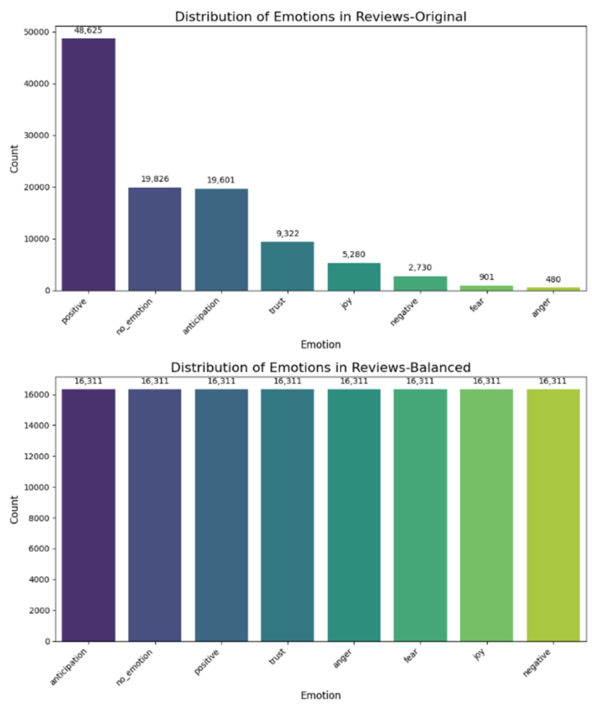


Fig. 2: *Emotion Class Distribution: Original vs. Balanced Dataset.*

3.5 Architecture and Hyperparameters of Our Multi-Channel Deep Learning Model

The architecture of the proposed TriFusion Attention Network model is depicted in Figure 3, focusing on building a robust emotion classification system for student feedback. We use a preprocessed sentence as input to the model. Subsequently, we generate word vectors using word embedding techniques. The model integrates several key components, including Bidirectional LSTM (BiLSTM) for capturing contextual information, an Attention mechanism for focusing on the most informative parts of the sequence, Bidirectional GRU (BiGRU) for modelling sequential dependencies, and Convolutional Neural Networks (CNN) for extracting local and position-invariant features. In the following paragraphs. We provide a detailed explanation of each phase and component.

To ensure optimal training and model performance, we carefully selected specific hyperparameters for each component. Table 2 summarizes these hyperparameters.

Table 2: *Model Hyperparameters and Training Settings.*

Hyperparameter	Value
Vocabulary Size	10,000
Max Sequence Length	100
Embedding Dimension	Varied (pre-trained optional)
Dropout (Embedding & Layers)	0.3
Bidirectional LSTM Units	128
Bidirectional GRU Units	64
CNN Filters / Kernel Size	64 / 5
Activation Functions	LeakyReLU, Softmax
Regularization (L2)	0.001
Optimizer	Nadam (LR = 5e-5)
Loss Function	Categorical Crossentropy
Batch Size	64
Epochs	10
Learning Rate Decay	ReduceLROnPlateau

The proposed combination is motivated by the complementary strengths of each channel. The BiLSTM captures long-range bidirectional dependencies, which are crucial for modeling discourse-level context in MOOC reviews. The BiGRU introduces a more efficient recurrent pathway with distinct gating dynamics, allowing the model to capture medium-range dependencies while stabilizing learning. The CNN block effectively extracts local features, such as n-gram–like sentiment expressions or intensifiers, which often indicate emotional polarity in short review segments. Finally, the attention mechanism acts as an integrator, selectively emphasizing the most salient features from both recurrent and convolutional outputs, thereby improving interpretability and robustness. together, this interaction enables the architecture to effectively capture both local and global emotional cues, addressing the noisy and context-dependent nature of learner-generated reviews.

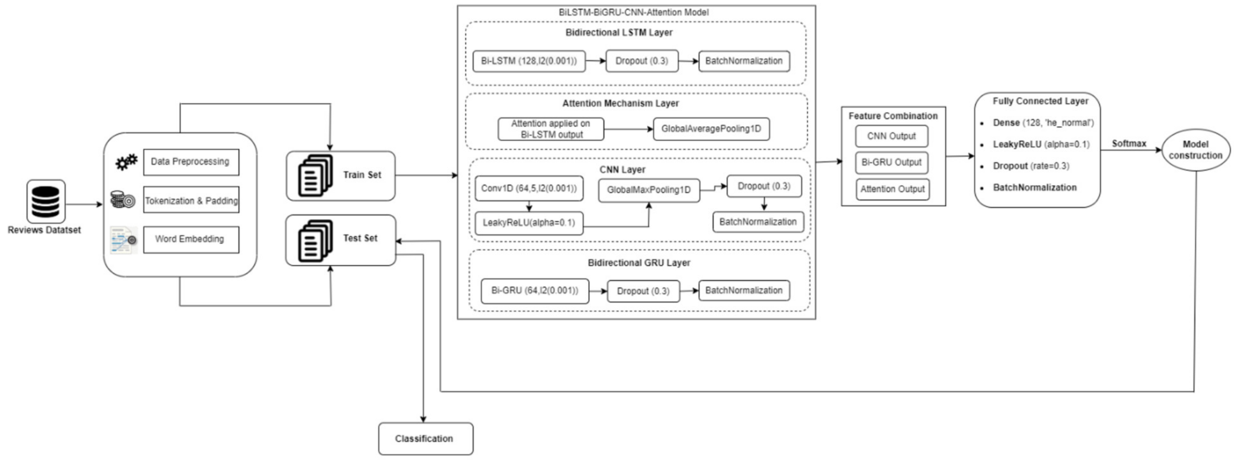


Fig.3: *The Architecture of the Hybrid Proposed Model.*

The hybrid model is represented mathematically as follows:

Denote the input sequence of tokens as $x = (x_1, x_2, ..., x_T)$, where T is the maximum sequence length.

