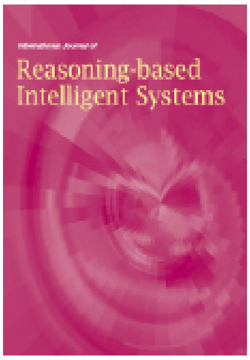


International Journal of Reasoning-based Intelligent Systems

ISSN online: 1755-0564 - ISSN print: 1755-0556
https://www.inderscience.com/ijris

Evaluation of teaching quality in accounting smart education classrooms driven by student expression feature recognition

Shu Chen

DOI: 10.1504/IJRIS.2025.10072264

Article History:

Received:	06 May 2025
Last revised:	31 May 2025
Accepted:	31 May 2025
Published online:	24 July 2025

Evaluation of teaching quality in accounting smart education classrooms driven by student expression feature recognition

Shu Chen

College of Accounting,
Zhanjiang University of Science and Technology,
Zhanjiang 524094, China
Email: chenshu3088@163.com

Abstract: The extensive use of artificial intelligence is making traditional quality evaluation methods progressively unfit for modern education's personalisation, intelligence, and real-time needs. Especially in accounting, students' expressiveness level and classroom participation influence their learning. For accounting smart education, student expression feature recognition technology is applied to offer a classroom teaching quality rating system. Two tests of the system were carried out in this work. First employing a confusion matrix, the model effectively detects the seven basic emotional states. The second experiment tracked participants' expression variations over a designated period and matched the manually annotated results with the outcomes of the expression recognition system. All experimental results show that the framework effectively and consistently detects student's expression states and catches dynamic classroom emotions, so optimising classroom teaching and learning. This paper presents fresh quality evaluation techniques in smart education and supports smart accounting education.

Keywords: expression feature recognition; convolutional neural networks; CNN; accounting smart education; classroom teaching quality evaluation.

Reference to this paper should be made as follows: Chen, S. (2025) 'Evaluation of teaching quality in accounting smart education classrooms driven by student expression feature recognition', *Int. J. Reasoning-based Intelligent Systems*, Vol. 17, No. 9, pp.1–11.

Biographical notes: Shu Chen received her Master's degree from the Hunan University in 2014. She is currently an Associate Professor at the Zhanjiang University of Science and Technology. Her research focuses on machine learning, accounting management and taxation studies.

1 Introduction

1.1 Research background

As big data find extensive use in many different sectors, the education sector is bringing smart improvements. Especially in the field of higher education, traditional teaching approaches and quality evaluation methods are progressively difficult to meet the needs of modern education for personalisation, intelligence, and real-time. Accounting is a theoretical and practical field that demands students to have great practical skill in addition to mastery of professional information (Giang, 2024). Often neglecting students' expressions and emotional reactions, which is challenging to fully represent the learning state, traditional accounting teaching quality evaluation mostly depends on students' performance and teachers' assessment (Jill et al., 2019).

As a significant use of artificial intelligence, expression recognition technology has drawn much interest in the field of education recently. Expressions naturally capture pupils' emotional fluctuations, classroom interest and attention (Mehta et al., 2022). Thus, expression recognition technology offers a fresh viewpoint for teaching quality evaluation, particularly in accounting education, where the variation of students' expressions directly influences the learning impact. By means of expression recognition, educators can get real-time expression feedback from their students, modify their instructional plans in response, and raise classroom interaction and teaching quality.

The emergence of smart education presents fresh possibilities and difficulties for conventional assessment of teaching effectiveness. Academic performance and teacher evaluation constitute the basis of the conventional assessment method, which makes it challenging to completely represent students' actual emotions and expression variances (Tang et al., 2025). By real-time analysis of students' facial expressions, expression recognition technology offers a more objective and thorough evaluation of instructional quality. This not only helps to offset the inadequacies of conventional approaches but also gives teachers real-time comments, which can maximise the material and raise the success of classroom instruction.

Based on student expressions, the system for evaluating classroom teaching quality, driven by recognition technology and smart education, can overcome the limitations of traditional evaluation methods. It accurately reflects students' learning status and changes in expression in real-time, thereby enhancing teaching quality. This study aims to explore how expression recognition technology can facilitate the dynamic optimisation of teaching quality in accounting education and provide innovative concepts for the advancement of smart education.

1.2 Related work

Expression recognition technology has great potential in several application scenarios, especially in the field of education, and has progressively become an important tool for enhancing teaching quality evaluation and learning effect. As a major research direction in the field of artificial intelligence. Early studies on expression identification mostly concentrated on face feature extraction methods based on conventional machine learning algorithms, like analyses of facial expressions through classification algorithms such support vector machine (SVM) and K nearest neighbour (KNN). These conventional techniques are able to make initial assessments on pupils' expressions under specific circumstances by use of facial feature points, local binary pattern (LBP), and other approaches (Kumar et al., 2023).

Deep learning methods, especially the extensive use of convolutional neural networks (CNN) and recurrent neural networks (RNN), have greatly raised the accuracy of expression recognition (Ahmed et al., 2023). Overcoming the limits of conventional manual feature extraction techniques, deep learning models can automatically learn multi-level features from a vast volume of visual data, so greatly boosting the accuracy and stability of expression identification. This lets expression recognition technology not only effectively examine students' emotional changes but also manage challenging classroom settings and offer real-time comments. When confronted with complicated facial expressions, varying lighting conditions, and occlusion, these techniques show poor resilience and accuracy, nevertheless.

Expression recognition technology is progressively being used in the field of education toward expression monitoring and classroom behaviour analysis. Teachers can better modify their instruction by knowing students' expressive reactions in the classroom, that is, interest, bewilderment, tiredness, and other emotional states, via real-time capture of their facial expressions. Enhancing classroom interaction and student involvement has proven notable benefits from this real-time change grounded on expression feedback (Al-Mansouri, 2024). Particularly in the context of smart education, expression-based teaching quality assessment has increasingly garnered significant interest in research. This approach considers students' emotional changes in real-time as a vital element of teaching quality assessment, thereby overcoming the limitations of traditional evaluation systems that rely heavily on students' performance.

