Mining Improvement Points for University Teaching Feedback Based on BERT Embedding and Sentiment Convolutional Network

Yiming Cai[1], Zhiteng Zhao[2], Meiqi Wang[3], Yi Yin[4]
[1]College of Politics and Public Administration, Shandong Youth University of Political Science, Jinan 250103, Shandong, China
[2]Department of Computer and Software Engineering, Shandong College of Electronic Technology, Jinan 250200, Shandong, China
[3]Pingyi County Audit Bureau, Linyi 273300, Shandong, China
[4]School of Art Design, Shandong Youth University of Political Science, Jinan 250103, Shandong, China

Corresponding author: Zhiteng Zhao (e-mail: 18653538400@163.com)

ABSTRACT University teaching feedback contains contextual semantics, implicit sentiment, and fragmented improvement suggestions, which makes automatic extraction difficult when only keyword or shallow syntactic features are used. To improve semantic understanding and emotion-aware localization, this study proposes a teaching-feedback mining model combining BERT embedding, a multi-channel sentiment convolutional network, emotional attention, and BiLSTM-CRF sequence labeling. After text cleaning, sentence segmentation, stop-word removal, and BIO-based sentiment annotation, BERT is used to generate deep contextual representations. Multi-scale convolutional kernels then extract local sentiment patterns, and an emotional attention mechanism enhances features highly related to improvement-point identification. The fused semantic and sentiment features are jointly modeled through a dual-task structure for sentiment classification and sequence labeling. Experiments on 18,000 teaching-feedback samples show that the model achieves an accuracy of 87.2%, an F1 score of 84.1%, and an improvement-point completeness of 80.5%, outperforming static word embeddings, LSTM embeddings, and single-task baselines. The proposed method provides an engineering-oriented text-signal processing framework for semantic feature extraction, sentiment modeling, and intelligent feedback-analysis systems.

INDEX TERMS bert embedding, sentiment convolutional network, semantic signal processing, bilstm-crf, text mining

I. INTRODUCTION

WITH the deepening of higher education quality assurance systems, student teaching feedback is becoming increasingly important in teaching evaluation and improvement. Its text comprehensively reflects students' views on teaching methods, course content, classroom interactions, and learning experiences, covering multidimensional information such as teaching organization, knowledge transfer, and learning outcomes [1-3]. Feedback is complex and emotionally intertwined, combining objective evaluations with subjective attitudes, placing higher demands on the quantitative evaluation of teaching quality [4-6]. Systematic analysis of feedback can provide valuable insights for refining classroom instruction and identifying specific areas for pedagogical adjustment [7-9]. Through deep mining and feature modeling, the proposed framework facilitates automated and refined evaluation, fosters scientific educational decision-making, and advances the sustainable development of teaching reform [10-12]. Therefore, building an efficient and intelligent mechanism to accurately extract improvement points has become an important direction in teaching quality research [13-15].

Teaching feedback texts in universities often contain many complex factors. Complex language, varied semantics, implicit emotions, and fragmented opinions increase the difficulty of automatically identifying improvement points [16-18]. Students' free expression results in loose sentences but strong semantic connections, making it difficult to capture deeper meaning through word frequency or superficial features [19-21]. Emotions and suggestions are intertwined, and emotional expressions interfere with the identification of key information [22-24]. Semantic ambiguity, logical jumps, and redundancy further render traditional keyword or syntactic methods inaccurate, easily missing valid content or misjudging non-substantive opinions [25-27]. Therefore, there is an urgent need to introduce deep semantic understanding and sentiment modeling to improve analytical effectiveness.

