MTL-BERTNet: A Multi-Task Learning Framework for Aspect and Sentiment Analysis in MOOC Reviews

[1]Raja Ouadad, and Hicham Mouncif

***Abstract*—In the context of large-scale online learning environments, analyzing student feedback is crucial for improving course content and learner engagement. This paper proposes MTL-BERTNet, a novel multi-task learning architecture that jointly performs aspect category classification and sentiment polarity detection from MOOC reviews. The model leverages contextual embeddings from a pre-trained BERT encoder and integrates a convolutional multi-head attention mechanism to capture subtle semantic nuances and inter-task dependencies. To further enhance shared representation learning across tasks, an inter-task matching layer (IML) is introduced. Experiments conducted on an imbalanced MOOC review dataset demonstrate strong performance, with macro F1-scores of 0.90 for aspect classification and 0.93 for sentiment prediction. These results highlight the effectiveness of jointly modeling aspects and sentiment, offering practical insights for improving course design, instructional quality, and learner satisfaction in MOOC platforms.**

***Index Terms*—aspect-based sentiment analysis, multi-task learning, MOOC reviews, BERT, deep learning, attention mechanism.**

I. INTRODUCTION

MAssive Open Online Courses (MOOCs) generate abundant user-generated reviews, offering insights into learners' experiences and expectations. Analyzing this feedback helps educators and course designers improve course quality, teaching methods, and engagement [1]. While traditional sentiment analysis identifies positive, negative, or neutral sentiments [2], it often misses aspect-specific details. Aspect-based sentiment analysis (ABSA) addresses this by linking sentiments to particular aspects—such as content, teaching style, or platform usability—providing a more detailed and actionable understanding [3].

Learners' feedback on online courses often addresses specific aspects—such as content quality, instructional clarity, course design, or overall structure—paired with a sentiment (positive, negative, or neutral) toward each aspect [4]. For instance, "The content was comprehensive and up to date" expresses positive sentiment toward the content, whereas "The structure was confusing and lacked coherence" conveys negative sentiment about course structure. This illustrates the close connection between aspect identification and sentiment polarity, which is crucial for fine-grained sentiment analysis in MOOCs [5][6].

Most prior work in sentiment analysis focuses on polarity detection alone [7], often neglecting the interdependence between sentiments and the aspects they reference. In learner reviews, sentiments are typically anchored to specific course dimensions, such as content or instructor performance. For example, "The instructor explained concepts poorly" links negative sentiment directly to the instructor aspect. Modeling such dependencies allows one task to inform the other, motivating multi-task learning approaches that exploit the correlation between aspect recognition and sentiment classification to improve performance in both tasks.

To address this issue, we suggest a unified multi-task learning (MTL) system that manages sentiment analysis and aspect categorization at the same time. The model can benefit from the connections between these tasks by training them together, which frequently produces better overall outcomes.

As a kind of implicit regularization that helps avoid overfitting, MTL's strength lies in its capacity to allow related tasks to guide one another [8]. By finding complementary patterns across tasks, this shared learning process helps the model generalize more successfully [9][10]. The advantages of MTL have been demonstrated in a variety of NLP applications, including sentiment and emotion classification, syntactic parsing, and machine translation [11]. In our work, because aspect evaluation and sentiment polarity are fundamentally coupled, MTL provides a viable strategy to exploit this interaction and improve both tasks simultaneously [12].

Unlike traditional multi-task learning (MTL) methods that divide the model into shared and task-specific parts, our framework introduces a new component called the Inter-Task Matching Layer (IML). Rather than relying solely on shared representations, the proposed IML explicitly models bidirectional interactions between aspect category classification and sentiment polarity detection through cross-task information exchange. This layer is designed to better capture and utilize the interplay between both tasks, enabling the model to learn richer and more discriminative task-specific representations [13]. Furthermore, to boost overall

[1] Raja Ouadad, Hicham Mouncif are affiated with laboratory of Innovation and Mathematics, Applications and Information Technology, Polydisciplinary Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco.
(E-mails: ouadadraja2@gmail.com, h.mouncif@usms.ma) Manuscript received October 23, 2025; revised XX,XX,XXXX

DOI: 10.36244/ICJ.2026.2.8

performance, our approach incorporates transfer learning by using the pre-trained BERT model, which is well-regarded for its ability to understand complex semantic and syntactic relationships within text. BERT was selected as a widely adopted and computationally efficient backbone, allowing us to focus on evaluating the effectiveness of the proposed multi-task architecture rather than the impact of different pre-trained language models. In addition, a hybrid CNN–Multi-Head Attention (CNN–MHA) module is employed to jointly capture local contextual patterns and long-range semantic dependencies within learner reviews.

In summary, the main contributions of this study are as follows:

1. We propose a novel BERT-enhanced Multi-Task Learning framework (MTL-BERTNet) designed to jointly perform sentiment polarity detection and aspect category classification on MOOC reviews.
2. A hybrid architecture utilizing convolutional operations and multi-head attention is employed to simultaneously model local patterns and broader context within learner reviews.
3. An Inter-Task Matching Layer is employed to capture dependencies and interactions between the sentiment analysis and aspect classification tasks.
4. We utilize the powerful contextual representation capabilities of pre-trained BERT to initialize the embedding layer, enabling the model to benefit from rich linguistic features.
5. Extensive experiments conducted on a real-world imbalanced dataset demonstrate that our proposed MTL-BERTNet model significantly outperforms several strong baselines in both sentiment and aspect classification tasks.

The remainder of this paper is structured as follows: Section II provides an overview of related work. Section III details the proposed multi-task learning framework. Section IV describes the experimental setup and presents the results and analysis. Finally, Section V presents the conclusion of the study and highlights potential avenues for future investigation.

II. RELATED WORK

Aspect-based Sentiment Analysis (ABSA), a subfield of Sentiment Analysis (SA), aims to identify sentiments associated with specific aspects of a target entity [14][19]. In MOOCs, ABSA enables the analysis of learners' opinions regarding cours content, instructional quality, course design, and platform usability, providing actionable insights for educational improvement [5] [14]. Existing approaches can be broadly categorized into traditional and deep learning methods, transformer-based models, and multi-task learning frameworks.

A. Traditional and Deep Learning Methods

While sentiment analysis has been widely applied across domains, research on educational ABSA remains relatively limited. Early studies relied on traditional machine learning techniques such as Naive Bayes and SVM for analyzing student feedback, achieving accuracies of 72.8% and 81%, respectively [15]. Subsequent work employed Stanford CoreNLP, OpenNLP, SentiWordNet, and ontology-based approaches for aspect extraction and sentiment classification in educational settings [16], [17].

Deep learning methods further improved performance by learning richer semantic representations. However, most of these approaches rely on sequential architectures (e.g., LSTMs) that may struggle to capture long-range dependencies and contextual relationships as effectively as transformer-based models. Sindhu et al. [18] proposed a two-layer LSTM architecture achieving 91% accuracy in aspect extraction and 93% in sentiment classification. Similarly, [19] reported F1-scores of 80% and 81% for aspect and sentiment prediction, respectively. Kastrati et al. [20] proposed a hybrid Word2Vec–CNN model for Coursera reviews, achieving about 80–82% macro-F1 and outperforming classical baselines by 6–10% in F1-score. A later study [21] used weak supervision and reported an F1-score of approximately 93.3%, showing strong performance while reducing annotation cost.