Apart from expressiveness studies, studies on teaching quality assessment are progressively leaning toward multidimensionality and multilayer. Although they can somewhat reflect the teaching effect, traditional approaches of evaluating teaching quality mostly depend on students' grades, instructors' evaluation and questionnaires, which usually lack real-time and comprehensiveness. Many studies have just started to investigate how to create a dynamic and complete teaching quality assessment model by aggregating students' expressions, behavioural data, and teaching interaction data (Kokoç et al., 2021). These new evaluation tools may evaluate teaching quality more holistically and precisely by including expression analysis, learning behaviour and teaching feedback, therefore enabling teachers to promptly change their classroom materials and approaches.

In the realm of accounting education, the application of expression recognition technology offers significant advantages. The accounting field is inherently operational and practical; thus, students' engagement and emotional expressions in the classroom greatly impact their learning outcomes. By utilising expression recognition technology, teachers can receive real-time feedback on students' emotional states and adjust their instructional methods accordingly, thereby enhancing the quality of their teaching. Consequently, there has been a growing body of research focused on integrating expression recognition technology with classroom teaching quality evaluation systems. This integration leverages students' emotional data to inform teaching adjustments, ultimately aiming to achieve personalised and intelligent improvements in teaching quality.

Although the use of expression recognition technology in education is still at the stage of ongoing research and development overall, its possibility to improve the accuracy, real-time and comprehensiveness of the teaching quality evaluation system has been extensively acknowledged. Expression recognition technology will become more and more relevant in smart education and teaching quality assessment as technology develops, particularly in fields with great practicality like accounting education, which can give teachers more accurate classroom feedback and so help to improve the teaching quality.

This work has as its innovative points following:

1. Experimental validation and confusion matrix analysis: high accuracy of the framework in identifying the seven fundamental emotional states is shown by experimental validation employing the CK+ face expression dataset. The performance of the model is validated by the confusion matrix analysis, which also offers trustworthy data support for evaluation of teaching quality.

2. Constructing an evaluation system for accounting teaching quality based on expression feature recognition: expression recognition technology was initially applied in the field of intelligent accounting education, leading to the development of a new evaluation framework for classroom teaching quality. This framework transcends the limitations of traditional

reliance on students' grades and teachers' subjective assessments, providing a novel approach and methodology for the dynamic optimisation of accounting education quality.

3 Collaborative work of five modules: realising the intelligence and personalisation of the teaching process, five modules work together to create a closed-loop system capable of real-time student expression data collecting, dynamic change of teaching strategies, and thorough evaluation of teaching quality.

4 Multi-dimensional teaching quality assessment model: a multi-dimensional and dynamic teaching quality assessment model is developed by combining students' expression status, learning behaviour characteristics, and classroom interaction data, so obtaining comprehensive and real-time teaching quality assessment.

2 Expression recognition technology

Using computer vision, expression recognition is a method that automatically detects and groups emotional elements from face photographs. Its main responsibility is to evaluate a person's emotional condition by means of facial expressions. Expression recognition technology has been extensively applied in many disciplines, mostly in the field of education, which can help teachers to understand the emotional changes and learning state of students in real-time. Facial expression is a direct reflection of emotional changes and can provide valuable information about an individual's psychological state.

Figure 1 LBP feature mapping

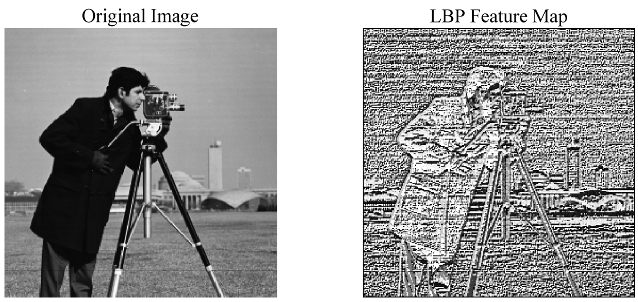
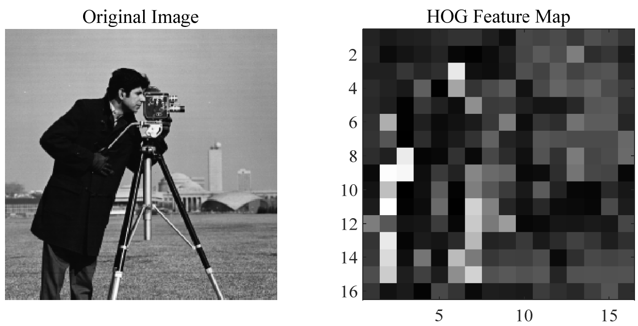


Figure 2 HOG characterisation map



Feature extraction is a necessary phase of expression recognition. One traditional approach of feature extraction is LBP. LBP computed for a given pixel position point $I(x, y)$ in a picture is:

$$LBP(x, y) = \sum_{p=0}^{P-1} s(I(p) - I(x, y)) \cdot 2^p \tag{1}$$

where s is a sign function with 0 otherwise and 1 when the neighbourhood pixel value $I(p)$ is higher than the centre pixel value $I(x, y)$. This allows the LBP to encode grey scale information in a particular area of an image to generate a binary value reflecting the texture properties of that area (Kaddar et al., 2018), see Figure 1.

Apart from LBP, histogram of oriented gradient (HOG) is another often utilised feature extraction technique. See Figure 2 for HOG forms that calculate the gradient direction and magnitude for every small region in the image, therefore characterising the local form of the image (Mukherjee et al., 2024).

One may write the computation of HOG characteristics as:

$$HOG(R) = \sum_{i=1}^{n} \left\| \nabla I(x, y) \right\| \cdot \theta \tag{2}$$

where $\|\nabla I(x, y)\|$ is the gradient magnitude at a given position in the image; θ is the direction of the gradient; n is the histogram's bucket count; $\nabla I(x, y)$ denotes the gradient at that moment.