In response to the difficulties in analyzing teaching feedback, some studies rely on traditional machine learning methods and artificial feature engineering. Masood K et al. used text and network mining to analyze social media data to extract course feedback [28]. Ren P et al. combined deep learning with dictionary methods to perform aspect-based sentiment analysis [29]. Jatain D et al. proposed a hybrid bio-inspired feature selection method to mine learner comments [30]. However, these methods have limitations in processing semantic diversity. Some studies applied CNN (Convolutional Neural Network) or RNN (Recurrent Neural Network) to capture local context or sequence dependencies. Ma X et al. combined K-means with recursive convolutional neural networks to achieve dimensionality reduction and classification of student feedback features [31]. Asad Abdi et al. constructed a CNN-BiLSTM-attention model for sentiment analysis [32]. However, this method still lacks global semantic modeling and has limited accuracy in improvement point identification. In recent years, pre-trained language models such as BERT have been applied to sentiment analysis and opinion mining. Zyout I et al. analyzed the behavioral mechanism of prompt-based pre-trained language models in sentiment analysis and emotion identification [33]. Jazuli A et al. implemented fine-grained sentiment analysis based on Indonesian BERT [34]. Shuqin H et al. combined BERT, BiLSTM, and attention mechanisms to capture the fine semantics of MOOC reviews [35]. Despite progress in semantic representation, there is a lack of coordination between sentiment identification and improvement point location. How to strengthen sentiment features based on deep semantic modeling and achieve joint modeling of the two is still an urgent problem to be solved. To address the problem of inaccurate improvement point identification caused by the difficulty in collaboratively modeling semantic understanding and sentiment features, this paper proposes a method that combines BERT embedding with a multi-channel sentiment convolutional network. The model uses BERT to obtain contextual semantic representations, extracts multi-scale sentiment features using heterogeneous convolutional kernels in multi-channel convolutional layers, and enhances information related to improvement points through a sentiment attention mechanism. Finally, the semantic and sentiment features are jointly input into a sequence labeling module to automatically locate and identify improvement points. This method establishes a synergistic mechanism between deep semantic understanding and sentiment modeling, significantly improving the accuracy and application value of identifying improvement points in teaching feedback.

II. MINING IMPROVEMENTS IN TEACHING FEEDBACK BASED ON BERT EMBEDDING AND MULTI-CHANNEL SENTIMENT CONVOLUTIONAL NETWORK

A. DATA PREPROCESSING AND INPUT REPRESENTATION

University teaching feedback text originates from course evaluation systems. The raw data suffers from redundant symbols, missing semantics, and sentiment noise. To improve the quality of the corpus, standardized cleaning and structured annotation are implemented. These procedures include sentence segmentation, the removal of invalid symbols and stop words, and the retention of domain-specific vocabulary. Subsequently, a unified labeling system is constructed, employing a BIO (Begin-Inside-Outside) annotation strategy to jointly encode the improvement point entities and sentiment. To mitigate the category imbalance inherent in sequence labeling, class weights are integrated into the loss function, assigning higher penalties to the rarer "Improvement" tags during training. Specifically, the label set is expanded to include sentiment-specific tags: B-IMP-POS, I-IMP-POS (Positive Improvement), B-IMP-NEG, I-IMP-NEG (Negative Improvement), and O (Outside). This allows the CRF layer to simultaneously decode boundary information and sentiment polarity. Sample ratio correction is used to mitigate category imbalance. Finally, the processed text is fed into the BERT model, where it is mapped into high-dimensional vectors using word, position, and segment embeddings, and semantic modeling is completed using a bidirectional Transformer.

B. SEMANTIC REPRESENTATION LAYER: BERT EMBEDDING

The semantic representation layer is based on BERT's bidirectional Transformer structure, which accurately captures the contextual dependencies of the teaching feedback text through deep semantic modeling. The input is linearly superimposed by word, position, and segment embedding and then enters the multi-layer self-attention network to generate a high-dimensional semantic representation. BERT's self-attention mechanism can achieve dynamic fusion of global dependency modeling and contextual information. Supposing the input sequence is $X = [x_1, x_2, \ldots, x_n]$, BERT first maps it to a word vector matrix and then generates a semantic feature matrix through the multi-head attention layer. The calculation process of a single attention head is shown in Formula (1):

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{1}$$

In Formula (1), Q, K, and V represent the query, key, and value matrices obtained by linearly transforming the input vector, respectively; d_k is the dimension of the key vector; the softmax function is used to normalize the weight distribution. This mechanism globally models sequence associations in a weighted manner, capturing the underlying dependency structure within a sentence. Multi-head attention outputs aggregate features in different semantic subspaces, and the semantic representation is updated through a feedforward network and layer normalization. The model parameters are optimized by maximizing conditional probability. The overall semantic representation structure and its connection to subsequent modules are shown in Figure 1.