- **Embedding Layer**

Each token x_t maps to a dense vector representation:

$$e_t = E(x_t) \tag{3}$$

Where $E \in \boldsymbol{R}^(v \times d)$ is the embedding matrix, V is the vocabulary size, and d is the embedding dimension. After applying a spatial dropout, the embedding undergoes regularization:

$$\tilde{e}_t = \text{SpatialDropout1D}(e_t) \tag{4}$$

- **Bidirectional LSTM Layer**

The sequence $\tilde{e}_t$ then passes through a Bidirectional LSTM.

At each time step, the LSTM computes:

$$\begin{aligned}
i_t &= \sigma(W_i \tilde{e}_t + U_i h_{t-1} + b_i) \\
f_t &= \sigma(W_f \tilde{e}_t + U_f h_{t-1} + b_f) \\
o_t &= \sigma(W_o \tilde{e}_t + U_o h_{t-1} + b_o) \\
g_t &= \tanh(W_g \tilde{e}_t + U_g h_{t-1} + b_g) \\
c_t &= f_t \odot c_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot \tanh(c_t)
\end{aligned} \tag{5}$$

The output from both directions is combined through concatenation:

$$h_t^{\text{biLSTM}} = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{6}$$

- **Attention Mechanism**

We apply an attention mechanism to the BiLSTM output:

$$\begin{aligned}
&Attention(h^{\text{biLSTM}}) \\
&= \text{softmax}\left(\frac{h^{\text{biLSTM}}(h^{\text{biLSTM}})^T}{\sqrt{d}}\right) h^{\text{biLSTM}}
\end{aligned} \tag{7}$$

We then apply Global Average Pooling to obtain a fixed-size vector:

$$\begin{aligned}
&h^{att} \\
&= \text{GlobalAveragePooling}(\text{Attention}(h^{\text{biLSTM}}))
\end{aligned} \tag{8}$$

- **CNN Block**

The BiLSTM output undergoes a convolution operation:

$$h^{cnn} = \text{LeakyReLU}(W_c * h^{\text{biLSTM}} + b_c) \tag{9}$$

Where W_c and b_c are the convolution kernel and bias, and * denotes 1D convolution.
Then, the model applies Global Max Pooling:

$$h^{cnn} = \text{GlobalMaxPooling1D}(h^{cnn}) \tag{10}$$

- **Bidirectional GRU Block**

In parallel, the BiLSTM outputs pass through a Bidirectional GRU:

$$\begin{aligned}
z_t &= \sigma(W_z[h_{t-1}, \tilde{e}_t] + b_z) \\
r_t &= \sigma(W_r[h_{t-1}, \tilde{e}_t] + b_r) \\
\tilde{h}_t &= \tanh(W_h[r_t \odot h_{t-1}, \tilde{e}_t] + b_h) \\
h_t &= (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t
\end{aligned} \tag{11}$$

The model concatenates the outputs from both directions:

$$h_t^{\text{biGRU}} = \left[\overrightarrow{h_t}; \overleftarrow{h_t}\right] \tag{12}$$

• **Feature Concatenation**

We concatenate the three feature vectors:

$$h^{concat} = [h^{cnn}; h^{\text{biGRU}}; h^{att}] \tag{13}$$

• **Fully Connected Layers**

The combined feature vector passes through a dense layer:

$$h^{dense} = \text{LeakyReLU}(W_d h^{concat} + b_d) \tag{14}$$

After dropout and batch normalization, the output layer produces the final probability distribution:

$$\tilde{y} = \text{Softmax}(W_o h^{dense} + b_o) \tag{15}$$

Where $\tilde{y} \in \boldsymbol{R}^C$ and C is the number of classes.

3.6 Experimental Setup

To evaluate the effectiveness of the proposed TriFusion Attention Network, we partitioned the dataset into 60% for training, 20% for validation, and 20% for testing. We trained the model on the training subset, optimized the hyperparameters using the validation data, and used the test subset for the final performance assessment. We explored two experimental setups: first, testing the model on the original imbalanced dataset, and second, on a balanced version of the dataset. We implemented all experiments in Python 3.9.11 using TensorFlow 2.18.0 (Keras 3.6.0) and scikit-learn 1.1.3, and we conducted them on a personal computer equipped with an Intel(R) Core (TM) i5-7300U processor, 16 GB of RAM, and a processing speed of 2.71 GHz.

3.7 Evaluation Metrics

A diverse array of evaluation metrics, including the Confusion Matrix, Matthews Correlation Coefficient (MCC), AUC-ROC, F1-score, Precision, and Recall, were used to evaluate the performance of the suggested model. Precision is the ratio of true positives to all predicted positives, whereas accuracy is the percentage of correctly predicted cases. The F1-score balances the trade-off between precision and recall by providing the harmonic mean of both variables. Recall shows the percentage of real positive instances that the model correctly identifies. The AUC-ROC assesses the model's capacity to discriminate between positive and negative classes across various thresholds. The MCC offers a more impartial assessment by considering both true and false negatives as well as true and false positives. Lastly, the Confusion Matrix supplies a thorough analysis of the classification results.