CNNs have transformed methods of expression recognition recently. CNNs can automatically learn multi-level characteristics of an image unlike conventional hand feature extraction techniques. Image feature extraction in CNN mostly depends on the convolution technique (Liu et al., 2021). The convolution operation can be written given an input picture I and a convolution kernel K as:

$$F = I * K \tag{3}$$

where * represents the convolution operation and F is the feature map produced following it. Sliding the convolution kernel helps the convolution procedure to extract the local image features, hence, filtering the image.

CNN's training procedure depends much on the pooling layer as well. Reducing the spatial dimensionality of the image by the pooling layer helps to maintain the most representative features by so lowering the computation required (Nirthika et al., 2022). Maximum pooling and average pooling are two popular forms of pooling procedures whereby maximum pooling chooses the maximum value in every small area and average pooling compiles the average value in the area. One may represent the pooling process as:

$$P = Pool(F) \tag{4}$$

where P is the combined feature map.

Usually, the fully connected layer of the CNN classifies the output once the convolution and pooling layers finish the feature extracting. Combining the acquired information, the fully connected layer finds the sentiment category the input image fits. Softmax is a usually used activation function to convert the network output scores into a probability distribution (Lin and Shen, 2018). The softmax function expressed is:

$$Y_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{5}$$

where Y_i represents the likelihood of the i^{th} emotion; z_i is the network's output score; j is the index of every emotion category. The softmax function generates normalised probability from the score, outputs the probability of every emotional category, and finally decides which most likely emotional category the input image has.

Apart from CNN, RNN and its variants such gated recurrent units (GRU) and long short-term memory networks (LSTM) are also applied in expression recognition (Hindarto, 2023). Given an input sequence of facial images X in RNN, the hidden state h_t of the network is constantly updated over time and its updating mechanism can be stated as:

$$X = \{x_1, x_2, ..., x_t\} \tag{6}$$

$$h_t = f\left(Wx_t + Uh_{t-1} + b\right) \tag{7}$$

where h_t as the hidden state at time t, W and U are the weight matrices; f is the activation function; b is the bias factor. Particularly appropriate for handling the challenge of expression recognition in films, the RNN is able to capture the dynamic changes of face expressions in time series by means of such a recursive mechanism.

RNN suffers in gradient vanishing or gradient explosion in the learning of long sequences even if it performs well in temporal data processing. LSTM network uses many gating methods to efficiently manage the input flow and guarantee that the model can learn the long-term relationships, so solving this problem. LSTM has a state update formula as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{8}$$

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{9}$$

$$o_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \tag{10}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{11}$$

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{12}$$

LSTM is able to choose which information needs to be kept and which may be forgotten by including input gates, forgetting gates and output gates, therefore essentially avoiding the gradient vanishing issue in conventional RNNs and learning face expression features over a long time period.

Expression recognition systems have essentially changed from early manual feature extraction approaches to end-to-end learning employing deep learning models. Expression recognition technology holds great promise for use in the field of education, particularly in the assessment of teaching quality in accounting education, given the ongoing development and optimisation of these technologies.

3 Construction and implementation of accounting wisdom classroom teaching quality evaluation framework

Driven by real-time student expression data collecting to support the dynamic adjustment of teaching decisions, expression recognition technology forms the fundamental building idea of the accounting smart classroom teaching quality evaluation framework. Figure 3 displays the framework:

3.1 Student expression data collection module

First, greyscaling transforms the student's facial image taken on camera into a greyscale image $I \in \mathbb{R}^{H \times W}$ whereby H and W respectively indicate the height and width of the image. After that, a facial detection algorithm removes the facial region; this process can be shown by a rectangle box where $R(x, y, w, h)$ has upper left corner coordinates (x, y) and width and height of the facial region respectively (Gao and Yang, 2022).

The picture is then feature-extracting CNN passed. Assuming I as the input to the convolutional layer, K as the convolutional kernel, and F as the output feature map, the convolution procedure can be stated as follows:

$$F_{i,j} = (I * K)_{i,j} = \sum_m \sum_n I_{i+m, j+n} K_{m,n} \tag{13}$$

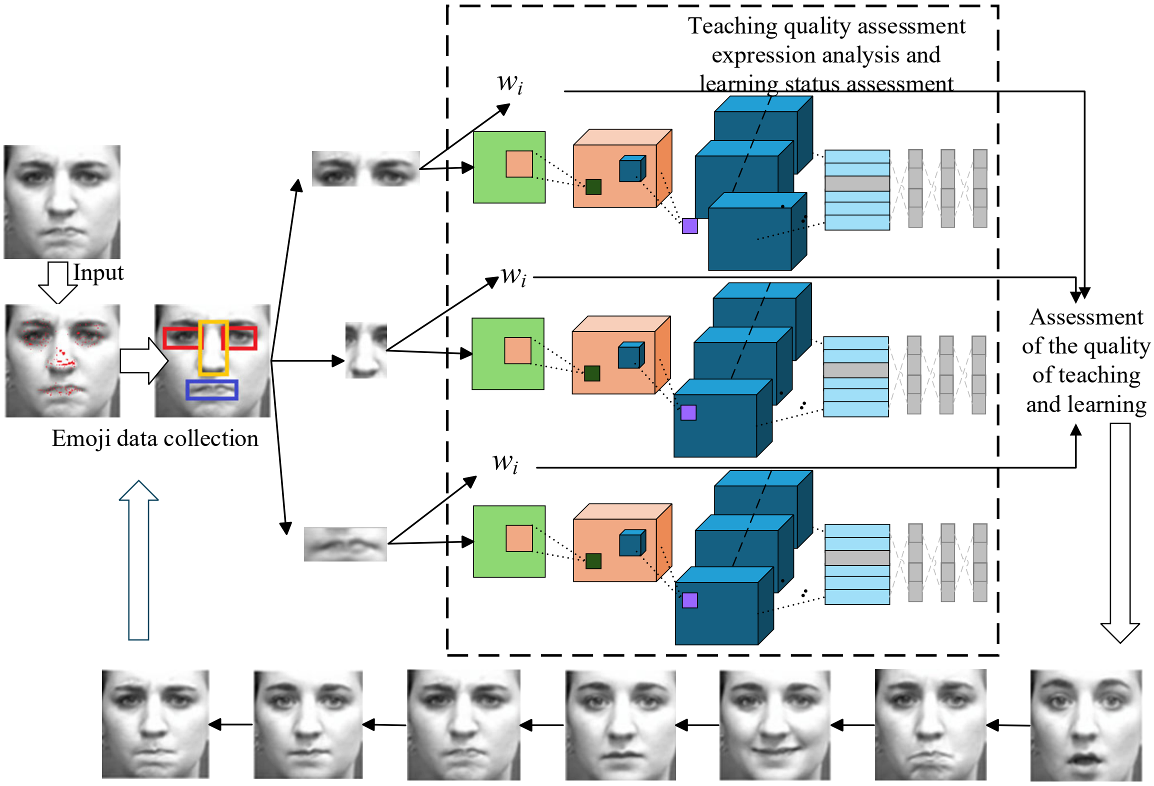
where $F_{i,j}$ is the feature map derived following convolution; $I_{i+m,j+n}$ is the local picture area; $K_{m,n}$ is the convolution kernel element. By means of local weighting and summation, the convolution operation generates the local features of the image, thereby extracting the feature information of face expression changes.