The model takes cleaned and segmented teaching feedback as input and generates a contextual semantic representation that fuses word vectors, position vectors, and sentence vectors through the BERT embedding layer. This semantic representation then enters a multi-channel sentiment convolutional network, where different convolution kernels extract multi-level sentiment features. These features are filtered and fused through the sentiment attention mechanism to form a comprehensive feature with enhanced sentiment. Within a dual-task framework, these fused features are fed into the BiLSTM-CRF sequence labeling module and the sentiment classification layer, respectively, to locate improvement points and identify sentiment polarity. The final output includes the improvement points and their sentiment labels.

C. EMOTIONAL FEATURE EXTRACTION LAYER: MULTI-CHANNEL CONVOLUTIONAL NETWORK

The sentiment feature extraction layer in this model plays a key role in converting the BERT semantic embedding vector into a fine-grained sentiment feature representation. This layer captures local patterns within the high-dimensional semantic tensor through a multi-channel convolutional structure to identify the implicit sentiment-related features in the teaching feedback text. The input BERT embedding matrix is set to $M \in R^{n\times d}$, where n represents the sentence length, and d represents the word vector dimension. The convolution operation extracts contextual sentiment features of different ranges using different convolution kernels in the time step dimension. The process can be formalized as Formula (2):

$$f_i^{(k)} = \sigma\left(W^{(k)} * M_{i:i+m_k-1} + b^{(k)}\right) \tag{2}$$

In Formula (2), $f_i^{(k)}$ represents the feature value extracted by the k-th convolution kernel at position i; $W^{(k)}$ is the convolution weight matrix; m_k represents the window size of the convolution kernel; $b^{(k)}$ is the bias term; $\sigma(\cdot)$ is the nonlinear activation function. After different convolution kernels extract features with different receptive fields on the same input, a global sentiment feature map is obtained through feature concatenation $F = [f^{(1)}, f^{(2)}, \ldots, f^{(k)}]$. To enhance the sentiment discriminability of the features, the convolution output undergoes batch normalization and maximum pooling to generate a fixed-length sentiment feature tensor for use in the attention fusion layer. The multi-channel structure can capture potential sentiment patterns from different semantic scales and improve the feature discrimination of improvement point identification.

D. EMOTIONAL ATTENTION MECHANISM AND FEATURE FUSION

The emotional attention mechanism and feature fusion layer apply a weighted aggregation strategy based on emotional correlation based on the output of multi-channel convolutional features to achieve deep fusion of semantic and emotional features. It is assumed that the emotional feature tensor output by the convolutional layer is $R = [r_1, r_2, \ldots, r_t]$, and r_t represents the local feature vector of the time step t. To enhance the attention to the emotional information related to the improvement point, the emotional attention weight α_t is applied to measure the importance of the features of each time step. Its calculation process is as shown in Formula (3):

$$e_t = v_a^\top \tanh(W_h r_t + W_s s), \quad \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^{N}\exp(e_k)} \tag{3}$$

In Formula (3), W_h is the feature transformation matrix; W_s is the emotional state mapping matrix; s is the global semantic context vector, which is obtained by applying global average pooling to the final hidden state output of the BERT embedding layer; v_a is the learnable attention vector; e_t is the feature importance score. This process achieves dynamic distribution of feature relevance through nonlinear mapping and softmax normalization, allowing the model to adaptively focus on areas highly correlated with emotional polarity in different semantic segments. The fused global emotional representation is generated by weighted summation, as shown in Formula (4):

$$c = \sum_{t=1}^{N} \alpha_t r_t \tag{4}$$

In Formula (4), c is the comprehensive feature vector after sentiment enhancement. This representation preserves semantic integrity while strengthening sentiment features that are highly correlated with the improvement point. To further integrate semantic and sentiment information, linear transformation is used to achieve cross-channel fusion.