B. BERT and Transformer-Based Models

More recent transformer architectures have continued to improve contextual representation learning; however, most studies focus on encoder improvements rather than explicitly modeling the interaction between aspect and sentiment prediction tasks. The introduction of BERT [13] and its variants, including SpanBERT [22] and RoBERTa [23], significantly advanced ABSA by providing contextualized word representations. In educational datasets, transformer-based models consistently outperformed conventional machine learning approaches. Fine-tuned IndoBERT achieved F1-scores of 0.897 and 0.882 for aspect extraction and sentiment classification, respectively [24], while multilingual transformer architectures demonstrated promising results for Arabic educational reviews [25].

Further studies confirmed the effectiveness of transformer-based approaches in educational feedback analysis. IndoBERT achieved 92.9% accuracy on university reviews [26], EduBERT reached 89.78% accuracy in sentiment and urgency classification [27], and hierarchical BERT architectures improved instructor evaluation tasks [28][29]. Hybrid approaches combining transformers and traditional classifiers also reported strong results, with RoBERTa achieving 95.5% sentiment classification accuracy [30].

Despite their effectiveness, most transformer-based approaches model aspect identification and sentiment classification independently, limiting their ability to explicitly exploit the strong dependency between these two tasks.

C. Multi-Task Learning for ABSA

Multi-Task Learning (MTL) has emerged as an effective paradigm for ABSA because it enables related tasks to share representations and improve generalization [8]. Attention-based MTL architectures have shown superior performance compared with single-task models by capturing inter-task dependencies more effectively [31], while task-specific attention mechanisms and cross-stitch networks further improve performance by enabling adaptive feature sharing across tasks, with reported gains on standard multi-task benchmarks such as text and vision classification datasets [32].

He et al. [33] proposed an MTL framework that jointly addresses Aspect Term Extraction (ATE), Aspect Polarity Classification (APC), and aspect–sentiment pair recognition. Zhao et al. [34] incorporated ATE as an auxiliary task and leveraged multi-head attention to strengthen dependency modeling. Other studies combined BERT with knowledge graphs and graph neural networks to enrich semantic representations and improve ABSA performance [35].

More recently, SP-MTL-BERT employed selective paraphrasing with nuance control to augment educational review datasets, reporting substantial gains in aspect recall and sentiment precision [36]. Although SP-MTL-BERT and the proposed framework both address aspect and sentiment analysis in educational reviews within a multi-task learning setting, their design objectives differ. SP-MTL-BERT primarily focuses on data augmentation and shared transformer representations, whereas MTL-BERTNet emphasizes explicit cross-task knowledge exchange through task-specific projections and a co-attention-based Inter-Task Matching Layer (IML). While existing MTL approaches generally rely on shared parameters, auxiliary-task supervision, or attention mechanisms to facilitate knowledge transfer, the dependency between aspect categories and sentiment polarity is often modeled only indirectly. To address this limitation, MTL-BERTNet incorporates the IML together with a hybrid CNN–MHA encoder, enabling both explicit inter-task interaction and the joint modeling of local contextual patterns and long-range semantic dependencies. As a result, the proposed framework provides a more direct mechanism for exploiting the mutual relationship between aspect categories and sentiment polarity.

III. PROPOSED METHOD

We present a multi-task learning framework for joint classification of aspect categories and sentiment polarity in MOOC reviews. Joint modeling captures interdependencies between sentiment and aspect information, improving overall performance in aspect-based sentiment analysis.

A. Formal Task Definition

Let $R = [x_1, x_2, \dots, x_N]$ represent a learner review composed of N tokens, where each token x_i denotes a word in the input sequence. Each review in the dataset is annotated with two types of labels:

• A sentiment polarity label, denoted by Y_{sent}, belonging to the set Y_{sent}, defined as:

Y_{sent} = {Positive (P), Neutral (NEU), Negative (N)}

• An aspect category label, denoted by Y_{aspect}, belonging to the set Y_{aspect}, defined as:

Y_{aspect} = {Content (C), Structure (S), Instructor (I), Course Design (SC), General Impression (ISC)}

The problem is formulated as a multi-task learning (MTL) framework where both tasks are jointly optimized, enabling the model to exploit semantic relationships between aspect and sentiment through shared contextual representations.

B. Model Architecture Overview

We propose MTL-BERTNet (Multi-Task Learning BERT Network), illustrated in Figure 1. The architecture consists of five main components:

- BERT Embedding Layer: Generates contextualized token embeddings $B_{emb} \in \mathbb{R}^{N \times d}$, where $d = 768$.
- Shared Representation Learning Module: A hybrid CNN–Multi-Head Attention (CNN–MHA) module extracts local and global contextual features, producing the shared representation C.
- Task-Specific Representation Learning Module: Two Fully Connected (FC) layers transform the shared representation into the sentiment representation $S_{rep}$and aspect representation A_{rep}.
- Inter-Task Matching Layer (IML): Performs bidirectional information exchange through Aspect-to-Sentiment (A2S) and Sentiment-to-Aspect (S2A) attention, enabling interaction between task-specific representations.
- Task-Specific Classification Layer: Uses Softmax classifiers to predict aspect categories and sentiment polarity.

For clarity, B_{emb}denotes BERT embeddings, $C$the shared feature representation, FC the Fully Connected layer, $S_{rep}$the sentiment representation, and $A_{rep}$the aspect representation.

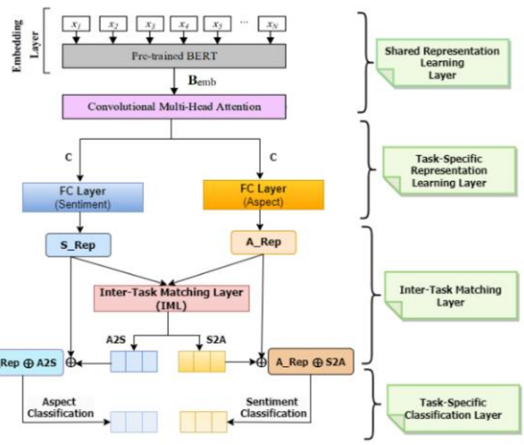

Fig.1. Overall architecture of the proposed Multi-Task Learning BERT Network (MTL-BERTNet), showing the flow from BERT-based embeddings through shared representation learning, task-specific encoding, and inter-task matching via cross-attention, followed by final classification layers for aspect and sentiment prediction.

C. Layer-Wise Framework and Learning Mechanism

The first stage of the proposed architecture aims to transform each review into a dense vector representation suitable for downstream learning.

Given an input sequence $R = [x_1, x_2, \dots, x_N]$ of N tokens, a frozen pre-trained BERT model [13] generates contextualized embeddings $B_{\text{emb}} \in \mathbb{R}^{N \times d}$ (with $d = 768$). Unlike static word representations, BERT produces context-dependent embeddings that capture lexical ambiguity and complex syntactic relations. Using BERT as a frozen feature extractor reduces computational cost, mitigates overfitting, and preserves pre-trained linguistic knowledge. This choice is motivated by the relatively limited size of MOOC review datasets and the need to ensure stable contextual representations while reducing training complexity, making it well-suited for our multi-task learning setting.