The feature map F is downsampled throughout the pooling layer and the maximum pooling operation is used to extract the maximum value of every local region as a representative of that region, hence lowering the amount of computation. One may define the pooling operation by the following equation:

$$P_{i,j} = \max\left(F_{i+m, j+n}\right), \quad \text{for } m, n \in \{0, 1, 2\} \tag{14}$$

where $P_{i,j}$ is the combined feature map element denoting the highest feature of every local area.

Figure 3 Framework for evaluating teaching quality in accounting smart classroom (see online version for colours)



Following feature extraction, the face features travel via a fully linked layer and softmax function for expression categorisation (Tang et al., 2021). The softmax function is obtained assuming the network output is z and k is the number of expression categories by:

$$P(y_i \mid X) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}} \tag{15}$$

where $\sum_{j=1}^{k} e^{z_j}$ is the exponential sum of the scores of all the categories, therefore assuring that the output probability adds to 1. $P(y_i| X)$ is the likelihood that the input feature X matches the expression category y_i, z_i is the score of the associated category.

In the process of data collection and storage, we strictly comply with ethical standards and relevant laws and regulations to ensure that all data is collected and used with the explicit consent of the participants, and that anonymisation and secure storage measures are taken. The cognitive state of the pupils is ultimately evaluated using the outcomes of expression categorisation (e.g., confused, focused, anxious, etc.). These findings are then given back to the classroom instructional system to enable the teacher to modify the teaching approaches.

3.2 *Module for the analysis of expression and assessment of learning status*

This module mostly relies on the recognition results of students' facial expressions mixed with learning behavioural traits to dynamically analyse the students' expression status, and then evaluate their learning development.

First, the expression analysis module generates numerical expression scores from the given student expression labels. Let the expression of a student at a given moment be $y_i \in \{1, 2, \ldots, m\}$, where m indicates the number of several expression types. A weighting function w_i can help one to depict the score of an expression category.

$$S_{\text{emotion}} = \sum_{i=1}^{m} w_i \cdot y_i \tag{16}$$

where S_{emotion} represents the composite score of the student's present expression state; w_i is the weight of the expression category y_i, therefore showing the degree of effect of the expression on the learning state.

The evaluation of learning status then rests on the combined computation of expression score S_{emotion} and students' behavioural traits including engagement (Chu et al., 2020). The learning state S_{learning} can be computed assuming students' classroom P_{class}, (e.g., answering questions, raising queries, etc.) is expressed as a numerical value:

$$S_{\text{learning}} = \alpha \cdot S_{\text{emotion}} + \beta \cdot P_{\text{class}} \tag{17}$$

Combining expressive feedback with learning activities, the formula more fully reflects students' learning situation in the accounting classroom.

Eventually, the system assesses whether students are in different learning states, such as active learning, confusion, or disengagement. This evaluation provides timely feedback to the teacher, enabling them to dynamically adjust their teaching strategies and optimise the effectiveness of classroom instruction. The system then classifies the students' learning status according to the value of S_{learning}.

3.3 Module for adapting teaching content and interaction

To raise the quality of instruction, the module changes the classroom content and interaction strategies in real-time based on the students' expression level, learning development, and established teaching goals. First, the students' adaptation score S_{adapt} is computed using the following formula using their expression scores and learning status:

$$S_{\text{adapt}} = \gamma \cdot S_{\text{learning}} + \delta \cdot S_{\text{previous}} \tag{18}$$

where S_{learning} is the student's present learning state; S_{previous} is the student's past learning state; γ and δ are weighting coefficients, therefore representing the influence weights of the current learning state and the historical state.

Subsequently, the system dynamically changes the manner of interaction and the difficulty of the instructional materials based on the adaption scores of the pupils (Tetzlaff et al., 2021). Assuming that D_{content} reflects the present difficulty of the teaching content, the system will modify it depending on the students' adaptability; the modified difficulty of the teaching material D_{adjusted} can be stated by the following formula:

$$D_{\text{adjusted}} = D_{\text{content}} \cdot \left(1 - \alpha \cdot \left(S_{\text{adapt}} - 0.5\right)\right) \tag{19}$$

where 0.5 denotes the neutral value of adaptation and S_{adapt} spans [0, 1]. When students' adaptability is high, the difficulty is adequately raised; vice versa, it is reduced, thereby guaranteeing that the complexity of the instructional content is appropriate for their present learning state.