4. RESULTS AND DISCUSSIONS

This section presents the assessment of the suggested hybrid model for classifying emotions in Coursera learner reviews. To evaluate the impact of data preprocessing on classification results, we first compared model performance before and after balancing the dataset. To evaluate the model's capacity to generalize across the training and validation sets, we produced overfitting graphs. We then demonstrated the benefits of the hybrid model by comparing its performance with conventional deep learning techniques. Lastly, a comparison with previous research showed that the suggested model outperforms the most advanced emotion classification techniques in terms of results.

4.1 Performance Evaluation

The proposed TriFusion Attention Network model, which integrates BiLSTM, BiGRU, CNN, and attention mechanisms, was evaluated for the task of emotion classification on learner reviews from Coursera. As summarized in Table 3, the model achieved an overall accuracy of 95.41% on the balanced dataset and 95.56% on the imbalanced dataset, with AUC-ROC scores of 0.9979 and 0.9881, respectively, demonstrating strong generalization across different data distributions. The Matthews Correlation Coefficient (MCC) further confirmed the model's reliability with scores of 0.9477 (balanced) and 0.9379 (imbalanced). A more detailed class-wise analysis (Table 4) showed that for the balanced dataset, the model consistently achieved high precision, recall, and F1-scores across all emotion classes, with powerful performance for major categories such as Joy, No Emotion, and Positive. In contrast, in the imbalanced dataset, while major in contrast, in the imbalanced dataset, the major classes maintained high predictive quality, whereas minority classes such as Anger and Fear showed a noticeable drop in recall. This effect can be attributed to their underrepresentation in the dataset compared to emotions like Joy or Trust. With fewer examples to learn from, the model struggles to capture the full linguistic diversity of these emotions. For instance, the model often misclassifies short reviews containing subtle signals of frustration (e.g., "the platform sometimes wastes my time") as neutral or negative, instead of explicitly recognizing them as Anger. The overall weighted average metrics nevertheless remained robust, highlighting the hybrid model's effectiveness in both balanced and real-world imbalanced settings.

Figure 4 illustrates the confusion matrices for the proposed TriFusion Attention Network model evaluated on both the balanced and imbalanced datasets. For the balanced dataset (left), the model demonstrates a high degree of classification accuracy across all emotion classes, with most samples correctly classified along the diagonal. Minor misclassifications

Table 3: *Overall Performance Metrics for Balanced vs. Imbalanced Data.*

Data Type	Accuracy	Precision	Recall	F1-Score	AUC-ROC	MCC
Balanced	0.9541	0.9544	0.9541	0.9538	0.9979	0.9477
Imbalanced	0.9556	0.9546	0.9556	0.9540	0.9881	0.9379

Table 4: *Comparison of Emotion Classification Metrics for Balanced and Imbalanced Datasets.*

Data Type	Balanced				Imbalanced			
Metric / Class	Precision	Recall	F1-Score	Support	Precision	Recall	F1-Score	Support
Anger	0.97	0.98	0.97	3291	0.83	0.19	0.31	100
Anticipation	0.95	0.91	0.93	3305	0.94	0.96	0.95	3991
Fear	0.96	0.99	0.97	3188	0.81	0.63	0.71	172
Joy	0.96	0.98	0.97	3253	0.94	0.97	0.96	1093
Negative	0.97	0.89	0.93	3198	0.85	0.75	0.80	534
No Emotion	0.96	0.99	0.97	3320	0.97	0.99	0.98	3927
Positive	0.96	0.94	0.95	3291	0.97	0.97	0.97	9666
Trust	0.91	0.96	0.94	3252	0.90	0.90	0.90	1868
Accuracy	-	-	0.95	26098	-	-	0.96	21351
Macro avg	0.95	0.95	0.95	26098	0.90	0.80	0.82	21351
Weighted avg	0.95	0.95	0.95	26098	0.95	0.96	0.95	21351

occur predominantly between semantically similar emotions, such as Anticipation and Trust, or Negative and Anger. For the imbalanced dataset, the model maintains strong performance on dominant classes—such as Positive, No Emotion, and Anticipation—which have larger sample sizes. However, the model shows a slight performance drop for under-represented classes such as Anger and Fear, mainly reflected in lower recall values, consistent with the class-wise metrics reported in Table 4. Overall, the confusion matrices confirm the robustness and reliability of the proposed hybrid model across varying data distributions.

These patterns highlight two primary challenges faced by the model. First, errors linked to class imbalance affect minority emotions such as Anger and Fear, which lack sufficient training examples to capture their full linguistic diversity. Second, the model experiences semantic confusion between closely related emotions, such as Anticipation and Trust. Both emotions frequently co-occur in forward-looking reviews; for instance, the sentence "*I look forward to completing more courses from this platform*" may be labeled as Anticipation but can also be interpreted as an expression of Trust. Such linguistic ambiguities partly explain the confusion observed in the matrices. Overall, the results confirm the robustness and reliability of the proposed hybrid model across varying data distributions despite these inherent challenges.

