Article in Press IJIMAI journal

A Realtime Classroom Assessment System for Analysis of Students' Evaluation of Teaching Through a Deep Learning and Emotional Contagion Mechanism

Kuan-Cheng Lin[1], Ya-Hsuan Lin[1], Mu-Yen Chen[2] *

[1] Department of Management Information Systems, National Chung Hsing University, Taichung (Taiwan)
[2] Department of Engineering Science, National Cheng Kung University, Tainan (Taiwan)

* Corresponding author: kclin@nchu.edu.tw (K. C. Lin), lnc3596@gmail.com (Y. H. Lin), mychen119@gs.ncku.edu.tw (M. Y. Chen).

Received 15 June 2024 | Accepted 2 January 2025 | Early Access 6 March 2025

Abstract

Student evaluations of teacher performance are often derived from end-of-semester assessments, significantly impacting the authenticity of teaching evaluations but failing to provide real-time feedback. In addition, teachers' emotional states affect student performance, including in terms of learning motivation and classroom participation, which reflect the students' emotional state. This teacher-student emotional contagion mechanism focuses on the interaction of teacher-student emotions and can be used to observe the quality of instructional performance. Therefore, automatically detecting teacher-student emotional interaction and then providing real-time class satisfaction feedback can provide teachers with a more effective basis for adjusting classroom content. This research proposes an end-to-end classroom real-time teaching evaluation system based on automatic facial-emotion recognition, which can accurately detect and directly analyze the emotions of students and teachers in streaming frames. The system consists of two parts: First, a YOLO model based on deep learning approaches is used to automatically detect the emotional states of teachers and students during the teaching process; Then, combining the emotional contagion mechanism with the teaching evaluation scale, teaching satisfaction can be predicted using a Long Short-Term Memory (LSTM) model to output a classroom satisfaction score within a fixed period. Further analysis of the testing dataset confirms that the model has a high reliability in predicting teaching satisfaction. Research results show the proposed system can achieve an emotional recognition accuracy rate of 98.1% for teachers and 99.5% for students based on the emotion datasets. Further development could potentially provide teachers with strategies to improve classroom teaching effectiveness, better understand students' emotions and learning motivation, and improve learning outcomes.

Keywords

Classroom Assessment Scoring System, Deep Learning, Emotional Contagion, Long Short-Term Memory, Teaching Evaluation.

DOI: 10.9781/ijimai.2025.03.002

I. Introduction

Traditional teaching models and classroom education have evolved into a simultaneous Online Distance Learning (ODL) model. Teachers' instructional behavior is crucial to students' emotional experience and impacts classroom participation. Like in-person classroom instruction, distance teaching can trigger strong emotional responses [1]. In synchronous distance teaching, teachers' emotions can transmit through the students' audio/visual experience, prompting positive learning emotions among students and thus positively impacting learning performance and class management.

In the teaching process, the transmission of knowledge is impacted by the teacher's emotional state [2]. As the objects of classroom interaction, students will invariably be influenced by the emotional state of their teachers. Such emotions can be directly transmitted from teachers to students through an emotional contagion in which humans automatically imitate and synchronize with other people's facial expressions, vocalizations, postures and actions in social environments to achieve emotional convergence [3]. Second, teacher emotion affects student performance. Students who have a positive relationship with their teachers will have improved well-being and learning outcomes [4]. Finally, according to Fredrickson's Broadening and Building Theory [5], teachers with positive emotional states are prone to employing a broader array of teaching strategies, are better at adapting to different teaching situations, and exhibit greater flexibility and creativity, thereby benefiting students' learning outcomes.

Please cite this article as:
K. C. Lin, Y. H. Lin, M. Y. Chen. A Realtime Classroom Assessment System for Analysis of Students' Evaluation of Teaching Through a Deep Learning and Emotional Contagion Mechanism, International Journal of Interactive Multimedia and Artificial Intelligence, (2025), http://dx.doi.org/10.9781/ijimai.2025.03.002

On the other hand, students' emotions can also harm teachers. Nurmi and Kiuru discussed the arousal effect of students on teachers, especially in terms of teachers' emotions [6]. Frenzel et al. introduced the mutual causal inference model, showing that teachers who provide high-quality teaching can also derive positive emotional feedback [7], while unengaging instructional experiences will sap teachers' positive emotions.

Previous classroom emotional research largely relied on self-reporting scales, as emotions are highly subjective [8]. However, the self-reporting assessment scale involves questions that result in subjects' engaging in self-deception under social expectations, such as teachers' beliefs that it is inappropriate to express anger or fear in the classroom [9].