E. IMPROVEMENT POINT IDENTIFICATION LAYER: JOINT MODELING OF CLASSIFICATION AND SEQUENCE LABELING

The improvement point identification layer takes the emotion-attention fusion feature as input and realizes the automatic identification of improvement points and emotional polarity discrimination in teaching feedback through the joint modeling of classification and sequence labeling. This layer consists of BiLSTM and CRF, and applies the sentiment classification subtask in parallel to form a multi-task learning framework. BiLSTM captures contextual dependencies, and its forward and backward hidden states are concatenated to form a comprehensive feature representation. CRF models the relationship between labels to improve the accuracy of improvement point prediction. CRF constrains the transition probability of adjacent labels by defining the state transition matrix A, and jointly considers the label consistency between contexts. Given an input sequence $H = [h_1, h_2, \ldots, h_t]$ and a corresponding label

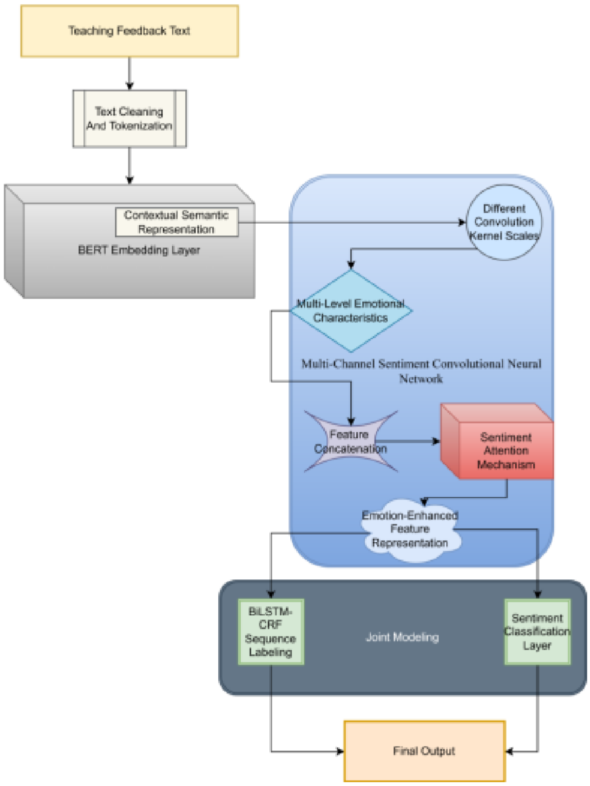

FIGURE 1. Method framework

sequence $y = [y_1, y_2, \ldots, y_t]$, its conditional probability is defined as Formula (5):

$$P(y \mid H) = \frac{\exp\left(\sum_{t=1}^{N}(A_{y_{t-1},y_t} + W_{y_t}^{\top} h_t + b_{y_t})\right)}{\sum_{y' \in Y(H)} \exp\left(\sum_{t=1}^{N}(A_{y'_{t-1},y'_t} + W_{y'_t}^{\top} h_t + b_{y'_t})\right)} \tag{5}$$

In Formula (5), A_{y_{t-1},y_t} represents the score transferred from label y_{t-1} to y_t; W_{y_t} is the weight vector of the corresponding label; b_{y_t} is the bias term; $Y(H)$ is the set of all possible label sequences. Parameter

optimization is completed by maximizing the log-likelihood function of the true label sequence. To further combine the sentiment classification task, a dual-task structure with a shared BiLSTM hidden layer is adopted to output the global sentiment classification result in the same feature space. The classification task calculates the sentiment polarity probability through the Softmax layer. The final multi-task joint loss is composed of the weighted sum of the sequence labeling loss and the sentiment classification loss, as shown in Formula (6):

$$\mathcal{L} = -\log P(y \mid H) + \lambda \, \mathrm{CE}\left(p(c \mid H), c^*\right) \tag{6}$$

In Formula (6), CE represents the cross-entropy loss; c^* is the true sentiment label; $p(c \mid H)$ is the sentiment polarity probability; λ is the weight balancing coefficient, which is used to adjust the relative influence of the two tasks. This joint modeling structure provides an efficient basis for the precise positioning of improvement points and trend analysis in university teaching feedback.

III. EXPERIMENTAL DESIGN

A. EXPERIMENTAL DATASET

The experiment processes the teaching feedback data and divides it into training, validation, and test sets according to representative proportions to ensure sample balance. The specific division is shown in Table 1.