In practical terms, distinguishing between closely related emotions such as Anger and Frustration has important implications for designing responsive educational support systems. For instance, expressions of Anger in learner reviews may signal dissatisfac-

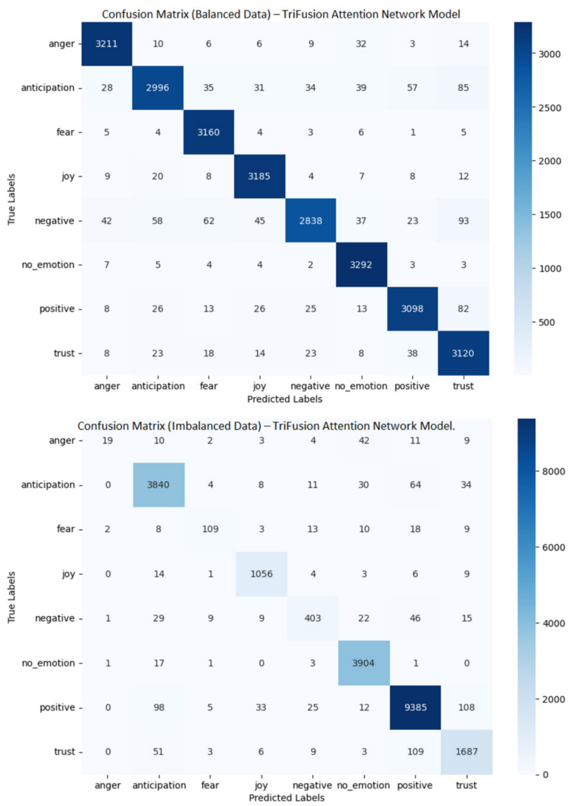


Fig.4: *Confusion matrices for the Proposed Model.*

tion with broader aspects of the learning experience, such as course organization, teaching quality, or perceived fairness. In such cases, institutional-level interventions—such as revising course content, providing instructor training, or collecting targeted feedback—may be appropriate. In contrast, Frustration often reflects challenges with specific learning tasks or technical difficulties, where more immediate and individualized responses, such as adaptive hints, step-by-step guidance, or peer support, would be more effective. By enabling platforms to differentiate these emotional states, our model provides actionable insights that can inform the development of adaptive interventions tailored to both the emotional and pedagogical needs of learners.

4.2 Interpretation and Analysis of Learning Curves

In this experiment, we visualized the training and validation accuracy and loss to comprehensively assess the performance of our hybrid model on both balanced and imbalanced datasets. The following figures present the results. In Figure 5 (balanced dataset), the training and validation accuracy steadily increased over epochs, reaching above 90% by the end of training. Simultaneously, both the training and validation loss show a continuous decrease, stabilizing at relatively low values. The close convergence of the training and validation curves, with minimal gaps between them, suggests that the model generalizes well without significant overfitting. This behaviour indicates that the hybrid architecture, combining Bidirectional LSTM, Attention, CNN, and Bidirectional GRU, effectively encapsulates the fundamental structures in the balanced data. In Figure 6 (imbalanced dataset), Figure 6 (imbalanced dataset) shows a similar trend: both training and validation accuracy progressively improve, while the corresponding losses decrease. However, is worth noting that the validation loss tends to be lower than the training loss from the early stages, and the validation accuracy slightly exceeds the training accuracy. This behaviour often occurs when the dataset imbalance leads the model to predict dominant classes more confidently during validation, resulting in lower validation loss. Despite this, the overall consistency between training and validation curves indicates good learning behaviour and robustness of the model, even when dealing with class imbalance.

Overall, the training and validation graphs demonstrate that the proposed hybrid model achieves stable and high performance across both balanced and imbalanced datasets, highlighting the effectiveness of combining recurrent, convolutional, and attention mechanisms within a single architecture.

4.3 Comparison with Simple Deep-learning Layers

To assess the effectiveness of the proposed hybrid model, we conducted a comparative analysis against several simple deep learning architectures, including LSTM, RNN, CNN-Text Classifier, BiLSTM, BiGRU, and CNN combined with LSTM. As shown in Tables 5 and 6, the proposed TriFusion Attention