From this contextual embedding matrix, a hybrid CNN–Multi-Head Attention (CNN–MHA) module extracts high-level semantic features (Figure 2). First, a 1D convolution with kernel size k and ReLU activation captures local dependencies (short phrases, n-grams):

$$X = \text{ReLU}(\text{Conv1D}(B_{\text{emb}}) + b), X \in \mathbb{R}^{N \times d}$$

where $b \in \mathbb{R}^d$ is a learnable bias. To complement these local features, a multi-head self-attention mechanism models global dependencies. Self-attention-based models [37] are effective at this task due to their parallelizable structure, and the multi-head attention mechanism [38] enables learning multiple semantic relationships simultaneously. Let H denote the number of attention heads, with head dimension $d_h = d/H$. For each head $h \in \{1, \dots, H\}$, the local representations X are projected into queries, keys, and values:

$$Q_h = XW_Q^h, K_h = XW_K^h, V_h = XW_V^h$$

The attention operation for head h is:

$$\text{Attention}_h(Q_h, K_h, V_h) = \text{softmax}\left(\frac{Q_h K_h^{\top}}{\sqrt{d_h}}\right) V_h \in \mathbb{R}^{N \times d_h}$$

where $\frac{Q_h K_h^{\top}}{\sqrt{d_h}}$ computes scaled pairwise similarities between tokens. The outputs of all heads are concatenated to form a unified representation:

$$C = \text{Concat}(C^1, C^2, \dots, C^H) \in \mathbb{R}^{N \times d},$$
$$C^h = \text{Attention}_h(Q_h, K_h, V_h)$$

Matrix C integrates both local and global features, serving as the shared representation for both tasks.

To specialize this shared knowledge, two linear projections map C into lower-dimensional task-specific spaces ($d' < d$):

$$S_{\text{rep}} = CW_s, A_{\text{rep}} = CW_a$$

with $W_s, W_a \in \mathbb{R}^{d \times d'}$, yielding $S_{\text{rep}}, A_{\text{rep}} \in \mathbb{R}^{N \times d'}$.

A co-attention mechanism then enables bidirectional information exchange between tasks. A normalized similarity matrix is first computed:

$$M = \frac{S_{\text{rep}} A_{\text{rep}}^{\top}}{\sqrt{d'}} \in \mathbb{R}^{N \times N}$$

Row-wise softmax normalization produces cross-task enriched representations:

$$\text{S2A} = \text{softmax}_{\text{row}}(M^{\top}) S_{\text{rep}} \in \mathbb{R}^{N \times d'},$$
$$\text{A2S} = \text{softmax}_{\text{row}}(M) A_{\text{rep}} \in \mathbb{R}^{N \times d'}$$

Here, S2A enriches aspect features with sentiment information, while A2S enriches sentiment features with aspect information.

For sentence-level prediction, global average pooling over the sequence length N aggregates token-level representations:

$$F_{\text{sent}} = \text{GlobalAvgPooling}(S_{\text{rep}} \oplus \text{A2S}) \in \mathbb{R}^{2d'},$$
$$F_{\text{aspect}} = \text{GlobalAvgPooling}(A_{\text{rep}} \oplus \text{S2A}) \in \mathbb{R}^{2d'}$$

where $\oplus$ denotes feature-wise concatenation. Final probability distributions are obtained via softmax classifiers:

$$P_{\text{sent}} = \text{softmax}(F_{\text{sent}} W_{\text{sent}} + b_{\text{sent}}) \in \mathbb{R}^{|Y^{\text{sent}}|},$$
$$P_{\text{aspect}} = \text{softmax}(F_{\text{aspect}} W_{\text{aspect}} + b_{\text{aspect}}) \in \mathbb{R}^{|Y^{\text{aspect}}|}$$

with $|Y^{\text{sent}}| = 3$ and $|Y^{\text{aspect}}| = 5$.

The model is trained via joint optimization, minimizing the sum of categorical cross-entropy losses:

$$\mathcal{L} = \mathcal{L}_{\text{sent}} + \mathcal{L}_{\text{aspect}},$$

$$\mathcal{L}_{task} = -\frac{1}{B} \sum_{n=1}^{B} \sum_{c=1}^{C} y_{n,c} \log(\hat{y}_{n,c})$$

where B is the mini-batch size, C the number of classes, $y_{n,c}$ the one-hot ground truth, and $\hat{y}_{n,c}$ the predicted probability. Equal weighting of both losses ensures balanced learning and effective knowledge transfer between sentiment and aspect classification.

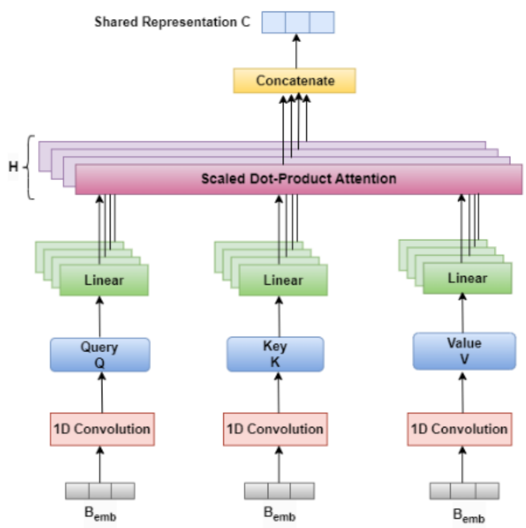

Fig. 2. Architecture of the Convolutional Multi-Head Attention Module (CNN-MHA).

IV . EXPERIMENTS

In this section, we present the experimental setup, dataset details, evaluation metrics, implementation settings, and a comprehensive analysis of the results obtained by the proposed MTL-BERTNet model. The primary objective of this experimental phase is to assess the effectiveness of our multi-task learning approach in jointly performing aspect category classification and sentiment polarity detection on MOOC reviews.

A. Dataset Description

The dataset used in this study consists of 24,255 student reviews collected from the Coursera platform through a web-scraping process. The data spans 15 distinct MOOCs covering multiple subject domains and was collected over a defined time period.

These 24,255 reviews include both authentic and synthetic samples. The dataset is composed of 21,940 (90.5%) authentic reviews and 2,315 (9.5%) synthetic reviews generated using a GPT-based data augmentation strategy.

The authentic reviews were originally collected and manually labeled by Kastrati et al. (2020) [20], and the dataset was obtained directly from the authors upon request, as it is not publicly available online. ensuring high-quality annotations across 15 courses. All reviews are written in English and represent realistic learner feedback from Massive Open Online Courses (MOOCs).

To enhance the representation of less frequent aspect–polarity combinations, synthetic reviews were generated using a GPT-based data augmentation approach. These samples were created using a structured instruction-based prompting strategy designed to produce realistic MOOC-style feedback while maintaining consistency with course-related aspects and sentiment labels. The goal of this augmentation was to increase data diversity and improve model generalization rather than to strictly balance the dataset.

Each instance in the dataset comprises three features: Comment, which contains the textual content of the review; Aspect, referring to one of five predefined course-related categories; and Polarity, indicating the sentiment expressed toward the aspect.

To address class imbalance, class weighting is applied during training, where weights are computed inversely proportional to class frequencies in the training set.

Reviews vary in length from 1 to 554 words, with an average of 23.26 words, a median of 13 words, and a standard deviation of 29.11, reflecting high variability in review length. Detailed descriptions of the Aspect and Polarity categories are provided in Table 1 and Table 2, respectively.

TABLE I
ASPECT CATEGORIES WITH DESCRIPTIONS AND EXAMPLES FROM THE DATASET.