TABLE 1. Dataset division statistics

Dataset Type	Sample Size	Average Sentence Length	Positive Sentiment Ratio (%)	Improvement Annotation Ratio (%)
Training Set	12000	34.8	52.3	41.6
Validation Set	3000	35.1	51.8	40.9
Test Set	3000	34.5	52.0	42.1

The training set contains 12,000 samples, and the validation set and the test set contain 3,000 samples each, with an average sentence length of 34.5-35.1 words, a positive sentiment ratio of 51.8%-52.3%, and an improvement point annotation ratio of 40.9%-42.1%. This ensures consistent text length, stable sentiment distribution, and sufficient improvement points, providing a balanced and representative data foundation for the model.

B. EXPERIMENTAL ENVIRONMENT

To balance convergence speed, gradient stability, and generalization capability, the experimental system configures key hyperparameters as follows: the learning rate is set to 2×10^{-5}, the batch size is 32, and the AdamW optimizer is employed. Training is conducted for 10 epochs with a Dropout rate of 0.1 and a maximum sequence length of 128. For multi-task joint modeling, the balancing coefficient λ in Formula (6) is set to 0.4. These settings ensure stable and reproducible training.

IV. RESULTS

A. ABLATION EXPERIMENT

1) Validation of the Semantic Representation Module

To systematically evaluate the performance of different semantic representation approaches in the task of improvement point identification, this paper compares the performance of feature representations based on static word embeddings (Word2vec and GloVe), Long Short-Term Memory (LSTM) sequence embeddings, and the BERT pre-trained language model embeddings implemented using the proposed method. In the experiments, using the same dataset and model training configuration, the accuracy, F1 score, and improvement point completeness of each semantic representation on the improvement point identification task are measured using a unified evaluation metric, thereby quantifying the differences in the ability of different semantic modeling approaches to capture textual features. The results of this analysis are shown in Figure 2.

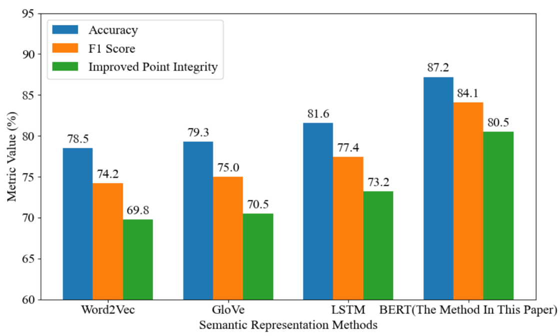

FIGURE 2. Comparison of identification effects of semantic representation methods

Word2Vec and GloVe achieve accuracy of 78.5% and 79.3%, F1-scores of 74.2% and 75.0%, and improvement point completeness of 69.8% and 70.5%, respectively. However, static word embeddings lack contextual information and therefore struggle to capture fine-grained improvement points. LSTM embeddings achieve an improved accuracy of 81.6%, an F1-score of 77.4%, and an improvement point completeness of 73.2%, with its recurrent structure enhancing the representation of sequence features. The integration of BERT embeddings into the proposed framework achieves an accuracy of 87.2%, F1-score of 84.1%, and improvement point completeness of 80.5%, respectively. This demonstrates the advantages of the bidirectional Transformer in modeling contextual semantics and long-term dependencies, and the significant performance advantage of pre-trained deep semantic models in identifying improvement points.

2) Contribution Evaluation of Emotional Attention Mechanism

In this experiment, to evaluate the impact of the emotional attention mechanism on identifying improvement points, a baseline model without the attention mechanism is constructed and compared with a model based on the proposed method, which adds emotional attention weighting during the convolutional feature fusion stage to enhance the representation of features related to improvement points. The experiments are conducted with the same dataset and training configuration. The effect of the attention mechanism on the output of multi-task joint modeling is analyzed by recording the accuracy, F1-Score, and improvement point integrity metrics of each model on the test set, as shown in Figure 3.

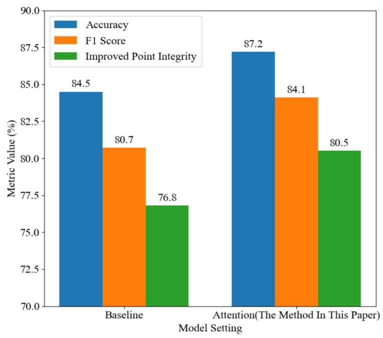

FIGURE 3. Contribution of emotional attention mechanism to improvement point identification

Accuracy of the baseline model is 84.5%, with an F1-score of 80.7% and improvement point completeness of 76.8%. The method proposed in this paper increases accuracy to 87.2%, F1-score to 84.1%, and improvement point completeness to 80.5%. There is a significant improvement in performance due to the attention mechanism's weighting of convolutional features. This enhances the signal related to the improvement point while suppressing irrelevant information, thereby strengthening the model's ability to locate text segments and discern sentiment. This validates the effectiveness of sentiment attention within multi-channel convolutional networks.