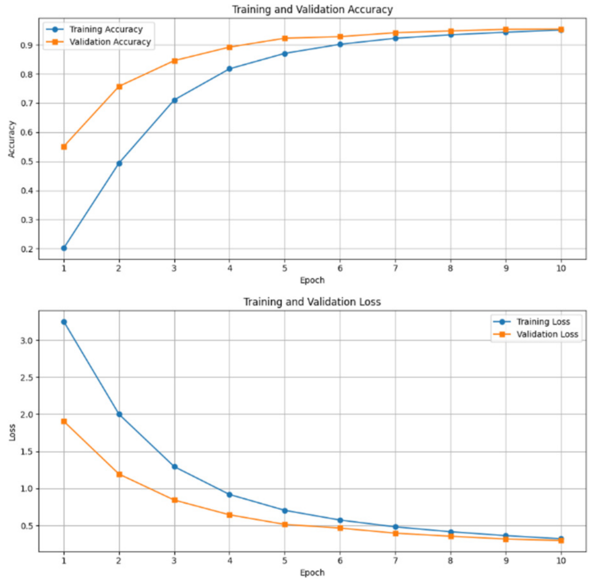


Fig.5: *Model Learning Curves Using the Balanced Dataset.*

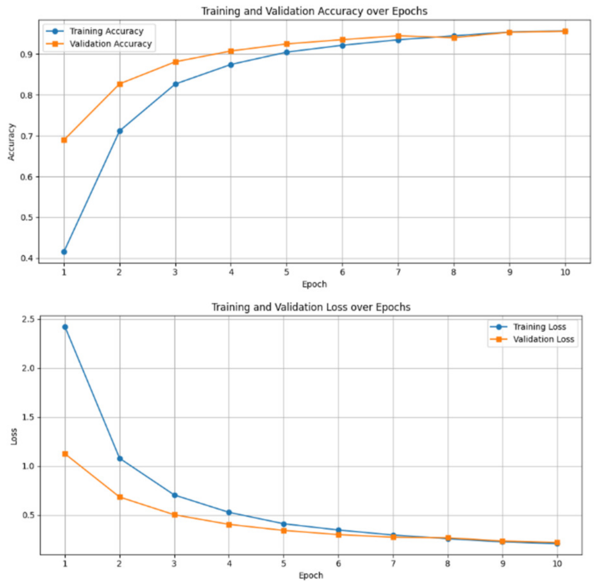


Fig.6: *Model Learning Curves Using the Imbalanced Dataset.*

Network consistently outperformed all other models on both balanced and imbalanced datasets across all evaluation metrics. On the balanced dataset, the proposed model obtained an accuracy of 95.41%, an F1-score of 95.38%, and an AUC-ROC of 0.9979, outperforming the best baseline (BiGRU) by approximately 1.5% in accuracy. Similarly, on the imbalanced dataset, the proposed model obtained an accuracy of 95.56%, an F1-score of 95.40%, and an AUC-ROC of 0.9881, again surpassing the next best model (BiGRU) by over 1.3% in accuracy. These results demonstrate the advantages of combining multiple recurrent, convolutional, and attention mechanisms for more robust emotion classification of learner reviews.

Table 5: *Comparative Performance of Deep Learning Algorithms and the Proposed Model on the Balanced Dataset.*

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	MCC
LSTM	0.6494	0.6317	0.6494	0.6240	0.9211	0.6037
RNN	0.8438	0.8430	0.8438	0.8426	0.9745	0.8217
CNN-Text Classifier	0.8917	0.8918	0.8917	0.8906	0.9909	0.8766
CNN + LSTM	0.9333	0.9338	0.9333	0.9329	0.9955	0.9240
BiLSTM	0.9351	0.9352	0.9351	0.9348	0.9936	0.9259
BiGRU	0.9392	0.9390	0.9392	0.9390	0.9949	0.9305
Our Model	**0.9541**	**0.9544**	**0.9541**	**0.9538**	**0.9979**	**0.9477**

Table 6: *Comparative Performance of Deep Learning Algorithms and the Proposed Model on the Imbalanced Dataset.*

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC	MCC
LSTM	0.5202	0.3486	0.5202	0.4127	0.5883	0.3004
RNN	0.8475	0.8217	0.8475	0.8339	0.9122	0.7865
CNN-Text Classifier	0.9087	0.8953	0.9087	0.8982	0.9586	0.8710
BiLSTM	0.9331	0.9189	0.9331	0.9254	0.9694	0.9059
CNN + LSTM	0.9357	0.9304	0.9357	0.9306	0.9777	0.9095
BiGRU	0.9423	0.9385	0.9423	0.9371	0.9810	0.9191
Our Model	**0.9556**	**0.9546**	**0.9556**	**0.9540**	**0.9881**	**0.9379**

As illustrated in Figures 7 and 8, the confusion matrices show that simple deep learning models—such as LSTM, RNN, and CNN-based architectures—cannot accurately classify all emotion classes, even on the balanced dataset, with noticeable confusion among similar categories. This difficulty becomes even more pronounced on the imbalanced dataset, where the model frequently misclassifies minority courses. In contrast, the proposed TriFusion Attention Network achieves a more consistent and balanced classification across all classes in both scenarios, demonstrating its superior ability to capture complex emotional patterns and effectively handle class imbalance.

4.4 Evaluative Comparison to Existing Literature

To further evaluate the efficiency of the proposed TriFusion Attention Network, we conducted a comparative study with recent works on emotion classification of learner reviews in online learning environments. Table 7 summarizes various studies focused on emotion detection in MOOCs, educational platforms, or student-generated content, indicating the datasets used, the number of emotion categories con-