Moreover, the key to raise classroom engagement and learning effect is the modification of teaching interaction. Teachers might increase students' enthusiasm in learning by suitable interactive modification assuming P_{class} as their present interactive involvement. One can describe the interaction adjustment technique $S_{\text{interaction}}$ by means of the following formula:

$$S_{\text{interaction}} = \lambda \cdot P_{\text{class}} + \theta \cdot \left(1 - S_{\text{adapt}}\right) \tag{20}$$

The former shows the effect of classroom involvement on the interaction approach where λ and θ are the weighting coefficients; the later represents the influence of learning adaptation.

The system automatically modifies the interaction method, (e.g., enhance the Q&A session, interactive group discussion, etc.) in response to low engagement or a confused expression of the students so enhancing their classroom engagement and learning interest.

Eventually, the system dynamically changes the classroom teaching approaches in real-time based on the modified teaching content and interactive strategies to guarantee that students can be in the best possible learning state in the accounting classroom, so enhancing the teaching quality and students' learning effect.

3.4 Teaching quality assessment module

Based on student expression level, learning status, and classroom interaction feedback, the module evaluates the whole efficacy of classroom teaching. It generates a teaching quality score by means of data analysis and processing from every module, therefore offering teachers with recommendations for optimisation.

Assuming that each student's expression score and learning status information affects teaching quality, the module first computes the overall teaching quality score, S_{quality}, by aggregating these data. Weighing averages helps one to represent the general teaching quality score:

$$S_{\text{quality}} = \alpha \cdot \frac{1}{N} \sum_{i=1}^{N} S_{\text{emotion}}^{(i)} + \beta \cdot \frac{1}{N} \sum_{i=1}^{N} S_{\text{learning}}^{(i)} \tag{21}$$

where N is the total number of pupils in the class; $S_{\text{emotion}}^{(i)}$ and $S_{\text{learning}}^{(i)}$ respectively indicate the expression score and learning status score of the i^{th} student. By means of this formula, the module may evaluate the expression fluctuation and learning situation of the entire classroom, so obtaining the general teaching quality score.

Apart from expression and learning environment, the frequency and nature of classroom interactions greatly affect the teaching quality (Xiao et al., 2023). Assuming I_{freq} as the frequency of interaction and I_{quality} as the quality of interaction score, the system aggregates these two measures using the following formula:

$$S_{\text{interaction}} = \gamma \cdot I_{\text{freq}} + \delta \cdot I_{\text{quality}} \tag{22}$$

The algorithm will score more highly for teaching quality when classroom interactions are high frequency and quality.

Ultimately, the system computes the final teaching quality score S_{final} by considering expressive feedback, learning state, classroom interactions, and is applied as a basis for assessing the teaching efficacy and so guiding teachers to modify their teaching tactics. The ultimate formula for teaching quality score is:

$$S_{\text{final}} = \lambda \cdot S_{\text{quality}} + \mu \cdot S_{\text{interaction}} \tag{23}$$

where λ and μ are adjustment coefficients expressing the relative influence of interaction quality and instructional quality on the ultimate ranking. This algorithm allows the system to create a complete teaching quality score for every class, therefore enabling teachers to evaluate the classroom's efficiency from an objective standpoint and create a basis for ongoing development.

3.5 Data storage and analysis module

To guarantee data dependability and real-time analysis, the module on data storage and analysis stores, manages, and analyses all sorts of data in the course of instruction and learning. The database table T_{data} stores all expression, learning status, and interaction data in indexed form under timestamp t:

$$T_{\text{data}} = \{(t, S_{\text{emotion}}, S_{\text{learning}}, I_{\text{interaction}})\} \tag{24}$$

where t is a timestamp marking the data collecting moment.

Following data storage, the module examines past performance to forecast student learning direction. Assuming X_t is a vector of student data at a given point, trend prediction is achieved by linear regression model analysis. Let $S_{\text{learning}}(t_0)$ be the learning state at the moment t_0. The linear regression equation allows one to obtain the expected value $S_{\text{learning}}(t_0 + \Delta t)$:

$$X_t = [S_{\text{emotion}}(t), S_{\text{learning}}(t)] \tag{25}$$

$$\begin{aligned} S_{\text{learning}}(t_0 + \Delta t) = \theta_0 + \theta_1 \cdot S_{\text{emotion}}(t_0) \\ + \theta_2 \cdot S_{\text{learning}}(t_0) \end{aligned} \tag{26}$$

where derived from training on past data, θ_0, θ_1, θ_2 are regression coefficients.

The module also detects anomalies to show changes in expression during the learning process. The distance-based anomaly identification technique outlined as follows allows one to compute the anomaly value X_{outlier} assuming historical data of N students:

$$X_{\text{outlier}} = \{X_i \mid \|X_i - \mu\| > k \cdot \sigma\} \tag{27}$$

where μ is the mean of the data collection, σ is the standard deviation, and k is a threshold defining that points greater than k times the standard deviation from the mean are regarded outliers.

By means of these data storage and analysis techniques, the module enables teachers to efficiently store and analyse instructional data, reflect students' learning status and expression variations in real-time, and assist in prompt identification and solution of learning challenges.

4 Experiment and result analysis

4.1 *Experimental data source*

This work selected for model training and evaluation testing the CK+ face expression dataset. Expanding on the Cohn-Kanade dataset, which comprises 327 emotion sequences spanning seven basic expression categories: angry, contemptuous, disgusted, afraid, glad, sad and startled, yields the CK+ dataset (Dalvi et al., 2021). See Figure 4 for rich graphic data these expression categories offer for examining students' emotional reactions in the accounting classroom.

Every picture in the collection was taken under settings tailored to an experiment to avoid obstruction, lighting, and location changes influencing expression detection. This qualifies the CK+ dataset ideal for developing and testing real-time pupil expression analysis techniques.

Although the CK+ dataset was not specifically designed for educational or classroom settings, its rich expression categories and high-quality images made it an effective tool for testing our framework. However, because the data were collected under control, non-classroom conditions, this may limit the generalisability of the results. Future research is needed to validate the generalisability of the model in a real classroom setting.