3) Synergistic Effect of Joint Modeling Strategies

This experiment compares the performance of single-task and dual-task joint modeling in improvement point identification and sentiment classification. By inputting fused features into the sequence labeling module and the sentiment classifier to form a dual-task output, the results are quantified in terms of accuracy, F1-score, and improvement point completeness compared to the single-task model. The results are shown in Figure 4.

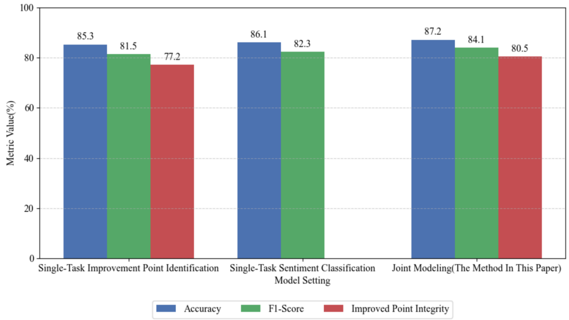

FIGURE 4. Comparison of dual-task joint modeling performance

Single-task improvement point identification achieves an accuracy of 85.3%, an F1 score of 81.5%, and an improvement point completeness of 77.2%. Single-task sentiment classification achieves an accuracy of 86.1% and an F1 score of 82.3%. Dual-task joint modeling achieves improvements to an accuracy of 87.2%, an F1 score of 84.1%, and an improvement point completeness of 80.5%, respectively. By sharing semantic and sentiment features, the capture of relevant patterns in improvement points and joint representation optimization enables more accurate localization, demonstrating the advantages of joint modeling in both improvement point identification and sentiment analysis.

V. CONCLUSIONS

The model for mining improvement points in university teaching feedback constructed in this paper uses BERT embedding as its core, combined with a multi-channel sentiment convolutional network and sentiment attention mechanism, and uses BiLSTM-CRF to locate improvement points and identify sentiment. Experiments show that the model outperforms baselines and single-task models in accuracy (87.2%), F1-Score (84.1%), and improvement point completeness (80.5%). It also enhances stability across different courses and sentiment distributions, enabling fine-grained extraction of student opinions and sentiment quantification, providing a reliable basis for teaching improvement and education quality monitoring.