It is challenging for researchers to objectively identify people's inner emotions [10]. Facial emotion-based emotion calculation can collect data without interrupting subjects' learning status and performance. The facial action coding system proposed by Ekman & Friesen includes a wide range of action units, and the combination of facial action units can generate the possible range of various facial expressions [11]. There is long-standing theoretical and convincing empirical support for analyzing facial expressions to identify discrete emotions. In learning environments with computers, it is especially suitable for capturing learners' emotions [12], so we can install video cameras to capture facial expressions of teachers and students in the classroom. These emotional expressions can then be modeled to explore the influence of teachers' emotional expression and the complex emotional communication relationship formed by dynamic feedback in the classroom.

Affective computing has become one of the most active research topics in education [13], [14]. Even without face-to-face interaction, teachers' emotions will produce emotional contagion to affect students' emotions. Students and teachers should establish a reciprocal relationship so that both parties can effectively understand and communicate to prevent the production and transmission of negative emotions (e.g., sadness and disgust) [15], [16]. This study uses a facial emotion recognition method based on affective computing to automatically detect emotional state and to directly measure the emotional sentiment of teachers and students, thus allowing for the identification of subsequent emotional contagion between teachers and students.

Based on this emotional contagion, a real-time teaching evaluation system is constructed, replacing the traditional teaching evaluation questionnaire. The end-to-end system can anticipate students' learning status and give real-time feedback. This research uses the lightweight YOLOv5 model [17] to instantly identify the emotional content of teachers' and students' facial expressions in the classroom. In addition, this research uses the long short-term memory (LSTM) model to learn static emotion recognition results with time series characteristics, and finally builds a dynamic emotional contagion mechanism to predict student satisfaction with classroom teaching. The feedback data can be used as the basis for teachers to improve teaching strategies and content and to accelerate knowledge transfer between lecturers and students, thereby enhancing the learning effectiveness of learners.

II. Literature Review

A. Emotional Contagion

Darwin proposed that people capture changes in the emotions and mental states of others with whom they interact. This unconscious transmission of mental states during social interaction is called emotional contagion or emotional crossover [18]. Finally, people tend to display emotions similar to others (emotional convergence), also known as primitive emotional contagion. Emotional contagion is so widespread it has been noted by researchers in diverse fields such as social psychology, neuroscience, communication studies, and industrial-organizational psychology. In this process, the emotions expressed by one person are perceived by another, mainly through facial expressions, and then transferred through natural imitation and feedback.

In schools, teachers are the primary caregivers of students [19], [20] and are an essential component in the development of attachment relationships. Based on emotional spread, crossover theory should be able to better explain the relationship between teachers' teaching emotions and students' learning emotions in terms of positive and negative emotional states. Frenzel et al. showed that in a mathematics class, the teachers' positive teaching emotions could predict the happiness of students' learning emotions [21]. Bakker also found that music teachers who experience "flow" (a state of total immersion and intense joy) are more likely to have students experience learning engagement, indicating that emotional transmission between teachers and students can explain the importance of teachers' emotions [22].

Although these studies bring emotional contagion into the classroom and provide teachers with practical teaching advice, they also bring limitations. First, the emotions of both students and teachers are self-reported, which is prone to standard method variation [23]. Even using teacher and student sources [24], it still relies on self-reports. Although this method is a simple, effective and non-invasive way to assess teacher emotions, participant responses are based on the retrospective judgment of personal criteria influenced by subjective emotional experience [25]. In contrast, this study uses affective computing to automatically detect emotional states and directly measure the emotions of teachers and students.

B. Teacher Emotion and Student Emotion

A teacher's emotional competence refers to one's recognizing and managing one's emotions and those of others. It involves regulating one's feelings and uplifting those nearby [26]. Teachers who maintain more positive emotions in the classroom will generate more innovative ideas and strategies, while negative emotions will reduce motivation [27]. Appropriate training and interventions have been found to improve human emotional ability [28]. Teachers with higher emotional regulation skills tend to manage emotions better and have higher job satisfaction and well-being [29] and better teaching quality [30]. To sum it up, teacher emotion is empirically and conceptually essential to understanding the true nature of the trajectory of knowledge accumulation.

In an educational environment, teachers' words and actions significantly impact students, and knowledge is acquired through transfer from the teacher to the students. The teacher-student relationship also features emotional convergence. In contrast, the emotional contagion of peers is passed on equally. Furthermore, dynamic rendering is also closely related to social intimacy. Dobransky and Frymier argue that the teacher-student relationship should include control, trust, and intimacy [31]. The quality of the teacher-student relationship is related to the degree to which the teacher invests in the relationship. Continuous contact and interaction are needed to enhance mutual friendship, and emotional contagion is more likely to occur.

Primitive emotional contagion occurs directly in the classroom, and students can unconsciously imitate the teacher's behavior, an instinct conveyed by brain structures [32]. Laird & Bresler also pointed out that the feedback of facial muscles will have a significant impact on the subject's own emotions. Students imitating the facial expressions of their teachers and send feedback to themselves [33]. In addition, Ekman believed that both affective and automatic nervous system activities were affected by facial expression feedback [34], finding that subjects asked to act out one of the six basic emotions (anger, happiness, fear, disgust, sadness and surprise) would lead to the corresponding emotion [35]. Ekman also noted that each emotion has specific characteristics and can express emotions differently. Each emotion is only regarded as a discrete category, and there are no multiple emotional states [36]. Therefore, we choose "discrete" emotions to distinguish individual facial expressions. In addition, the present research aims to analyze the contagion of positive and negative emotions for the relationship between teachers' and students' emotions. The present research also combines the primary emotion classification of students and teachers into positive and negative emotions for discussion.