Code	Aspect	Description	Example
I	Instructor	Reviews that focus on instructor-related elements such as teaching style, delivery, communication, or subject expertise.	“funny and engaging instructor breaking down the basics of physics in a simple and easy to understand way!”
C	Content	Feedback concerning the course material, topics covered, assignments, or overall subject matter.	“fantastic, i never took a physics course when i was getting my undergrad degree. this course has been great in helping me to understand "how things work". it was really well put together and easy to understand.”
S	Structure	Comments addressing the course’s organization, sequencing, or clarity in presenting learning objectives.	“very limited information on wordpress. then added some lectures on web design that could have used information relevant to wordpress such as image compression plugins instead of tinypng.com (see last component).”
SC	Design	Reviews reflecting both the structure and content, emphasizing how they work together to shape the course.	“learning material is too good and very well organised too. overall, i have good learning experience here. thanks coursera.”
ISC	General	Broad or overall evaluations that reference multiple aspects of the course, such as content, delivery, and structure.	“everything is awesome .thank you for everything!”

TABLE II
POLARITY CATEGORIES WITH DESCRIPTIONS AND EXAMPLES FROM THE DATASET.

Code	Polarity	Description	Example
N	Negative	Reviews expressing dissatisfaction, criticism, or unfavorable sentiment toward an aspect.	“very bad tool use for uploading assignment.”
NEU	Neutral	Comments that are objective, balanced, or without strong positive or negative sentiment.	“you should provide support for python 2 and python 3.”
P	Positive	Feedback that conveys satisfaction, praise, or favorable opinions regarding an aspect.	“this course is very useful!”

The distribution of reviews across aspect categories is illustrated in Figure 3. As the figure shows, there is a significant imbalance in the number of reviews assigned to each aspect. For instance, the Content category contains 12,600 reviews, whereas the Structure category includes only 1801 reviews.

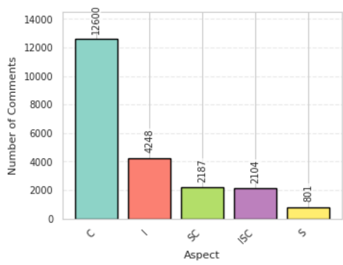

Fig. 3. Distribution of Aspects Mentioned.

The sentiment polarity for each aspect can be classified as positive (P), negative (N), or neutral (NEU). The distribution of reviews across these polarity categories is presented in Figure 4.

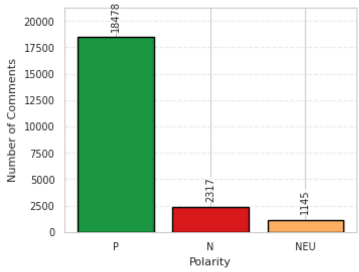

Fig. 4. Distribution of Sentiment Polarities.

B. *Exploratory Analysis of Aspect–Sentiment Interactions*

The exploratory analysis of the MOOC review dataset revealed imbalances between aspects and sentiment polarity. Content (C) is predominantly associated with positive sentiment, whereas Structure (S) and Instructor (I) exhibit a more balanced distribution, including negative and neutral comments. The sentiment–aspect correlation heatmap confirms these patterns, showing high polarity density for positive evaluations in Content and Instructor, while Structure and Design (SC) display a wider sentiment spread.

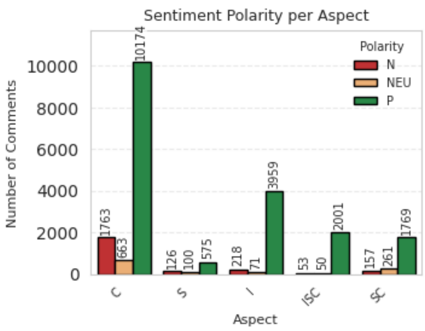

Fig. 5. Sentiment Polarity per Aspect.

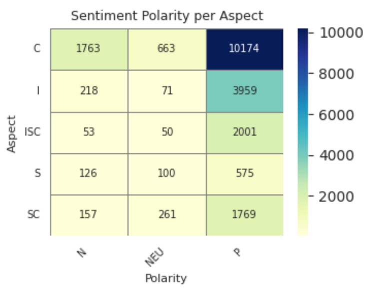

Fig. 6. Sentiment-Aspect correlation heatmap.

Furthermore, the average comment length varies significantly across aspect–sentiment pairs (Figure 7). For instance, negative comments associated with Content and Instructor tend to be substantially longer, potentially reflecting detailed critiques. Conversely, positive comments under General (ISC) or Structure (S) aspects are notably shorter, indicating more concise satisfaction expressions. These insights underscore the nuanced interdependence between sentiment polarity, aspect category, and review verbosity, providing valuable guidance for designing aspect-aware sentiment models capable of capturing such complex patterns.

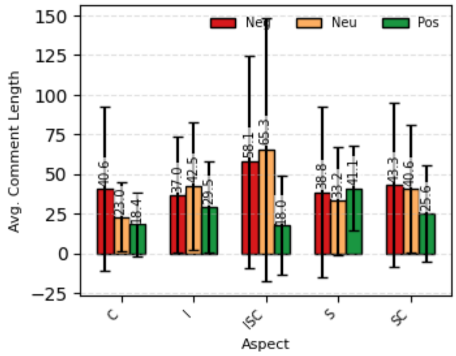

Fig. 7. Average Comment Length by Aspect and Sentiment Polarity.

The exploratory analysis guided key design choices in the proposed model. The imbalance among sentiment–aspect pairs motivated the use of a multi-task learning framework for shared knowledge between aspect and sentiment prediction. Correlations across tasks justified adding an inter-task matching module to enhance semantic representation, while variations in review length and writing style highlighted the need to combine convolutional layers for local context with multi-head attention for long-range dependencies. Together, these insights shaped a robust architecture suited for aspect-based sentiment analysis in MOOCs.

C. *Experimental Setup*

All learner reviews were cleaned by removing HTML tags and extra spaces, then stored in a dedicated column for tokenization. Aspect and polarity labels were converted to integer indices, and the dataset was stratified by aspect–polarity pairs to ensure balanced class representation. Rare combinations with fewer than two samples were discarded. The final split allocated 72% for training, 13% for validation, and 15% for testing. This configuration ensures sufficient training data while maintaining separate validation and test sets for model selection, hyperparameter tuning, and unbiased evaluation.

Reviews were tokenized with the BERT-base-uncased tokenizer and standardized to 128 tokens. Token IDs and attention masks served as model inputs through a custom PyTorch Dataset for multitask learning.

The MTL-BERTNet model employs frozen BERT embeddings followed by a convolutional layer (kernel size = 3, ReLU) and a four-head attention module. Shared features are processed by two task-specific branches linked through inter-task matching layers, enabling mutual learning between aspect and sentiment predictions.

The CNN kernel size of 3 is selected to capture local trigram-level contextual patterns, which is commonly effective for short-text classification tasks such as user reviews. The four-head attention mechanism is used as a lightweight configuration to balance representational capacity and computational efficiency compared to full multi-head settings in standard Transformer architectures.

Training used the AdamW optimizer (learning rate = 2e-5, batch size = 32) for up to ten epochs with early stopping (patience = 2) based on validation macro F1. The learning rate is chosen following standard practice for fine-tuning transformer-based models and is validated using the validation set performance. Class weights were applied to mitigate imbalance in aspect prediction. Experiments were run on GPU with fixed seeds for reproducibility. Evaluation metrics included accuracy, macro F1-score, confusion matrices, and ROC curves. Model parameters are summarized in Table 3.