AUTHOR CONTRIBUTIONS

Conceptualization –Yiming Cai, Zhiteng Zhao, Meiqi Wang and Yi Yin; methodology – Yiming Cai, Zhiteng Zhao and Yi Yin; investigation – Yiming Cai, Zhiteng Zhao, Meiqi Wang and Yi Yin; writing-original draft preparation – Yiming Cai, Zhiteng Zhao, Meiqi Wang and Yi Yin. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] R. A. M. de Kleijn, “Supporting student and teacher feedback literacy: an instructional model for student feedback processes,” Assessment & Evaluation in Higher Education, vol. 48, no. 2, pp. 186-200, 2023, doi: 10.1080/02602938.2021.1967283.
[2] R. Zedan, “Student feedback as a predictor of learning motivation, academic achievement and classroom climate,” Education and self-development, vol. 16, no. 2, pp. 27-46, 2021, doi: 10.26907/esd.16.2.03.
[3] D. Carless and N. Winstone, “Teacher feedback literacy and its interplay with student feedback literacy,” Teaching in higher education, vol. 28, no. 1, pp. 150-163, 2023, doi: 10.1080/13562517.2020.1782372.
[4] A. A. Lipnevich and J. K. Smith, “Student-feedback interaction model: Revised,” Studies in Educational Evaluation, vol. 75, pp. 101208-101218, 2022, doi: 10.1016/j.stueduc.2022.101208.
[5] J. H. Nieminen and D. Carless, “Feedback literacy: A critical review of an emerging concept,” Higher Education, vol. 85, no. 6, pp. 1381-1400, 2023, doi: 10.1007/s10734-022-00895-9.
[6] D. Carless, “From teacher transmission of information to student feedback literacy: Activating the learner role in feedback processes,” Active Learning in Higher Education, vol. 23, no. 2, pp. 143-153, 2022, doi: 10.1177/1469787420945845.
[7] M. A. Adarkwah, “The power of assessment feedback in teaching and learning: a narrative review and synthesis of the literature,” SN Social Sciences, vol. 1, no. 3, pp. 75-83, 2021, doi: 10.1007/s43545-021-00086-w.
[8] Z. Gan, Z. An, and F. Liu, “Teacher feedback practices, student feedback motivation, and feedback behavior: how are they associated with learning outcomes? Frontiers in psychology,” 2021; 12:697045-697055, doi: 10.3389/ fpsyg.2021.697045.
[9] R. Ajjawi, F. Kent, J. Broadbent, et al., “Feedback that works: A realist review of feedback interventions for written tasks,” Studies in Higher Education, vol. 47, no. 7, pp. 1343-1356, 2022, doi: 10.1080/03075079.2021.1894115.
[10] E. Lukwaro, K. Kalegele, and D. Nyambo, “A review on nlp techniques and associated challenges in extracting features from education data,” International Journal of Computing and Digital Systems, vol. 16, no. 1, pp. 962-972, 2024.
[11] M. Liang, “Leveraging natural language processing for automated assessment and feedback production in virtual education settings,” Journal of Computational Methods in Sciences and Engineering, vol. 25, no. 3, pp. 2502-2515, 2025, doi: 10.1177/14727978251314556.
[12] F. Qi, Y. Gao, M. Wang, T. Jiang, and Z. Li, “Data mining of online teaching evaluation based on deep learning,” Mathematics, vol. 12, no. 17, pp. 2692-2702, 2024, doi: 10.3390/math12172692.
[13] R. Nawaz, Q. Sun, M. Shardlow, et al., “Leveraging AI and machine learning for national student survey: actionable insights from textual feedback to enhance quality of teaching and learning in UK’s higher education,” Applied Sciences, vol. 12, no. 1, pp. 514-524, 2022, doi: 10.3390/app12010514.
[14] Sikandar Seemab, Muhammad Saqib Zaigham, Aisha Bhatti, and Hashim Khan and Uzma Arshad Mughal, “Sentiment analysis of student feedback in higher education using Natural Language Processing (NLP),” Contemporary Journal of Social Science Review, vol. 2, no. 04, pp. 1741-1750, 2024.
[15] R. Harrison, L. Meyer, P. Rawstorne, et al., “Evaluating and enhancing quality in higher education teaching practice: A meta-review,” Studies in Higher Education, vol. 47, no. 1, pp. 80-96, 2022, doi: 10.1080/03075079.2020.1730315.
[16] T. Shaik, X. Tao, Y. Li, et al., “A review of the trends and challenges in adopting natural language processing methods for education feedback analysis,” Ieee Access, vol. 10, pp. 56720-56739, 2022, doi: 10.1109/ACCESS.2022.3177752.
[17] A. Botelho, S. Baral, J. A. Erickson, P. Benachamardi, and N. T. Heffernan, “Leveraging natural language processing to support automated assessment and feedback for student open responses in mathematics,” Journal of computer assisted learning, vol. 39, no. 3, pp. 823-840, 2023, doi: 10.1111/jcal.12793.
[18] Z. Kastrati, F. Dalipi, A. S. Imran, et al., “Sentiment analysis of students’ feedback with NLP and deep learning: A systematic mapping

study," Applied Sciences, vol. 11, no. 9, pp. 3986-3996, 2021, doi: 10.3390/app11093986.

[19] K. Sangeetha and D. Prabha, "Understand Students Feedback Using Bi-Integrated CRF Model Based Target Extraction," Computer Systems Science & Engineering, vol. 40, no. 2, pp. 735-745, 2022, doi: 10.32604/csse.2022.019310.

[20] C. Dervenis, G. Kanakis, and P. Fitsilis, "Sentiment analysis of student feedback: A comparative study employing lexicon and machine learning techniques," Studies in Educational Evaluation, vol. 83, pp. 101406-101416, 2024, doi: 10.1016/ j.stueduc.2024.101406.