C. Classroom Assessment Scoring System (CLASS)

In 1995, Rosalind Picard of the MIT Multimedia Laboratory proposed the concept of Affective Computing [37]. Affective computing mainly focuses on how machines can be used to perceive human intentions and emotions and on establishing an appropriate emotion recognition model to meet the needs of different users. At present, the more common methods include facial recognition and detection of users' physiological state to identify their emotional state [38]. While collecting physiological state information through sensors is more objective and has higher accuracy than facial recognition, it requires subjects to wear sensors, and is thus not appropriate for use in actual classrooms. Therefore, this study uses facial expression recognition to analyze the emotional changes of students. This approach produces dynamic data that can be divided into three recognition mechanisms: Visual, Audio, and Audiovisual. In this study, images are used as the method for emotional calculation.

Some authors [16] found a positive correlation between classroom affective interaction, teaching methods, learning quality, and students' performance and engagement based on facial affective computing. The present research proposes a prototype for an emotional arousal scoring system based on facial and speech emotion, along with and head posture to obtain quantitative indicators for the teaching quality evaluation of teachers delivering online instruction. The experimental setup includes three pairs of online learning videos (6 videos in total), each of which is 10 minutes long. All courses are taught in one-to-one form in Chinese language, and the content focuses on high school mathematics, specifically elliptic problems in conic curves and equations [16]. In addition to evaluating students' learning processes, the system also considers the interaction between teachers and students. The output indicators include students' Affective Frequency Index, Affective Correlation Index, and Affective Arousal Level, and we verify the correlation between the Granger causality test of the teacher-student emotional sequence and the sequence.

III. The Present Study

Given the lack of behavioral measures and justification methods for teacher and student sentiment, the present study aims to identify teachers' and students' emotions through affective computing. The current research uses affective computing to perceive the emotions of teachers and students, monitor their emotional states during the teaching process, and examine the correlation between teacher and student emotions and teaching satisfaction. The facial detection emotion method can anticipate students' learning status, allowing for real-time feedback, and thus obviating the need for self-reporting. This could potentially provide teachers with real-time information about students with sub-optimal learning status, allowing them to dynamically correct or adjust their teaching methods or focus, thereby improving learning outcomes. The observed facial emotions can also be used to assess learner engagement and satisfaction, thereby providing an effective feedback mechanism for improving teaching quality and teacher-student interaction.

This research aims to:

1. Build an end-to-end real-time classroom assessment scoring system.
   a) Use YOLO to build a teacher and student facial emotion recognition model.
   b) Establish a teaching satisfaction model through deep learning and emotional contagion of teachers and students.
2. Use the teaching evaluation scale scores to verify the effectiveness of the instant teaching evaluation system.

IV. Method

A. Participants

5 teachers (2 males and 3 females) conducted the teaching for 25 graduate students (13 males and 12 females) in fields related to deep learning.

The experimental environment is long-distance synchronous teaching. The resulting database is called the "Emotion and Teaching Satisfaction Database". The teaching content of the course is divided into sub-units. Teacher-student interaction was recorded on video, with students asked to complete a teaching evaluation (using a five-point Likert assessment scale) after each unit as the basis for teaching satisfaction scores. Fig. 1 shows the experimental process.

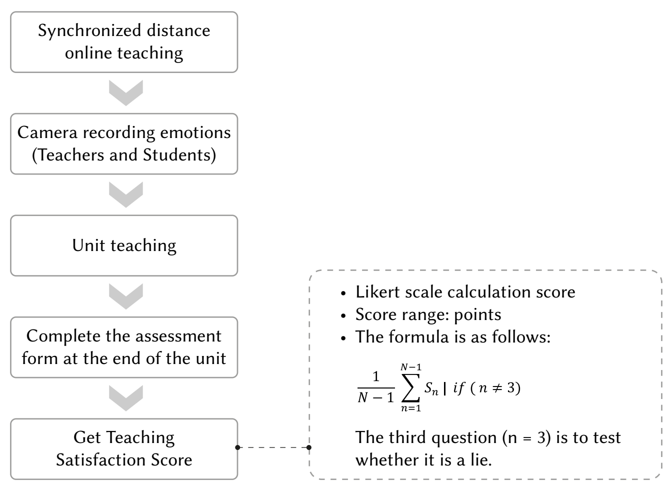


Fig. 1. Experimental design.