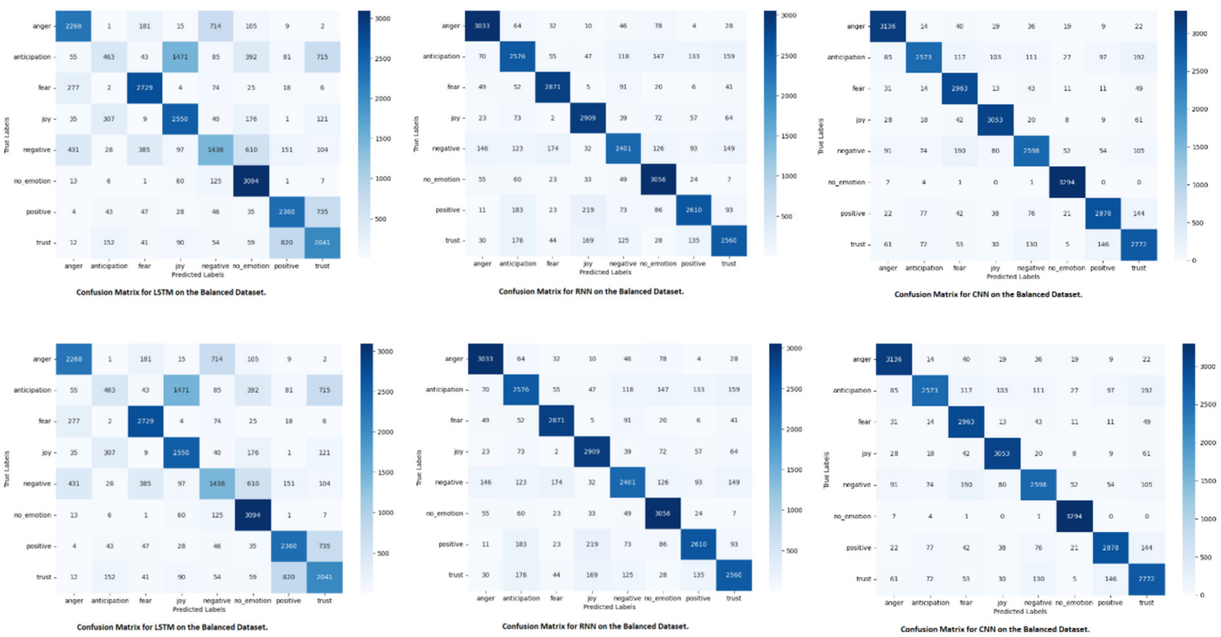


Fig.7: *Confusion Matrices of Various Models on Balanced Dataset.*

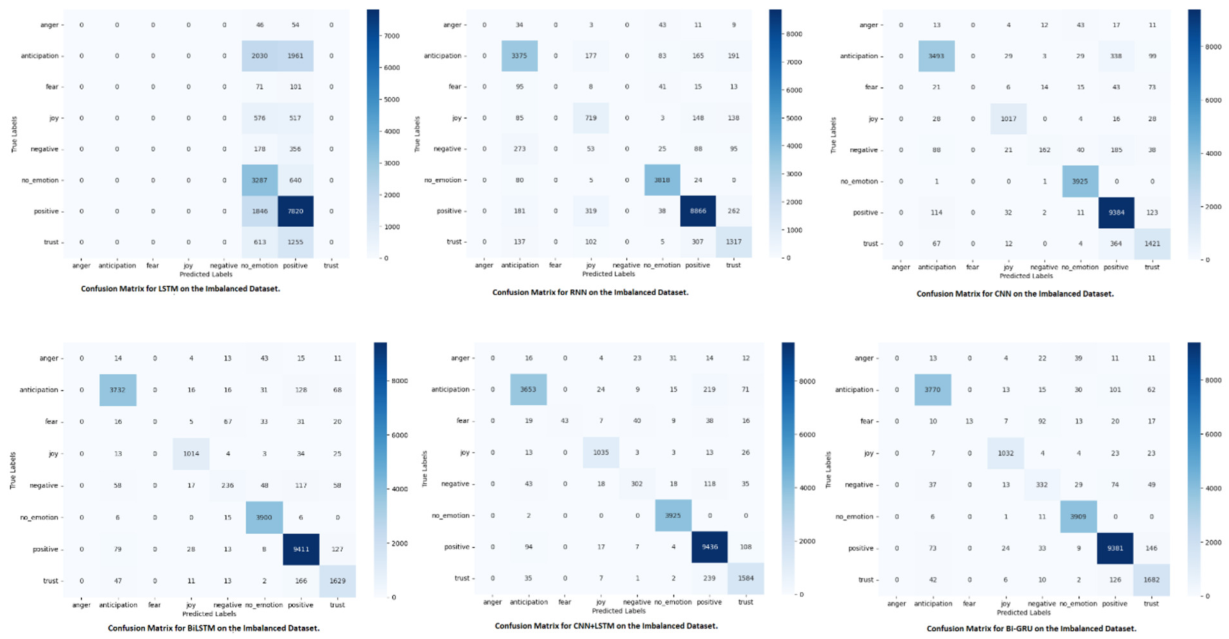


Fig.8: *Confusion Matrices of Various Models on Imbalanced Dataset.*

Table 7: *Comparative Evaluation of Emotion Classification Models on Educational and Review Datasets.*

Method	Dataset Description	Number of Emotions	Accuracy
Our Model	**107,018 reviews from Coursera**	**8(Joy,Trust,Anticipation,Positive , Anger,Fear,Negative,No emotion)**	**95%**
CNN+LSTM[25]	Corpus of opinions related to computer programming (7,777 records)	5 (frustrated, bored, excited, engaged, neutral)	Polarity: 88.26%, Emotion Recognition: 60.0%
EvoMSA + BERT [27]	SentiText and EduSere datasets	2 (positive/negative) for SentiText; 4 (frustration, boredom, excitement, engagement) for EduSere	93% (SentiText), 84% (EduSere)
A-CNN [28]	More than 200,000 online student comments (Tencent Classroom, Chinese University MOOC, and blogs)	9 (boredom, shame, enjoyment, hope, disappointment, Anger, anxiety, joy, relaxation)	89%
BiLSTM + Simple Copy Method [30]	6694 MOOC course reviews (“Fundamentals of College Computer Application” + other universities)	2 (Positive, Negative)	$> 90\%$
Random Forest (RF) [32]	E-Learners Academic Reviews (ELAR) dataset (80,940 reviews)	5 (Excitement, Happy, Satisfied, Not Satisfied, Frustration)	94%
Adaptive Deep Learning Framework (BERT + CNN, Tri-modal Fusion) [33]	Student social media dataset (text, auditory and facial emotion data)	6 (Sadness, Joy, Love, Anger, Fear, Surprise)	74.62%

sidered, and their reported accuracy. Although the datasets and the number of classified emotions differ, this comparison provides a meaningful benchmark. The results show that the proposed model achieves superior or comparable performance, demonstrating its robustness and strong potential for emotion analysis in MOOCs and online education contexts.