4.2 *Experimental environment setup*

As shown in Table 1, the experiments were carried out in an environment with high-performance computing capabilities, configured with a high-performance central processor, sufficient memory, and an advanced graphics processing unit, so guaranteeing the high efficiency of the training and testing of the accounting smart classroom teaching quality assessment framework.

Table 1 Experimental environment

Name	*Configuration*
CPU	i7-9700K
RAM	64G
Graphics card	GeForce RTX 3080Ti
Deep learning framework	Pytorch
Programming language	Python
CUDA version	10.1

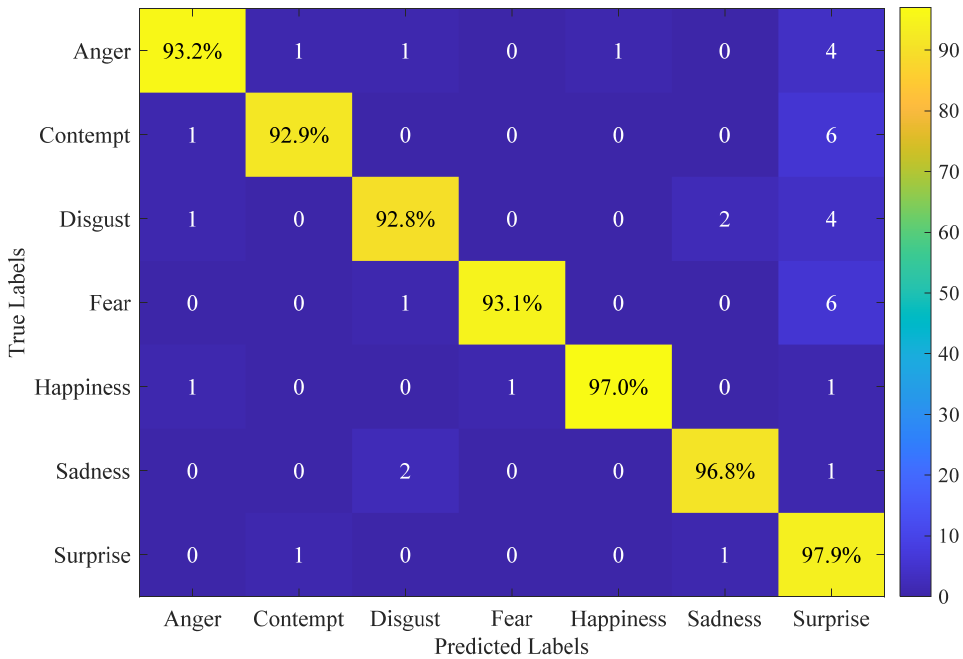
The iteration period of the CK+ dataset training was set to 70, the random cropping size was 44 × 44, Adam was used as the optimisation function, the initial learning rates were all 0.001, and the cosine annealing learning rate decay strategy was used also in order to exclude the influence of chance on the experimental results, the validation was carried out using the ten-fold cross-valuation method for the CK+ dataset (Polovnikov et al., 2021). Experimental results show that the system is able to operate efficiently in real-time feedback with an average processing time of 0.1 seconds and a delay of less than 0.05 seconds.

4.3 *Experimental procedure*

This study firstly trains and evaluates the accounting smart classroom teaching quality evaluation framework by using the CK+ face expression dataset, which not only guarantees the model's capacity to identify the stationary expression images, but also provides a comprehensive confusion matrix (see Figure 5) to shows the model's performance on different emotional categories (Krstinić et al., 2020). This process helps verify the accuracy and reliability of the model in recognising the seven basic emotional states of students: anger, contempt, disgust, fear, happiness, sadness, and surprise. This validation procedure ensures that the framework can provide teachers with accurate feedback on student expressions, thereby enabling the dynamic adjustment of teaching strategies and offering robust data support for the subsequent evaluation of teaching quality.

Figure 4 Example of face expression dataset

Figure 5 Confusion matrix for emotion recognition (see online version for colours)

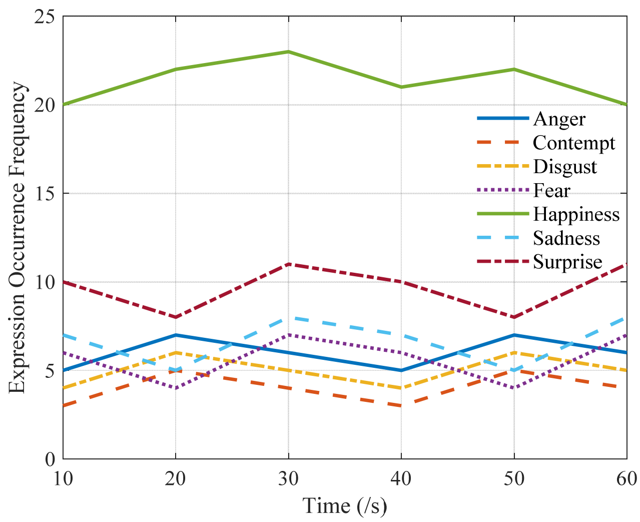


With regard to two emotions (contempt and disgust), the framework attained over 93% accuracy, suggesting its great ability in identifying these unpleasant emotions with comparable facial traits. The uniqueness of the anger and fear expressions may be the reason the identification accuracy for these emotions was somewhat greater than 93%. This helps the framework to recognise and record them easier. These results highlight even more the excellent performance of the framework in identifying positive and powerful emotions since the recognition accuracy for the happiness emotion was 97.0% while the recognition accuracy for the sadness and surprise emotions was as high as 96.8% and 97.9%, respectively.

In general, the framework proved great accuracy and dependability in identifying the seven fundamental emotional states. The confusion matrix's diagonal elements show the number of accurately categorised samples; their high accuracy values show that, generally, the framework can identify phrases. The non-diagonal elements show the number of misclassified samples, and their rather low values highlight the framework's robustness even further.