TABLE III
MODEL AND TRAINING CONFIGURATION.

Component	Value
Tokenizer	BERT-base-uncased
Max Input Length	128 tokens
Embedding Source	Frozen BERT
CNN Kernel Size	3
CNN Activation	ReLU
Attention Heads	4
Inter-Task Matching	MLP (ReLU)
Optimizer	AdamW
Learning Rate	2e-5
Batch Size	32
Epochs	10 (with early stopping)
Early Stopping Patience	2
Loss Function (Aspect)	Cross-Entropy with class weights
Loss Function (Polarity)	Cross-Entropy
Stratified Splitting	By Aspect–Polarity combination
Evaluation Metrics	Accuracy, F1 (macro), ROC, CM
Reproducibility Settings	Random seed = 42
Device	GPU

V. RESULTS AND DISCUSSION

To evaluate the effectiveness of our proposed MTL-BERTNet architecture, we conducted a thorough training procedure across 10 epochs using stratified train–validation–test splits. This stratification preserved the joint distribution of aspect–sentiment label pairs, which is particularly important given the inherent class imbalance in MOOC reviews. To mitigate overfitting, we applied early stopping based on the macro-averaged F1-score of the aspect classification task. The evolution of training and validation metrics—including loss, accuracy, and F1-score—for both tasks is presented in Figure 8.

As shown in Figure 9, the model achieved rapid improvements during the first few epochs, with performance peaking at Epoch 3. Beyond this point, marginal gains in validation F1-score were observed, leading to early stopping after epoch 5. This pattern indicates strong generalization and mitigated overfitting.

The final selected model (epoch 3) attained a validation accuracy of 90.88% for aspect prediction and 96.31% for polarity prediction, along with macro F1-scores of 0.8911 and 0.9403, respectively. These results demonstrate the model's capacity to effectively capture both structural and sentiment-specific information from MOOC reviews.

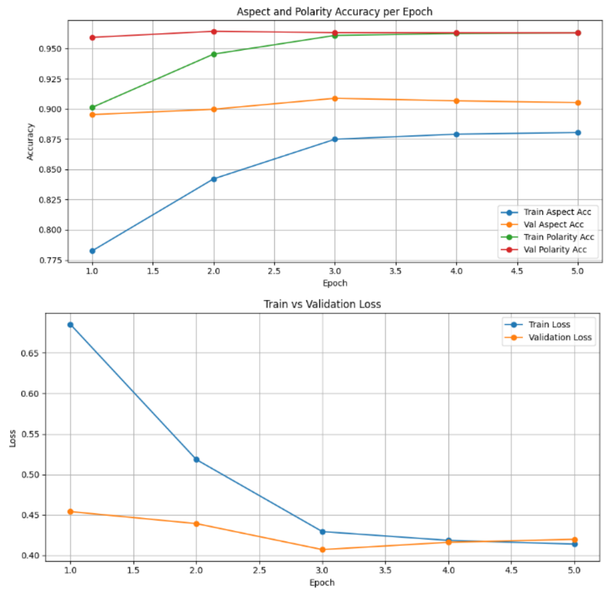

Fig. 8. Model Accuracy and Loss Trends During Training.

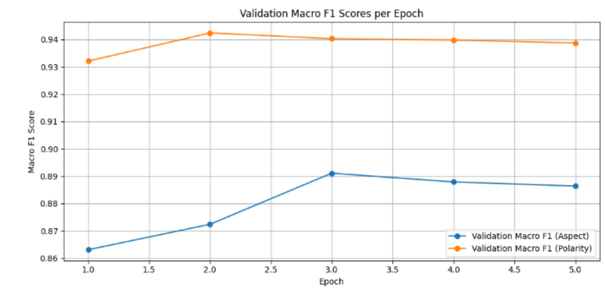

Fig. 9. Validation Macro F1 Score Progression.

To further evaluate generalization, the best model was tested on the held-out test set. The classification performance is summarized in Table 4. The model achieved:

- 92% overall accuracy on aspect classification with a macro F1-score of 0.90, and
- 96% accuracy on polarity classification with a macro F1-score of 0.93.

Notably, the model exhibited strong performance on dominant aspect classes like Content (C) and SC, achieving F1-scores above 0.94. Performance was comparatively lower for underrepresented or semantically diffuse classes such as Structure (S) and Instructor (I). In the sentiment task, the model delivered exceptional performance on Positive (P) comments (F1 = 0.98) and remained robust on more ambiguous classes like Neutral (NEU) and Negative (N).

TABLE IV
TEST SET CLASSIFICATION REPORT.

Aspect Class	Precision	Recall	F1-Score	Support
Content (C)	0.93	0.95	0.94	1891
Instructor (I)	0.85	0.85	0.85	652
General(ISC)	0.95	0.90	0.92	331
Structure (S)	0.83	0.77	0.80	270
Design(SC)	0.96	0.97	0.97	495
Macro Avg	0.91	0.89	0.90	3639

Polarity Class	Precision	Recall	F1-Score	Support
Negative (N)	0.86	0.89	0.88	516
Neutral (NEU)	0.96	0.91	0.93	352
Positive (P)	0.98	0.98	0.98	2771
Macro Avg	0.93	0.93	0.93	3639

The confusion matrices in Figure 10 illustrate MTL BERTNet's performance on the test set, highlighting challenges related to class imbalance. For aspect classification, the model accurately identifies high-resource categories such as Content (C) and SC, with 1792/1891 and 479/495 correct predictions, respectively. Misclassifications occur mainly between Content and Instructor (74 instances) and between Structure and Instructor (11 instances), while 42 Structure samples were predicted as Content, reflecting semantic overlaps between instructional content and pedagogical structure. The underrepresented Instructor class was frequently confused with Content (78) and Structure (8), indicating the difficulty of modeling sparsely represented or diffuse concepts. Despite these issues, recall and precision remain high across all aspect classes.

For polarity prediction, Positive (P) samples were classified with high accuracy (2708/2771), while some confusion exists between Neutral (NEU) and Negative (N). Specifically, 47 Negative samples were predicted as Positive and 10 as Neutral, and 14 Neutral instances were classified as Positive, highlighting the subjective nature of sentiment expression.

These observations underscore the benefit of joint learning in MTL-BERTNet, as shared representations capture interdependencies between aspects and sentiment. For example, structural criticism often conveys subtle negativity, and instructor praise implies positivity even with minimal sentimental cues, helping the model handle nuanced sentiment aspect interactions.

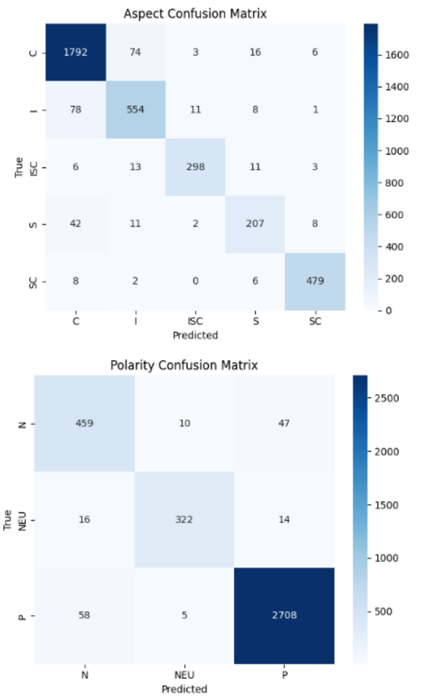

Fig. 10. Confusion Matrices for Aspect and Polarity Tasks.