The Emotion and Teaching Satisfaction Database classifies emotions as joy, neutrality, anger, disgust, fear, sadness, and surprise [11]. The emotional data derived from teachers in the experiments are 3770 still images of the teachers' facial expressions produced while speaking, while an additional 1230 images were taken of students. Neutral emotions account for the largest proportion of images, followed by joy, sadness, anger, fear, disgust and surprise. Fig. 2 shows that teachers and students showed similar distributions of various emotional patterns. We can assume that the transmission of knowledge during the teaching process will be accompanied by student perception of the teachers' emotional state [2].

B. Experimental Procedure

The system construction process is divided into four steps and is explained in sequence as follows:

Step 1. This research recorded videos of multiple teachers and students engaged in remote synchronous teaching. Following each course unit, students were asked to complete an online assessment of the teaching content and performance. Students are asked to rate

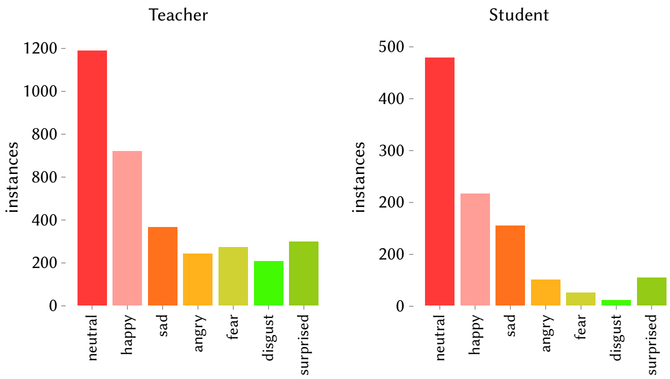


Fig. 2. Distribution of emotional states in the classroom.

the teacher's teaching content to obtain course satisfaction scores and establish an "emotion and teaching satisfaction database".

Step 2. A model trained in WIDER FACE [39] was used to assist in automatic labeling, obtain the facial coordinate values and reference emotion categories of teachers' and students' classroom emotions, and then re-label the emotion states through the basic emotion operation definition to establish classroom sample labels. In this study, the classroom emotion data set is used as training data to establish real-time emotion recognition models suitable for multiple teachers and students.

Step 3. We then discuss the emotional rendering of teachers and students in the course and the impact of classroom emotions on teaching satisfaction. First, we align the synchronized teacher and student videos, and respectively input the streams into the teacher and student classroom emotion recognition models. At this stage, the prediction results of the basic emotions are output, the teacher and student emotion recognition categories are arranged in sequence according to the continuity of the images, and the two output results are combined.

Step 4. Finally, the results of the image sequence are used as training data to establish a satisfaction score prediction model. In the verification stage, the analysis model and the actual evaluation scale score demonstrate the model's validity. This research aims to combine the above frameworks to construct an end-to-end real-time teaching evaluation system based on teacher-student emotional rendering. The research process of Steps 3 and 4 is shown in Fig. 3.

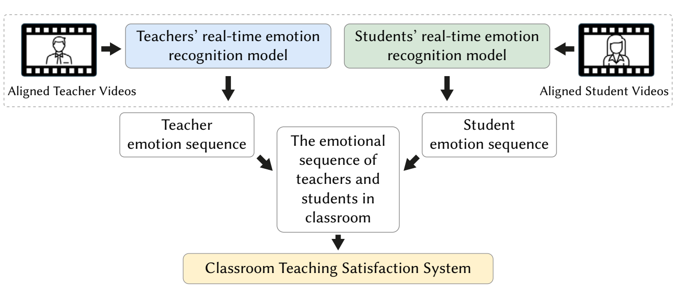


Fig. 3. Proposed system framework.

C. Instruments

1. End-to-end System Overview

YOLO's image capture function can receive the URL of a specific IP camera. The primary implementation method uses Real Time Streaming Protocol (RTSP) for still and video images transmitted over the network. The system can use an IP camera over wireless Internet to monitor remote teachers and students. The IP camera converts image signals into packets for wireless real-time transmission under the RSTP protocol. Next, multiple threads are used to simultaneously execute the YOLO real-time emotion recognition model in the classroom for teachers and students, while detecting the classroom images of teachers and students in real-time (1fps), and exporting the YOLO model's discrete emotion sequence output into the trained LSTM for a fixed period. The model then predicts a teaching effectiveness score, as shown in Fig. 4.

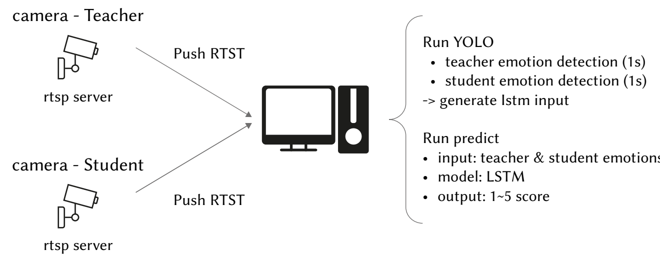


Fig. 4. End-to-end system data transfer.