5. CONCLUSIONS

In this study, we introduce a hybrid deep learning framework that integrates BiLSTM, CNN, Bi-GRU, and an attention mechanism for emotion classification of learner reviews in MOOCs. The model demonstrated robust performance on both balanced and imbalanced Coursera review data, surpassing existing state-of-the-art techniques across multiple evaluation metrics, including accuracy, Precision, Recall, F1-Score, AUC-ROC, and MCC. By effectively capturing sequential, local, and contextual features, our model addresses key challenges in text classification, offering a robust solution for various NLP applications. Looking ahead, we plan to integrate advanced techniques such as pre-trained contextual embeddings and transfer learning to scale the model for big-data environments. We will also test the model's generalizability across a broader range of educational datasets and MOOC platforms to ensure its effectiveness in diverse linguistic and academic contexts.

AUTHOR CONTRIBUTIONS

Raja Ouadad conducted the programming, performed the experiments, and wrote the manuscript. Hicham Mouncif supervised the project, validated the results, and provided critical feedback on the manuscript. Both authors contributed to the visualization, manuscript revision, and approved the final version of the paper.

References

[1] K. F. Hew and W. S. Cheung, "Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges," *Educational Research Review*, vol. 12, pp. 45–58, Jun. 2014.

[2] S. Zheng, M. B. Rosson, P. C. Shih and J. M. Carroll, "Understanding Student Motivation, Behaviors and Perceptions in MOOCs," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, in CSCW '15*, New York, NY, USA: Association for Computing Machinery, pp. 1882–1895, Feb. 2015.

[3] S. S. I. ElShafie, S. S. Ismail, K. A. H. Bahnasy and M. M. Aref, "Convolutional Neural Network Multi-Emotion Classifiers," *Jordanian Journal of Computers and Information Technology (JJCIT)*, vol. 5, no. 2, pp. 97–97, Jul. 2019.

[4] S. M. Mohammad, "9 - Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text," in *Emotion Measurement*, H. L. Meiselman, Ed., Woodhead Publishing, pp. 201–237, 2016.

[5] H. Alhuzali and S. Ananiadou, "SpanEmo: Casting Multi-label Emotion Classification as Span-prediction," in *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, P. Merlo, J. Tiedemann, and R. Tsarfaty, Eds., Online: Association for Computational Linguistics, pp. 1573–1584, Apr. 2021.

[6] S. M. Mohammad and P. D. Turney, "Crowdsourcing A Word–Emotion Association Lexicon," Accessed: Apr. 28, 2025. [Online]. Available: `https://onlinelibrary.wiley.com/doi/10.1111/j.1467-8640.2012.00460.x`

[7] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, A. Moschitti, B. Pang, and W. Daelemans, Eds., Doha, Qatar: Association for Computational Linguistics, pp. 1746–1751, Oct. 2014.

[8] A. Vaswani *et al.*, "Attention is all you need," in *Proceedings of the 31st International Conference on Neural Information Processing Systems, in NIPS'17*, Red Hook, NY, USA: Curran Associates Inc., pp. 6000–6010, Dec. 2017,.

[9] J. Mourad, T. Hiba, R. Yassir and H. Imad, "ER-ABQS: entity resolution based on active machine learning and balancing query strategy," *Journal of Intelligent Information Systems*, vol. 62, no. 5, pp. 1347–1373, Mar. 2024.

[10] M. Jabrane, H. Tabbaa, A. Hadri, and I. Hafidi, "Enhancing Entity Resolution with a hybrid Active Machine Learning framework: Strategies for optimal learning in sparse datasets," *Information Systems*, vol. 125, , Nov. 2024.

[11] R. Ouadad and H. Mouncif, "Sentiment Analysis of Students Feedback in Online Courses Using Supervised, Ensemble, and Transfer Learning Methods," in *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security, in NISS '24*, New York, NY, USA: Association for Computing Machinery, pp. 1–7, Aug. 2024.

[12] P. Rajpurkar, J. Zhang, K. Lopyrev and P. Liang, "SQuAD: 100,000+ Questions for Machine Comprehension of Text," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, J. Su, K. Duh, and X. Carreras, Eds., Austin, Texas: Association for Computational Linguistics, pp. 2383–2392, Nov. 2016.

[13] S. Peng and L. Cao, "Emotion Classification in Textual Conversations Using Deep Broad Learning," in *Textual Emotion Classification Using Deep Broad Learning*, Springer, Cham, pp. 119–133, 2024.

[14] N. Altrabsheh, M. Cocea and S. Fallahkhair, "Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback," *2014 IEEE 26th International Conference on Tools with Artificial Intelligence*, Limassol, Cyprus, pp. 419-423, 2014.

[15] S. Poria, E. Cambria, and A. Gelbukh, "Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, L. Màrquez, C.

Callison-Burch, and J. Su, Eds., Lisbon, Portugal: Association for Computational Linguistics, pp. 2539–2544, Sep. 2015.

[16] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion Detection in Text: A Review," *arXiv preprint arXiv:1806.00674*, 2018

[17] H. Makhoukhi and S. Roubi, "Multi-Label Emotion Classification of Online Learners' Reviews Using Machine Learning," in *Proceedings of the 2024 5th International Conference on Education Development and Studies, in ICEDS '24*, New York, NY, USA: Association for Computing Machinery, pp. 59–64, Jul. 2024.