To gain deeper understanding of how the model handles minority aspect classes, we visualized a filtered confusion matrix for Structure (S) and ISC, two of the least frequent categories, as shown in Figure 11. Despite their low representation, the model correctly identified 196 out of 197 Structure samples and 304 out of 311 ISC samples, with only one Structure sample misclassified and seven ISC samples predicted as Structure. This impressive performance demonstrates that the model's attention and encoding layers are effective at capturing subtle context features, even when class frequencies are low. It also highlights the impact of macro-F1 optimization and stratified sampling during training, which preserve balance and encourage generalization across classes of varying sizes.

Collectively, these findings indicate that MTL-BERTNet is capable of learning both dominant and minority patterns in imbalanced sentiment–aspect datasets. Nonetheless, the remaining misclassifications underscore the need for further refinement, including context-aware attention, data augmentation techniques for rare classes, and integration of external knowledge resources to aid in disambiguating closely related concepts.

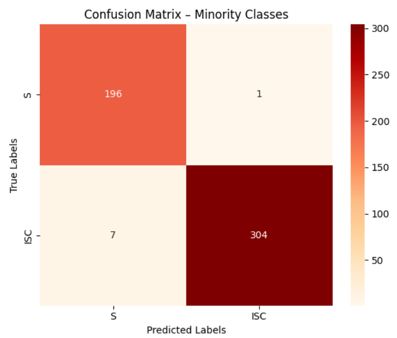

Fig. 11. Filtered Confusion Matrix for Minority Aspect Classes.

To further assess the discriminative capacity of the proposed MTL-BERTNet model, we conducted a multi-class Receiver Operating Characteristic (ROC) analysis for both aspect and polarity classification tasks, as illustrated in Figure 12. These curves provide a comprehensive view of the model's performance across all classes beyond simple accuracy or F1 metrics, particularly in the presence of class imbalance.

As shown in the Figure, the model achieved a macro-average AUC of 0.983 and a micro-averaged AUC of 0.990 for aspect prediction. The class-specific curves highlight that the model performs particularly well on dominant and structurally distinct categories such as SC (AUC = 0.999) and ISC (AUC = 0.994). Even for more semantically overlapping categories like Content (C) and Instructor (I), the AUCs remain strong at 0.981 and 0.976, respectively, indicating a high true positive rate across thresholds.

In the polarity classification task, the ROC analysis further underscores the robustness of the model. With a macro-average AUC of 0.992 and micro-average AUC of 0.996, the model exhibits outstanding predictive performance. Class-specific

AUC values are consistently high: 0.988 for Negative (N), 0.995 for Neutral (NEU), and 0.992 for Positive (P). This high separability suggests that the learned representations effectively capture the sentiment polarity nuances present in learner reviews.

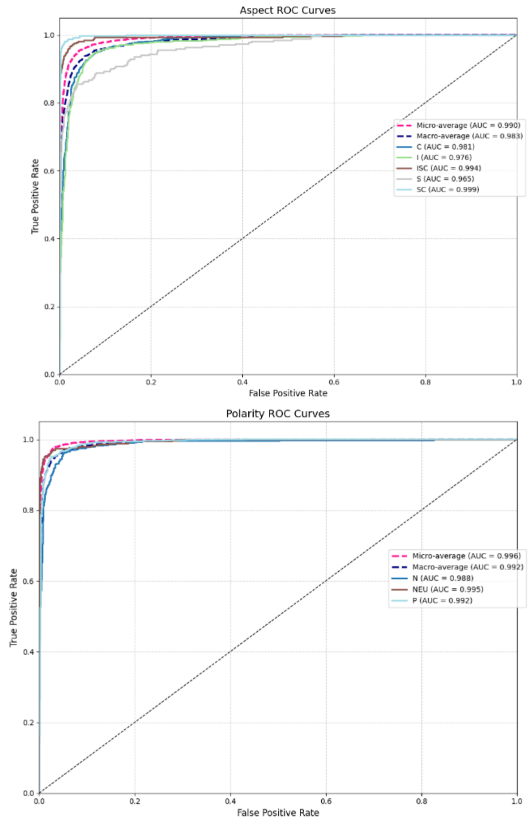

Fig. 12. Multi-Class ROC Curves for Aspect and Polarity Classification (MTL-BERTNet).

These results underscore the effectiveness of our model in capturing the intricate interdependence between aspect classification and sentiment polarity, particularly within the nuanced context of educational reviews. The integration of convolutional filters and multi-head attention mechanisms enhances the model's ability to detect fine-grained linguistic patterns and emphasize contextually salient tokens. Moreover, the inter-task matching layer facilitates cross-task interaction, enabling the model to associate sentiment expressions with their corresponding aspects more coherently. This synergy proves especially valuable in domains like online learning, where user feedback is often dense with implicit opinions and overlapping semantic cues.

From a practical standpoint, these findings have significant implications for the analysis and personalization of MOOCs. Accurate identification of both what learners are commenting on (e.g., course structure, content quality, instructor effectiveness) and how they feel about it (positive, neutral, or negative sentiment) enables more targeted course refinement and improved learner support. For example, consistently low sentiment linked to a specific aspect like "Structure" can alert educators to redesign the learning flow, while strong positive feedback on "Content" can reinforce current pedagogical strategies. Ultimately, by leveraging models like MTL-BERTNet, MOOC platforms can transform unstructured learner feedback into actionable insights, enhancing course quality, increasing engagement, and fostering better learning outcomes.

A. Comparison with Baseline Models

We evaluated the proposed MTL-BERTNet model against several baselines, including classical machine learning algorithms (SVM, Logistic Regression, Random Forest) trained on TF-IDF feature representations of the input text, and deep learning models (CNN, LSTM, BiLSTM). For traditional models, each review was first preprocessed (tokenization, lowercasing, stop-word removal), and then transformed into a TF-IDF weighted bag-of-words vector representation capturing unigram and bigram statistics. Classical methods showed moderate performance for aspect classification (macro F1: 0.51–0.58), with SVM providing the best trade-off between accuracy and generalization. Deep learning models improved sentiment prediction (e.g., LSTM: 88.56% accuracy, 0.77 macro F1) but struggled with less frequent aspect classes, resulting in macro F1 below 0.70.

In contrast, MTL-BERTNet, which combines shared representations, convolutional encoding, and multi-head attention, outperformed all baselines, achieving 92.00% aspect accuracy (macro F1 = 0.90) and 96.00% sentiment accuracy (macro F1 = 0.93). These results demonstrate the effectiveness of multi-task learning in capturing dependencies between aspects and sentiments in MOOC reviews.

TABLE V
PERFORMANCE COMPARISON OF BASELINE MODELS VS. MTL-BERTNET.