2. Deep Learning Technology: YOLO Model

Taking the YOLO series as the representative of the classic one-stage detectors, the entire network structure consists only of convolutional layers and input images. The target category and position are directly returned after the convolution operation. This method has fast inference speed but low accuracy. It applies an end-to-end neural network to divide the image into grid regions while predicting the boxes in each region of the grid, treating the object detection problem as a regression problem. The model only needs to perform one operation. Each grid is only responsible for objects whose centers are within the grid and predicts the coordinates of the boxes and their scores. Each resulting rectangular box corresponds to a five-dimensional output, coordinates and confidence. The model uses the YOLOv5s architecture to establish emotion recognition models for teachers and students. The teacher and student classroom emotion datasets are divided into training, validation and testing sets with a ratio of 8:1:1, and the COCO pre-training weights are loaded into the PyTorch framework for transfer learning.

3. Deep Learning Technology: LSTM Model

This study's teaching satisfaction prediction model is constructed based on a many-to-one LSTM. We align a segment of the teacher images and multiple student images for each unit chronologically. Since the video is recorded at 30fps, the maximum speed of the emotion recognition model to predict an emotional state once is 0.03 seconds. In this study, an image is predicted at 1-second intervals, and two streaming images of teachers and students are respectively input to the teacher's and the student's emotion recognition models for emotion recognition (basic emotion) based on the corresponding teacher input. Two segments of the same time sequence are obtained. If there is no detection at the beginning of the video segment, the facial emotion is used to replace the facial emotion with neutral emotion, and if it is in the segment, it is replaced with the value of the previous emotion sequence. The classroom videos for each unit are mostly about 5 minutes long after alignment, thus we can obtain teacher and student emotional sequences with a length of 300 seconds. We then send the discrete emotions with the time series relationship into the LSTM model to achieve the prediction of satisfaction scores as shown in Fig. 5.

D. Data Analysis

1. Evaluation of Emotion Recognition

Below we provide a brief description of the indicator definitions used to assess the YOLO model's performance:

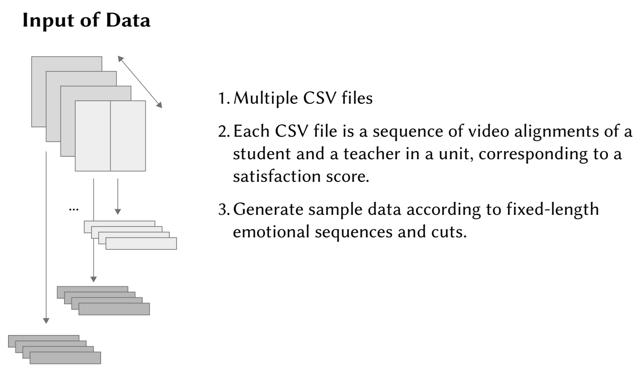


Fig. 5. LSTM input data processing method.

- Box loss measures the difference between the predicted bounding box and the ground-truth bounding box.
- Obj loss measures the difference between the predicted objectness and the true objectness. Objectness is defined as whether there is an object in the image.
- Cls loss measures the error between the predicted class of a detected object and its actual class.
- mAP_0.5 is the mean precision defined by the VOC dataset.
- mAP_0.5:0.95 is the mean precision defined by the COCO dataset.

In the study, objective evaluation metrics such as Precision, Recall, mAP (mean average precision), and F1 score were used to evaluate the performance of the trained classroom emotion recognition model, calculated using Eqs. (1) to (4):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{2}$$

$$\text{mAP} = \frac{1}{C}\sum_{K=i}^{N} P(k)\Delta R(k) \tag{3}$$

$$F1 = \frac{2}{\frac{1}{\text{Precision}}+\frac{1}{\text{Recall}}} \tag{4}$$

Among them, TP represents the number of correctly identified target emotions; FP represents the number of backgrounds incorrectly identified as emotions; FN represents the number of unrecognized emotions; C represents the number of target emotion categories; N represents the number of thresholds, K is the threshold, P(k) is precision, and R(k) is recall.

2. Evaluation of Satisfaction Score

Root mean squared error (RMSE) is a statistical indicator commonly used for testing neural network models, and is used here to analyze the degree of agreement between the predicted value and the actual value. The smaller the RMSE value, the higher the consistency between predicted and actual values. The smaller the deviation between the predicted and actual values, the more accurate and reliable the model's prediction result is. Therefore, RMSE can accurately reflect model performance, and is calculated using Eq. (5):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(\text{Predicted}_i - \text{Actual}_i)^2}{N}} \tag{5}$$

To further verify the feasibility of the satisfaction score model, this study divides the satisfaction score into five levels including A, B, C, D, and E. Class interval = total interval/number of groups. The number of groups is 5. The sentiment sequence is input into the LSTM model to predict the score. Based on the group distance, the satisfaction prediction results and the actual value of the evaluation scale were divided into different intervals to analyze the mean and standard deviation. In addition, statistical satisfaction scores and prediction consistency (number of errors ≤ 0.55) are used to judge the model's quality.