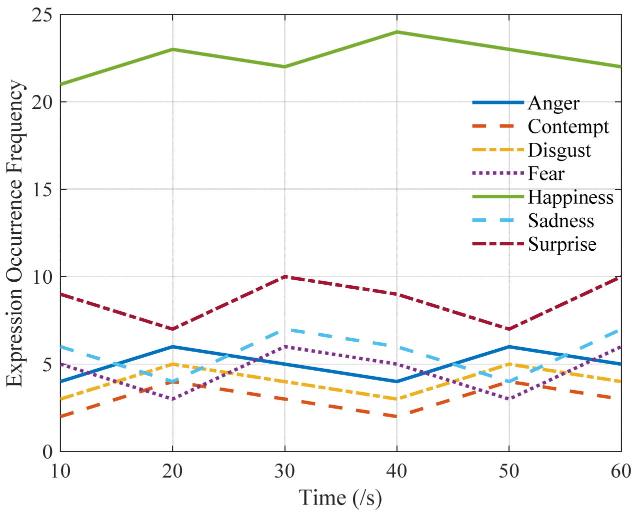
Furthermore, the experimental results demonstrate great accuracy and reliability in identifying students' emotional states. The proposed framework for assessing the quality of teaching and learning in the smart accounting classroom relies on obtaining students' expressive feedback in real-time, adjusting teaching strategies accordingly, and enhancing classroom interaction and overall teaching quality. By utilising expression recognition technology, teachers can gain a better understanding of their students' emotional changes, allowing for more targeted instructional adjustments that improve the overall effectiveness of their teaching.

Figure 6 Distribution curve of recognition results of expression recognition framework (see online version for colours)



This study utilises image sequences from a dataset as a data source to capture changes in participants' expressions over a specified time period, thereby facilitating a more comprehensive evaluation of the framework's performance in dynamic environments. Two groups were formed from the data: one group was produced by the expression recognition framework (results of recognition framework) and the other group by hand annotation (manual recognition results) (Li and Deng, 2019). To guarantee the continuity and integrity of the data, the experimental time span was from 10 to 60 seconds, and the data was noted every 10 seconds. Figures 6 and 7 exhibit the experimental outcomes.

Figure 7 Distribution curve of recognition results from manual assessment (see online version for colours)



Consistent with the good classroom setting, the frame identification results revealed that the Happiness feeling had the most frequency of occurrence. With an average frequency of 21.67 events, the frequency of the happiness feeling changed in the frame identification between 20 and 23 times. This outcome suggests that the model can more effectively reflect students' positive classroom sentiments.

Reflecting the complexity and fluidity of classroom emotions, other emotions, such as anger, disdain, disgust, fear, sadness, and surprise, also displayed suitable crossing at some times. For instance, the findings of the frame identification revealed that at some time points Fear and Surprise were near in frequency, implying that students may show similar emotional reactions in some situations. Furthermore, the frequency fluctuations of anger and contempt in the framework recognition match the manual recognition results; yet the model is more sensitive to the identification of these emotions at specific time points.

Generally, especially in the high frequency recognition of Happiness, the framework recognition results match the manual recognition findings in terms of emotional distribution. This implies that more precisely the emotional dynamics in the classroom can be captured and reflected by the chosen emotion recognition paradigm. On some emotional categories, however, the manual recognition results cross over more often, suggesting that hand annotators have an edge in catching minute emotional variations. Still, frame recognition's performance captures variations in emotional dynamics, which is sufficient especially in the recognition of joyful emotions.

All the foregoing experimental results show that the suggested framework for assessing accounting smart classroom teaching quality has great dependability and accuracy in identifying student's expressive states. Getting students' expressive feedback in real-time, changing teaching tactics in response, and enhancing classroom interaction and teaching quality depend on this great accuracy. By means of expression recognition technology, teachers can better grasp the emotional fluctuations of their students in the classroom, therefore enabling more focused teaching modifications to raise the general teaching efficacy.

5 Conclusions and prospect

The paper develops a classroom teaching quality evaluation system driven by accounting smart education with feature identification technology for student expression. By gathering expression data in real-time and integrating it with learning behavioural traits, the system dynamically evaluates students' learning state and subsequently modifies teaching content and interaction strategies to fully assess teaching quality. The experimental results imply that the framework may effectively and consistently identify students' expression states, record dynamic classroom emotions, and provide real-time feedback for teachers to improve classroom instruction.

Although it has limitations, this study demonstrated that expression recognition technology can enhance instruction in accounting classrooms. While students' emotional responses may be influenced by more complex factors within the classroom, the experimental data is derived from a single dataset that encompasses only a few basic expression categories. Furthermore, this research focuses primarily on facial emotion detection, overlooking other important aspects such as body language, tone of voice, and additional multimodal information that could indicate a student's learning state. Even though the testing results suggest that the framework exhibits strong emotional detection accuracy, the long-term stability of this technology in dynamic environments needs further validation. Future directions of inquiry consist in:

1 Multimodal data fusion: by means of multimodal data, such as facial expressions, body language, and voice intonation, it can more precisely represent students' learning status and expression reactions, therefore offering teachers greater educational feedback (Peng and Nagao, 2021).

2 Long-term stability testing: long-term testing across a greater spectrum of classroom contexts to confirm the dependability and stability of the system in many educational settings. Long-term experimental data collecting and analysis help to better refined the model parameters thereby strengthening the system's resilience.

3 In-depth analysis of expression feedback: to attain ongoing development in teaching quality, further in-depth studies of the relationship between students' expression feedback and teaching effectiveness were carried out to expose the teaching guidelines behind the expression data.

4 Cross-disciplinary application research: investigating the application possibilities of the framework in other domains, like language instruction and scientific education, can prove the universality and adaptability of the framework and offer a larger range of application scenarios for the growth of smart education.