Model	Aspect		Sentiment	
	Accuracy	Macro F1	Accuracy	Macro F1
Decision Tree	61.61%	52%	79.61%	61%
Naive Bayes	65.73%	46%	84.53%	64%
SVM	71.56%	58%	87.14%	71%
Random Forest	70.71%	58%	85.68%	68%
Logistic Regression	69.17%	53%	87.11%	72%
CNN	72.31%	64.83%	87.35%	74.19%
LSTM	74.13%	67.94%	88.56%	77.22%
BiLSTM	73.33%	67.65%	87.93%	75.93%
Our Model	92.00%	90%	96.00%	93%

B. Ablation Study

To assess the contribution of each component within our proposed MTL-BERTNet architecture, we conducted a comprehensive ablation study by evaluating several model variants. The BertSTL model serves as a single-task baseline using BERT, applied separately to either aspect or sentiment classification. We tested the Vanilla-MTL-BERTNet version, which removes the Inter-Task Matching Layer (IML), to evaluate the contribution of cross-task knowledge sharing. The MTL-BERTNet-CNN variant excludes the Multi-Head Attention module to assess the importance of capturing global contextual dependencies, while MTL-BERTNet-NonBERT replaces BERT with pre-trained GloVe embeddings to examine the impact of contextualized representations.

The ablation study is designed to isolate the contribution of each architectural component rather than to maximize absolute performance improvements, which is consistent with standard practice in multi-task learning evaluation.

Compared with the BertSTL baseline, the Vanilla-MTL-BERTNet model improves the Aspect Macro F1-score from 86.10% to 88.76% (+2.66%), demonstrating the effectiveness of the multi-task learning framework. Incorporating the CNN–MHA module further increases the Macro F1-score to 89.42% (+0.66%), highlighting the benefit of jointly capturing local and global contextual dependencies. Adding the Inter-Task Matching Layer yields a further improvement to 90.00% (+0.58%), indicating that cross-task information exchange between aspect and sentiment representations provides complementary knowledge that enhances prediction performance. Overall, these results demonstrate that each component contributes positively and incrementally to the effectiveness of MTL-BERTNet, with the complete model achieving the highest accuracy and Macro F1-score across both tasks.

TABLE VI
ABLATION STUDY RESULTS ON ASPECT AND SENTIMENT CLASSIFICATION.

Model	Aspect		Sentiment	
	Accuracy	Macro F1	Accuracy	Macro F1
BertSTL (Single-task BERT)	88.52%	86.10%	92.14%	90.78%
Vanilla-MTL-BERTNet	90.34%	88.76%	93.85%	92.01%
MTL-BERTNet-CNN	91.21%	89.42%	94.02%	92.38%
MTL-BERTNet-NonBERT (GloVe)	89.65%	87.32%	92.77%	91.24%
MTL-BERTNet (Our Model)	92.00%	90%	96.00%	93%

VI. CONCLUSION

In this work, we introduced MTL-BERTNet, a multi-task learning model that uses BERT to handle aspect classification and sentiment analysis on MOOC reviews simultaneously. Our model captures greater semantic linkages in learner feedback by including an inter-task matching layer and exchanging learnt representations across tasks. The effectiveness and dependability of MTL-BERTNet are demonstrated by experimental results, which routinely outperform single-task techniques and several top baseline models.

These results demonstrate how useful multi-task learning frameworks are for better understanding the thoughts and concerns of MOOC participants, which can support course improvement and more personalized learning experiences.

Future work will evaluate MTL-BERTNet on larger and more diverse educational datasets across multiple MOOC platforms to further assess its robustness and generalization capability. We also plan to integrate explainability techniques to provide more interpretable and actionable insights for educators and platform developers. In addition, exploring domain adaptation methods and incorporating real-time feedback represent promising directions to enhance the model's practicality and real-world applicability.

REFERENCES

[1] 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, Aug. 2024, pp. 1–7. DOI: 10.1145/3659677.3659733.

[2] B. Liu, "Sentiment Analysis: A Fascinating Problem," in Sentiment Analysis and Opinion Mining, B. Liu, Ed., Cham: Springer International Publishing, 2012, pp. 1–8. DOI: 10.1007/978-3-031-02145-9_1.

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, P. Nakov and T. Zesch, Eds., Dublin, Ireland: Association for Computational Linguistics, Aug. 2014, pp. 27–35. DOI: 10.3115/v1/S14-2004.

[4] J. Zeng, K. Luo, Y. Lu, and M. Wang, "An Evaluation Framework for Online Courses Based on Sentiment Analysis Using Machine Learning," *Int. J. Emerg. Technol. Learn.* IJET, vol. 18, pp. 4–22, Sep. 2023, DOI: 10.3991/ijet.v18i18.42521.

[5] W. Awadh, R. Sulaiman, and M. Mahmoud, "Aspect-based sentiment analysis in MOOCs: a systematic literature review introducing the MASC- MEF framework," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 37, Mar. 2025, DOI: 10.1007/s44443-025-00018-1.

[6] S. Deshpande *et al.*,"Elevating educational insights: sentiment analysis of faculty feedback using advanced machine learning models," *Adv. Contin. Discrete Models*, vol. 2025, Apr. 2025, DOI: 10.1186/s13662-025-03933-9.

[7] O. Habimana, Y. Li, R. Li, X. Gu, and G. Yu, "Sentiment analysis using deep learning approaches: an overview," *Sci. CHINA Inf. Sci.*, vol. 63, no. 1, Dec. 2019, DOI: 10.1007/S11432-018-9941-6.

[8] S. Ruder, "An Overview of Multi-Task Learning in Deep Neural Networks," Jun. 15, 2017, *arXiv: arXiv:1706.05098*. DOI: 10.48550/arXiv.1706.05098.

[9] H. Zhang, L. Xiao, W. Chen, Y. Wang, and Y. Jin, "Multi-Task Label Embedding for Text Classification," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, E. Riloff, D. Chiang, J. Hockenmaier, and J. Tsujii, Eds., Brussels, Belgium: Association for Computational Linguistics, Oct. 2018, pp. 4545–4553. DOI: 10.18653/v1/D18-1484.

[10] R. Caruana, "Multitask Learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997, DOI: 10.1023/A:1007379606734.

[11] A. Maurer, M. Pontil, and B. Romera-Paredes, "The Benefit of Multitask Representation Learning".

[12] G. Balikas, S. Moura, and M.-R. Amini, "Multitask Learning for Fine- Grained Twitter Sentiment Analysis," in *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, in SIGIR '17. New York, NY, USA: Association for Computing Machinery, Aug. 2017, pp. 1005–1008. DOI: 10.1145/3077136.3080702.

[13] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*: Human Language Technologies, Volume 1 (Long and Short Papers), J. Burstein, C. Doran, and T. Solorio, Eds., Minneapolis, Minnesota: Association for Computational Linguistics, Jun. 2019, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

[14] C. Abdessamad and N. El Faddouli, "Sentiment Analysis on Massive Open Online Courses (MOOCs): Multi-Factor Analysis, and Machine Learning Approach," *Int. J. Inf. Commun. Technol. Educ.*, vol. 18, pp. 1–22, Jan. 2022, DOI: 10.4018/IJICTE.310004.

[15] N. D. Valakunde and M. S. Patwardhan, "Multi-aspect and Multi-class Based Document Sentiment Analysis of Educational Data Catering Accreditation Process," in *2013 International Conference on Cloud & Ubiquitous Computing & Emerging Technologies*, Nov. 2013, pp. 188–192. DOI: 10.1109/CUBE.2013.42.