In addition to the above methods, the verification phase is considered. The Pearson correlation coefficient is also used to assess how well the model relates to the ground truth. The Pearson correlation coefficient is used to reflect statistics of linear correlation between two variables [40], with a value between -1 and 1, calculated using Eq. (6):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{6}$$

where x and y represent the two variables, x_i is the value of the actual score; y_i is the value predicted by the model; n is the total number of samples. The higher the absolute value of r_{xy}, the higher the correlation between the actual score and the model prediction. A coefficient value of 0.7 or greater is considered to indicate a high degree of correlation, 0.3-0.69 is a moderate correlation, and below 0.3 is a low correlation.

V. Results

A. Students' Classroom Emotion Recognition

The classroom teaching situation observed 25 students for seven emotional states, producing a total of 1230 student classroom emotion data items which were divided into training, verification and testing sets at a ratio of 8:1:1. The student emotion recognition model is applied to the test set validation (123 images). Table I shows that the average precision of the seven emotions is between 0.939 and 1. The average recall is between 0.984 and 1, while mAP@0.5 is 0.995, and mAP@.5:.95 falls between 0.99 and 0.995. The recall of neutral emotions is slightly lower than that of the other categories which were classified into other emotional categories.

TABLE I. Experimental Results of Student Emotion Recognition Model

Class	Labels	Precision	Recall	mAP@0.5	mAP@.5:.95
all	123	0.984	0.998	0.995	0.994
neutral	60	1	0.984	0.995	0.994
happy	26	1	1	0.995	0.991
sad	20	0.976	1	0.995	0.99
angry	6	0.998	1	0.995	0.995
fear	3	0.972	1	0.995	0.995
disgust	1	0.939	1	0.995	0.995
surprised	7	1	1	0.995	0.995

B. Teachers' Classroom Emotion Recognition

The teachers' classroom emotional data set included a total of 3,770 samples, approximately three times that of the students' classroom emotions. In classroom teaching, the teacher's speech expression involves the movement of facial muscles which affect mouth shape, thereby increasing the complexity of emotion recognition. Therefore, we use richer facial expressions to train the teachers' classroom emotion recognition model. We observed five teachers for seven emotional states, with the resulting samples divided into training, validation and test subsets at a ratio of 8:1:1.

The teacher emotion recognition model, is applied to validate the test set with a total of 377 images. Table II shows that the average detection precision for the seven emotional states is between 0.939 and 0.993, with an average recall between 0.933 and 1. mAP falls between 0.942 and 0.995, and mAP@.5:.95 falls between 0.817 and 0.965. The lowest mAP@0.5 is for angry, mainly because other emotion categories were misclassified as angry.

TABLE II. Experimental Results of Teacher Emotion Recognition Model

Class	Labels	Precision	Recall	mAP@0.5	mAP@.5:.95
all	377	0.981	0.98	0.981	0.875
neutral	144	0.991	1	0.994	0.965
happy	84	0.988	0.964	0.986	0.952
sad	40	0.975	0.974	0.979	0.881
angry	31	0.939	0.985	0.942	0.834
fear	29	0.993	1	0.995	0.827
disgust	20	0.991	1	0.995	0.817
surprised	30	0.989	0.933	0.975	0.848

C. Satisfaction Prediction Model

RMSE was used to evaluate the model's prediction performance, and the average deviation of the predicted value from the actual value was calculated to reflect the model's prediction quality, with a smaller value indicating better prediction results. Table III shows the loss functions for predicting teaching satisfaction scores for different emotion sequences models (i.e., length sequence, step size, emotion category). The basic emotions include anger, disgust, fear, happiness, sadness, surprise, and neutral. Binary emotions include positive and negative emotions. Table III shows that the best model for the experimental data is the basic emotion, with an emotion sequence of 160, a step size of 30, and an RMSE value of 0.1721. The RMSE after normalization of the satisfaction score is 0.1741, and the RMSE of both models is less than 0.2.

TABLE III. RMSE Data Comparison

Length + Step	Basic emotion	Basic emotion (minmax)	Binary emotion	Binary emotion (minmax)
80 + 30	0.2817	0.2487	0.3391	0.3307
160 + 30	**0.1721**	**0.1741**	0.2506	0.2434
160 + 60	0.2170	0.2322	0.3066	0.2192

Table IV shows the correlation between the model prediction results of different emotion sequences (length sequence, step size, emotion category) and the actual teaching satisfaction scores. The best model for the experimental data is the basic emotion, with an emotion sequence of 160 and a step size of 30. The Pearson correlation coefficients before and after the normalization of the satisfaction scores are 0.9796 and 0.9781, respectively.