Acknowledgements

This work is supported by the Guangdong Education Science Planning Project (Higher Education Special Project) (No. 2024GXJK257), the construction project of teaching quality and teaching reform of undergraduate colleges in Guangdong Province 'Research on the implementation path of innovation and entrepreneurship education reform of accounting major under the background of new business' (Guangdong Education High Letter [2024] No. 9, No. 1198), the characteristic innovation project of ordinary universities in Guangdong Province (No. 2023WTSCX157), and 2024 Annual School-level Undergraduate Education and Teaching Quality Project 'Integrated Innovation Textbook: Entrepreneurship Management' (No. WCRHJC-24690).

Declarations

All authors declare that they have no conflicts of interest.

References

Ahmed, S.F., Alam, M.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. and Gandomi, A.H. (2023) 'Deep learning modelling techniques: current progress, applications, advantages, and challenges', *Artificial Intelligence Review*, Vol. 56, No. 11, pp.13521–13617.

Al-Mansouri, J.O. (2024) 'The impact of real-time feedback on optimizing teachers' classroom teaching pace', *Research and Advances in Education*, Vol. 3, No. 11, pp.45–48.

Chu, H-C., Tsai, W.W-J., Liao, M-J., Chen, Y-M. and Chen, J-Y. (2020) 'Supporting e-learning with emotion regulation for students with autism spectrum disorder', *Educational Technology & Society*, Vol. 23, No. 4, pp.124–146.

Dalvi, C., Rathod, M., Patil, S., Gite, S. and Kotecha, K. (2021) 'A survey of AI-based facial emotion recognition: features, ml & dl techniques, age-wise datasets and future directions', *IEEE Access*, Vol. 9, pp.165806–165840.

Gao, J. and Yang, T. (2022) 'Face detection algorithm based on improved TinyYOLOv3 and attention mechanism', *Computer Communications*, Vol. 181, pp.329–337.

Giang, T.H. (2024) 'Enhancing professional skills for accounting students to meet the program's graduation standards', *European Journal of Contemporary Education and E-Learning*, Vol. 2, No. 3, pp.226–246.

Hindarto, D. (2023) 'Comparison of RNN architectures and non-RNN architectures in sentiment analysis', *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, Vol. 7, No. 4, pp.2537–2546.

Jill, M., Wang, D. and Mattia, A. (2019) 'Are instructor generated YouTube videos effective in accounting classes? A study of student performance, engagement, motivation, and perception', *Journal of Accounting Education*, Vol. 47, pp.63–74.

Kaddar, B., Fizazi, H. and Boudraa, A-O. (2018) 'Texture features based on an efficient local binary pattern descriptor', *Computers & Electrical Engineering*, Vol. 70, pp.496–508.

Kokoç, M., Akçapınar, G. and Hasnine, M.N. (2021) 'Unfolding students' online assignment submission behavioral patterns using temporal learning analytics', *Educational Technology & Society*, Vol. 24, No. 1, pp.223–235.

Krstinić, D., Braović, M., Šerić, L. and Božić-Štulić, D. (2020) 'Multi-label classifier performance evaluation with confusion matrix', *Computer Science & Information Technology*, Vol. 1, pp.1–14.

Kumar, S., Sagar, V. and Punetha, D. (2023) 'A comparative study on facial expression recognition using local binary patterns, convolutional neural network and frequency neural network', *Multimedia Tools and Applications*, Vol. 82, No. 16, pp.24369–24385.

Li, S. and Deng, W. (2019) 'Blended emotion in-the-wild: multi-label facial expression recognition using crowdsourced annotations and deep locality feature learning', *International Journal of Computer Vision*, Vol. 127, No. 6, pp.884–906.

Lin, G. and Shen, W. (2018) 'Research on convolutional neural network based on improved Relu piecewise activation function', *Procedia Computer Science*, Vol. 131, pp.977–984.

Liu, Y., Pu, H. and Sun, D-W. (2021) 'Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices', *Trends in Food Science & Technology*, Vol. 113, pp.193–204.

Mehta, N.K., Prasad, S.S., Saurav, S., Saini, R. and Singh, S. (2022) 'Three-dimensional DenseNet self-attention neural network for automatic detection of student's engagement', *Applied Intelligence*, Vol. 52, No. 12, pp.13803–13823.

Mukherjee, A., Islam, M.Z. and Ali, L.E. (2024) 'Human iris classification through histogram of oriented gradient features with various distance metrics', *Machine Graphics and Vision*, Vol. 33, Nos. 3/4, pp.97–124.

Nirthika, R., Manivannan, S., Ramanan, A. and Wang, R. (2022) 'Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study', *Neural Computing and Applications*, Vol. 34, No. 7, pp.5321–5347.

Peng, S. and Nagao, K. (2021) 'Recognition of students' mental states in discussion based on multimodal data and its application to educational support', *IEEE Access*, Vol. 9, pp.18235–18250.

Polovnikov, V., Alekseev, D., Vinogradov, I. and Lashkia, G.V. (2021) 'DAUNet: deep augmented neural network for pavement crack segmentation', *IEEE Access*, Vol. 9, pp.125714–125723.

Tang, J., Liang, J., Yu, T., Xiong, Y. and Zeng, G. (2021) ‘Trip destination prediction based on a deep integration network by fusing multiple features from taxi trajectories’, *IET Intelligent Transport Systems*, Vol. 15, No. 9, pp.1131–1141.

Tang, X., Gong, Y., Xiao, Y., Xiong, J. and Bao, L. (2025) ‘Facial expression recognition for probing students’ emotional engagement in science learning’, *Journal of Science Education and Technology*, Vol. 34, No. 1, pp.13–30.

Tetzlaff, L., Schmiedek, F. and Brod, G. (2021) ‘Developing personalized education: a dynamic framework’, *Educational Psychology Review*, Vol. 33, pp.863–882.

Xiao, M., Tian, Z. and Xu, W. (2023) ‘Impact of teacher-student interaction on students’ classroom well-being under online education environment’, *Education and Information Technologies*, Vol. 28, No. 11, pp.14669–14691.