[21] K. Okoye, A. Arrona-Palacios, C. Camacho-Zuñiga, et al., "Towards teaching analytics: a contextual model for analysis of students' evaluation of teaching through text mining and machine learning classification," Education and Information Technologies, vol. 27, no. 3, pp. 3891-3933, 2022, doi: 10.1007/s10639-021-10751-5.

[22] Tracii Ryan, Michael Henderson, and Kris Ryan and Gregor Kennedy, "Feedback in higher education: Aligning academic intent and student sensemaking," Teaching in Higher Education, vol. 29, no. 4, pp. 860-875, 2024, doi: 10.1080/13562517.2022.2029394.

[23] M. Alzaid and F. Fkih, "Sentiment analysis of students' feedback on e-learning using a hybrid fuzzy model," Applied Sciences, vol. 13, no. 23, pp. 12956-12966, 2023, doi: 10.3390/app132312956.

[24] C. Dann, P. Redmond, M. Fanshawe, et al., "Making sense of student feedback and engagement using artificial intelligence," Australasian Journal of Educational Technology, vol. 40, no. 3, pp. 58-76, 2024, doi: 10.14742/ajet.8903.

[25] C. R. Drue and C. A. Bifulco, "Beyond the numbers: How directors and chairs interpret student feedback to equitably evaluate teaching," Journal of Academic Ethics, vol. 23, pp. 1499-1521, 2025, doi: 10.1007/s10805-025-09611-5.

[26] Z. Liu, C. Wen, Z. Su, et al., "Emotion-semantic-aware dual contrastive learning for epistemic emotion identification of learner-generated reviews in MOOCs," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 11, pp. 16464-16477, 2023, doi: 10.1109/TNNLS.2023.3294636.

[27] A. S. Sunar and M. S. Khalid, "Natural language processing of student's feedback to instructors: A systematic review," IEEE Transactions on Learning Technologies, vol. 17, pp. 741-753, 2023, doi: 10.1109/TLT.2023.3330531.

[28] K. Masood, M. A. Khan, U. Saeed, et al., "Semantic analysis to identify students' feedback," The Computer Journal, vol. 65, no. 4, pp. 918-925, 2022, doi: 10.1093/comjnl/bxaa130.

[29] P. Ren, L. Yang, and F. Luo, "Automatic scoring of student feedback for teaching evaluation based on aspect-level sentiment analysis," Education and information technologies, vol. 28, no. 1, pp. 797-814, 2023, doi: 10.1007/s10639-022-11151-z.

[30] D. Jatain, M. Niranjanamurthy, and P. Dayananda, "A hybrid bio-inspired fuzzy feature selection approach for opinion mining of learner comments," SN Computer Science, vol. 5, no. 1, pp. 135-145, 2024, doi: 10.1007/s42979-023-02526-1.

[31] X. Ma, "English Teaching in Artificial Intelligence-based Higher Vocational Education Using Machine Learning Techniques for Students' Feedback Analysis and Course Selection Recommendation," Journal of Universal Computer Science (JUCS), vol. 28, no. 9, pp. 898-908, 2022, doi: 10.3897/jucs.94160.

[32] Asad Abdi, Gayane Sedrakyan, Bernard Veldkamp, and Jos van Hillegersberg and Stéphanie M van den Berg, "Students feedback analysis model using deep learning-based method and linguistic knowledge for intelligent educational systems," Soft Computing, vol. 27, no. 19, pp. 14073-14094, 2023, doi: 10.1007/s00500-023-07926-2.

[33] I. Zyout and M. Zyout, "Sentiment analysis of student feedback using attention-based RNN and transformer embedding," Int J Artif Intell, vol. 13, no. 2, pp. 2173-2184, 2024, doi: 10.11591/ijai.v13.i2.pp2173-2184.

[34] G. Suthakar A R and Sandesh B J .Enhanced image captioning with positional and dual attention using deep convolution long short-term memory and emotional feedback mechanism, "Multimedia Tools and Applications," 2026; 85(3). doi : 10.1007/s11042-026-21394-4.

[35] H. Shuqin and C. Raga R, "A deep learning model for student sentiment analysis on course reviews," IEEE Access, vol. 12, pp. 136747-136758, 2024, doi: 10.1109/ACCESS.2024.3463793.