[16] L. Balachandran and A. Kirupananda, "Online reviews evaluation system for higher education institution: An aspect based sentiment analysis tool," in *2017 11th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, Dec. 2017, pp. 1–7. DOI: 10.1109/SKIMA.2017.8294118.

[17] N. Soe and P. T. Soe, "Domain Oriented Aspect Detection for Student Feedback System," pp. 90–95, Nov. 2019, DOI: 10.1109/aitc.2019.8921372.

[18] I. Sindhu, S. M. Daudpota, K. Badar, M. Bakhtyar, J. Baber, and M. Nurunnabi, "Aspect-based opinion mining on student's feedback for faculty teaching performance evaluation," *IEEE Access*, vol. 7, pp. 108 729–108 741, 2019.

[19] G. S. Chauhan, P. Agrawal, and Y. K. Meena, "Aspect-Based Sentiment Analysis of Students' Feedback to Improve Teaching–Learning Process," in *Information and Communication Technology for Intelligent Systems,* Springer, Singapore, 2019, pp. 259–266. DOI: 10.1007/978-981-13-1747-7_25.

[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 ACM Other conferences., 2020, pp. 510–514. DOI: 10.1145/3404555.3404633.

[21] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly Supervised Framework for Aspect-Based Sentiment Analysis on Students' Reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106 799–106 810, 2020, DOI: 10.1109/ACCESS.2020.3000739.

[22] M. Joshi, D. Chen, Y. Liu, D. S. Weld, L. Zettlemoyer, and O. Levy, "SpanBERT: Improving Pre-training by Representing and Predicting Spans," Jan. 18, 2020, *arXiv: arXiv:1907.10529*. DOI: 10.48550/arXiv.1907.10529.

[23] Y. Liu *et al*., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," Jul. 26, 2019, *arXiv: arXiv:1907.11692*. DOI: 10.48550/arXiv.1907.11692.

[24] A. Jazuli, Widowati, and R. Kusumaningrum, "Aspect-based sentiment analysis on student reviews using the Indo-Bert base model," presented at the E3S Web of Conferences, Aug. 2023, p. 02004. DOI: 10.1051/e3sconf/202344802004.

[25] K. Alshaikh, O. Almatrafi, and Y. Abushark, "BERT-based Model for Aspect-Based Sentiment Analysis for Analyzing Arabic Open-ended Survey Responses: A Case Study," *IEEE Access,* vol. PP, pp. 1–1, Jan. 2023, DOI: 10.1109/ACCESS.2023.3348342.

[26] I. Id, H. Saputra, S. Syamsudhuha, R. Kurniawan, and Y. Andriyani, "Sentiment analysis of student evaluation feedback using transformer-based language models," *Indones. J. Electr. Eng. Comput. Sci*., vol. 36, p. 1127, Nov. 2024, DOI: 10.11591/ijeecs.v36.i2.pp1127-1139.

[27] B. Clavié and K. Gal, "EduBERT: Pretrained Deep Language Models for Learning Analytics," Dec. 02, 2019, *arXiv: arXiv:1912.00690*. DOI: 10.48550/arXiv.1912.00690.

[28] W. Wang, H. Zhuang, M. Zhou, H. Liu, and B. Li, "What Makes a Star Teacher? A Hierarchical BERT Model for Evaluating Teacher's Performance in Online Education," Dec. 03, 2020, *arXiv: arXiv:2012.01633*. DOI: 10.48550/arXiv.2012.01633.

[29] M. Herath, K. Chamindu, H. Maduwantha, and S. Ranathunga, "Dataset and Baseline for Automatic Student Feedback Analysis," in *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, J. Odijk, and S. Piperidis, Eds., Marseille, France: European Language Resources Association, Jun. 2022, pp. 2042–2049. Accessed: Jul. 21, 2025. [Online]. Available: https://aclanthology.org/2022.lrec-1.219/

[30] A. Koufakou, "Deep learning for opinion mining and topic classification of course reviews," *Educ. Inf. Technol*., vol. 29, no. 3, pp. 2973–2997, Feb. 2024, DOI: 10.1007/s10639-023-11736-2.

[31] X. Li and W. Lam, "Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, M. Palmer, R. Hwa, and S. Riedel, Eds., Copenhagen, Denmark: Association for Computational Linguistics, Sep. 2017, pp. 2886–2892. DOI: 10.18653/v1/D17-1310.

[32] P. Liu, X. Qiu, and X. Huang, "Adversarial Multi-task Learning for Text Classification," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), R. Barzilay and M.-Y. Kan, Eds., Vancouver, Canada: Association for Computational Linguistics, Jul. 2017, pp. 1–10. DOI: 10.18653/v1/P17-1001.

[33] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "An Interactive Multi-Task Learning Network for End-to-End Aspect-Based Sentiment Analysis," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, A. Korhonen, D. Traum, and L. Màrquez, Eds., Florence, Italy: Association for Computational Linguistics, Jul. 2019, pp. 504–515. DOI: 10.18653/v1/P19-1048.

[34] G. Zhao, Y. Luo, Q. Chen, and X. Qian, "Aspect-based sentiment analysis via multitask learning for online reviews," *Knowl.-Based Syst*., vol. 264, p. 110 326, Mar. 2023, DOI: 10.1016/j.knosys.2023.110326.

[35] Z. He, H. Wang, and X. Zhang, "Multi-Task Learning Model Based on BERT and Knowledge Graph for Aspect-Based Sentiment Analysis," *Electronics*, vol. 12, no. 3, Art. no. 3, Jan. 2023, DOI: 10.3390/electronics12030737.

[36] S. Gul, M. Asif, F.-E.- Amin, K. Saleem, and M. Imran, "Advancing Aspect-Based Sentiment Analysis in Course Evaluation: A Multi-Task Learning Framework With Selective Paraphrasing," *IEEE Access,* vol. 13, pp. 7764–7779, 2025, DOI: 10.1109/ACCESS.2025.3527367.

[37] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "DiSAN: Directional Self-Attention Network for RNN/CNN-Free Language Understanding," Nov. 20, 2017, *arXiv: arXiv:1709.04696*. DOI: 10.48550/arXiv.1709.04696.

[38] A. Vaswani *et al*., "Attention Is All You Need," Aug. 02, 2023, *arXiv: arXiv:1706.03762*. DOI: 10.48550/arXiv.1706.03762.

[39] A. W. Yu *et al*., "QANet: Combining Local Convolution with Global Self-Attention for Reading Comprehension," Apr. 23, 2018, *arXiv: arXiv:1804.09541*. DOI: 10.48550/arXiv.1804.09541.

[40] B. Yang, L. Wang, D. Wong, L. S. Chao, and Z. Tu, "Convolutional Self-Attention Networks," Apr. 05, 2019, *arXiv: arXiv:1904.03107*. DOI: 10.48550/arXiv.1904.03107.

[41] N. Majumder, S. Poria, H. Peng, N. Chhaya, E. Cambria, and A. Gelbukh, "Sentiment and Sarcasm Classification With Multitask Learning," *IEEE Intell. Syst*., vol. 34, no. 3, pp. 38–43, May 2019, DOI: 10.1109/MIS.2019.2904691.

[42] C. Xiong, V. Zhong, and R. Socher, "Dynamic Coattention Networks For Question Answering," Mar. 06, 2018, *arXiv: arXiv:1611.01604*. DOI: 10.48550/arXiv.1611.01604.

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 e-mail: 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 e-mail: h.mouncif@usms.ma.