TABLE IV. Pearson Correlation Coefficient Data Comparison

Length + Step	Basic emotion	Basic emotion (minmax)	Binary emotion	Binary emotion (minmax)
80 + 30	0.9407	0.9515	0.9141	0.9138
160 + 30	**0.9796**	**0.9781**	0.9594	0.9616
160 + 60	0.9642	0.9578	0.9277	0.9643

Tables III and IV show the error and correlation between the predicted value of each model and the actual value. The basic emotional model and the positive and negative emotional models are compared using the same training data. The basic emotional model with a sequence length of 80 or 160 and a step size of 30 outperforms the positive and negative emotional models. When observing the changes in the emotional sequence, since the positive and negative sequences have only binary values, few features can be expressed through emotional contagion. Relative to the correspondence between basic emotions, the seven discrete emotions are more refined and suitable for observing the rendering phenomenon in the classroom and applying it to assessment of teaching satisfaction.

Next, we use a finer degree of granularity to observe the basic sentiment model predictions with RMSE below 0.2. When the emotion sequence is set to 160 and the step size is set to 30, the RMSE is 0.1721 and 0.1741. However, it is impossible to know whether the predicted situation conforms to the actual situation, that is, whether the model can distinguish the satisfaction score. First, the actual teaching satisfaction scores are divided into five groups from low to high, where the total distance is 2.75, and the group distance is 0.55. The total number of samples assigned to A~E is respectively 67, 54, 60, 60, and 36.

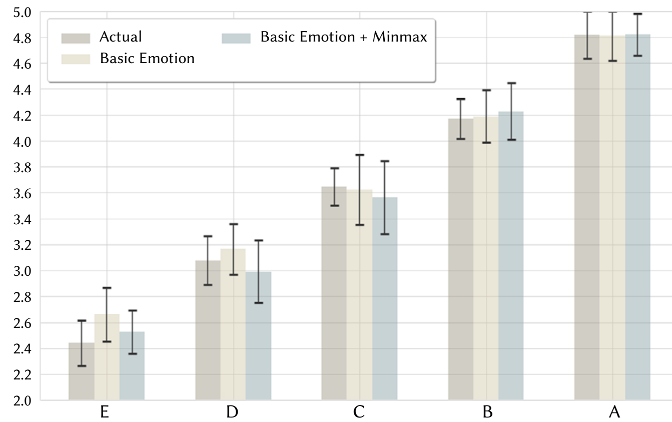


Fig. 6. Standard deviation and mean of emotion sequence 160 + step 30 model.

From the left to the right of Fig. 6, E is most prone to overestimation in all intervals. As shown in Table V, the average value of basic sentiment plus Minmax model overestimation error is 0.2357, and almost all predicted values are for overestimation, with an average overestimation error of the basic sentiment model of 0.1161. The average overestimation error of the Minmax model in the D area is 0.0964, while the average underestimation error is 0.1703. The average error of overestimation is smaller than that of underestimation, while the basic sentiment model is overestimated. The average error is 0.1427, and the average underestimation error is 0.0828. The average overestimation error is more significant than that of underestimation, which is opposite to the situation of the Minmax model where underestimation in the C area is severe, and the standard deviation of its distribution is the highest of all standard deviations, while the underestimation is the highest among all standard deviations. The MinMax model's underestimation in area C is particularly severe. The standard deviation of its distribution is the largest of all, and the underestimation error is the worst among all interval predictions, with an average value of 0.1811. However, without applying MinMax, the average values of overestimation and underestimation errors for the basic sentiment in area B do not exceed 0.15. Additionally, the averages for both models are slightly higher than the actual average values, indicating a tendency to overestimate. In contrast, area A shows the best performance among the five intervals. For both overestimation and underestimation in area A, the average error for the two models is less than 0.1. However, the basic sentiment model performed slightly worse in the low partition (E). Therefore, the MinMax model was selected as the best-performing model.

This study also explores potential over- and underestimation of the predicted value. The results predicted by the test set are divided into five-level groups, where class interval = total distance/number of groups. Given five sets, the set distance is 0.55. Therefore, 0.55 is set as the threshold value, and the statistical prediction value is the

TABLE V. The Ratio of High and Low Error ≥ Threshold Value

Interval	A	B	C	D	E
		Basic emotion			
Overestimated mean error	0.0521	0.1321	0.0891	0.1427	**0.2357**
Underestimated mean error	0.0554	0.0665	0.1711	0.0828	**0.0205**
		Basic emotion + Minmax			
Overestimated mean error	0.0658	0.1322	0.0686	0.0964	**0.1161**
Underestimated mean error	0.067	0.0839	0.1811	0.1703	**0.0469**