[18] K. Z. Aung and N. N. Myo, "Sentiment analysis of students' comment using lexicon based approach," *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, Wuhan, China, pp. 149-154, 2017.

[19] R. Faizi, "Using Sentiment Analysis to Explore Student Feedback: A Lexical Approach," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 18, no. 09, Art. no. 09, May 2023.

[20] Z. Kastrati, B. Arifaj, A. Lubishtani, F. Gashi and E. Nishliu, "Aspect-Based Opinion Mining of Students' Reviews on Online Courses," in *Proceedings of the 2020 6th International Conference on Computing and Artificial Intelligence, in ICCAI '20*, New York, NY, USA: Association for Computing Machinery, pp. 510–514, Aug. 2020.

[21] R. Sadigov, E. Yıldırım, B. Kocaçınar, F. Patlar Akbulut and C. Catal, "Deep learning-based user experience evaluation in distance learning," *Cluster Computing* , vol. 27, pp. 443–455, 2024.

[22] B. Patrick, E. Skinner and J. Connell, "What Motivates Children's Behavior and Emotion? Joint Effects of Perceived Control and Autonomy in the Academic Domain," *Journal of Personality and Social Psycholply*, vol. 65, pp. 781–91, Oct. 1993.

[23] R. Pekrun, T. Goetz, A. Frenzel, P. Barchfeld and R. Perry, "Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ)," *Contemporary Educational Psychology*, vol. 36, pp. 36–48, Jan. 2011.

[24] F. Tian *et al.*, "Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems," *Knowledge-Based Systems*, vol. 55, pp. 148–164, Jan. 2014.

[25] R. Oramas Bustillos, R. Zatarain Cabada, M. L. Barrón Estrada and Y. Hernández Pérez, "Opinion mining and emotion recognition in an intelligent learning environment," *Computer Applications in Engineering Education*, vol. 27, no. 1, pp. 90–101, Jan. 2019.

[26] P. Kumar, "Exploring the Students Feelings and Emotion Towards Online Teaching: Sentimental Analysis Approach," *IFIP Advances in Information and Communication Technology*, vol. 617, pp. 137–146, Dec. 2020.

[27] M. L. Barrón Estrada, R. Zatarain Cabada, R. Oramas Bustillos and M. Graff, "Opinion mining and emotion recognition applied to learning environments," *Expert Systems with Applications*, vol. 150, p. 113265, Jul. 2020.

[28] X. Feng, Y. Wei, X. Pan, L. Qiu and Y. Ma, "Academic Emotion Classification and Recognition Method for Large-scale Online Learning Environment—Based on A-CNN and LSTM-ATT Deep Learning Pipeline Method," *International Journal of Environmental Research and Public Health*, vol. 17, no. 6, p. 1941, Mar. 2020.

[29] A. Tzacheva and A. Easwaran, "Emotion Detection and Opinion Mining from Student Comments for Teaching Innovation Assessment," *International Journal of Education (IJE)*, vol. 09, no. 02, pp. 21–32, Jun. 2021.

[30] S. Ji and T. Fangbi, "Emotion Analysis Model of MOOC Course Review Based on BiLSTM," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 08, Art. no. 08, Apr. 2021.

[31] M. Z. Asghar *et al.*, "An Efficient Classification of Emotions in Students' Feedback using Deep Neural Network," in *2022 13th International Conference on Information and Communication Systems (ICICS)*, , pp. 186–191, Jun. 2022.

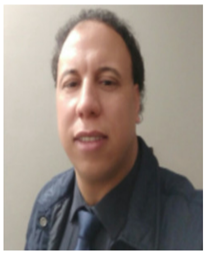
[32] P. S. Kasliwal, R. Gunjan and V. Shete, "Computation of E-learners Textual Emotion to Enhance learning Experience," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 10s, Art. no. 10s, Aug. 2023.

[33] V. Kamakshamma, "Adaptive Deep Learning Framework for Emotion Classification in Student Data Using Transformer Models and Advanced Evaluation Metrics," *Journal of Information Systems Engineering and Management*, vol. 10, pp. 276–286, Mar. 2025.

[34] M. Hosseinzadeh *et al.*, "Data cleansing mechanisms and approaches for big data analytics: a systematic study," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, pp. 1–13, Nov. 2021.

[35] S. M. Mohammad and P. D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, Aug. 2013.

Raja Ouadad received the M.Sc. degree in Computer Science from the Faculty of Sciences and Techniques, Beni Mellal, Morocco, in 2022. Currently, she is a Ph.D. candidate in the Department of Mathematics and Informatics at the Polydisciplinary Faculty, Sultan Moulay Slimane University, Morocco. Her research interests include sentiment analysis, artificial intelligence, machine learning, and deep learning algorithms. She is the corresponding author of this article and can be contacted at email: ouadadraja2@gmail.com



Hicham Mouncif is a Full Professor and Ph.D. Supervisor in the Department of Computer Sciences, Polydisciplinary Faculty of Beni Mellal, University Sultan Moulay Slimane. He has published numerous academic papers in distinguished journals based on teaching and research experience. As the Coordinator of the Computer Systems Engineering Master's Program, his research interests include educational technologies, machine learning, transportation networking, and routing protocols. He can be contacted at email: h.mouncif@usms.ma.