TABLE VI. The Ratio of High and Low Error ≥ Threshold V

Length + Step	**Basic emotion**		**Basic emotion (minmax)**		**Binary emotion**		**Binary emotion (minmax)**	
Proportion	Overestimate	Underestimate	Overestimate	Underestimate	Overestimate	Underestimate	Overestimate	Underestimate
80 + 30	1.4%	4.82%	1.4%	3.42%	1.4%	7.04%	2.21%	6.43%
160 + 30	**0.36%**	**0.72%**	**0.36%**	**1.08%**	0%	2.88%	1.44%	0.72%
160 + 60	0.59%	1.19%	1.78%	2.38%	2.38%	2.38%	1.78%	0.59%

proportion of the number of high and low estimates of the actual score ≥ 0.55. As shown in Table VI, when the basic emotion and emotion sequence is 160, and the step size is 30, the number of overestimation greater than or equal to the threshold is only 0.36%, and the number of underestimation greater than or equal to the threshold is 0.72%; after the satisfaction score is normalized, the number of overestimation greater than or equal to the threshold is 0.36%, and the number of underestimation greater than or equal to the threshold is 1.08%.

VI. Discussion

Questionnaires are generally used to solicit student assessments of teaching quality [15]. However, traditional questionnaire surveys suffer from two limitations in this context. First, since the survey time points fall before the midterm exam and before the final exam, teachers can only obtain teaching feedback at fixed time points. In addition, students may be motivated to not respond truthfully for fear of affecting their final course grade. At the same time, if the evaluation is a consideration for teacher career development and advancement, teachers will tend to seek to please students rather than improve the quality of teaching, and thus defeating the purpose of the evaluation. However, providing teachers with immediate and objective feedback on teaching satisfaction would effectively help teachers improve their teaching methods [41].

Teachers' emotions have a crucial impact on students' emotions in the classroom. Previous studies on classroom emotions mainly used self-reporting scales to assess emotions [42], [43], but responses to such surveys will be affected by personal subjective emotional experience or memory bias. A more objective alternative is to use facial emotion recognition [34], [36], paired with real-time automatic emotion detection from streaming images of teachers' and students' faces, thus overcoming problems raised by complex facial image preprocessing [12].

However, the quantitative scores output by the traditional "Classroom Assessment Scoring System (CLASS)" [16], [44], [45], [46], [47] are relatively subjective scores customized by researchers, and there is no standard for comparison. Due to the scarcity of teaching materials, it is impossible to more objectively adjust and verify the thresholds and score conversion formulas set in the scoring process. This research focuses on the emotional contagion of teachers and students in the teaching evaluation system. The proposed approach thus presents improved application value and practicability.

This research analyzes the facial expressions of students and teachers in classroom videos, creating models for the identification and prediction of emotional changes of students and teachers during class. Compared with existing teaching evaluation methods, the proposed approach offers two advantages. First, the evaluation scores based on detected facial expressions are more objective and reliable than questionnaire responses. Second, the proposed system evaluates teaching performance in real-time allowing for timely adjustment. The system consists of two parts: First, the YOLO model is used to automatically detect the emotions of teachers and students during the teaching process; Then, combining the emotional contagion mechanism with the teaching evaluation scale, the teaching satisfaction prediction model constructed by LSTM will quickly output the classroom satisfaction score. This allows teachers to dynamically adjust their teaching methods or mode of interaction with students to effectively improve learning outcomes.

VII. Conclusions

This research proposes an end-to-end classroom assessment system constructed by deep learning technology, with two specific research contributions.

First, we establish a YOLO model for real-time teacher and student classroom emotion recognition. After training using image-based emotion data of five classroom teachers, the accuracy rate of emotion model recognition reaches 98.1%, while accuracy for the students' model reaches 99.5%.

Second, we conduct correlation analysis between the time series LSTM model based on the emotional contagion of teachers and students and the classroom satisfaction scale. The RMSE of this study's best regression satisfaction prediction model is 0.1741, with a mean and standard deviation of 3.73 (+/-0.83). The mean and standard deviation of the accurate five-point scale was 3.75 (+/- 0.81). The Pearson correlation coefficient between the prediction results of the best model and the actual scale scores is 0.9781, which verifies that the model has a certain degree of reliability in predicting teaching satisfaction.

With the application of artificial intelligence technology, the system score can more directly reflect the learning situation than the traditional student evaluations of teacher performance. The proposed system directly provides real-time reference information for those involved in classroom teaching activities, allowing both teachers and students to dynamically adjust to improve learning efficiency. The proposed end-to-end system is both real-time and objective. With appropriate environmental conditions and equipment, facial emotion recognition does not require additional classroom observations during the teaching process. Remote online synchronous learning is emerging as a mainstream means of instruction, and the proposed system can provide important real-time feedback to better monitor learning status in remote classrooms, allowing teachers to monitor their teaching impact and make adjustments on the fly.

Future work can expand on the proposed approach and address certain research limitations.

- Support multi-person detection in traditional classrooms:

  While the present research focuses on application to remote classrooms, the techniques explored can also be applied to in-person instruction, capturing student faces by webcam. Additional improvements can help the system accurately identify emotional state even if the subject is wearing a surgical mask.

- Add facial emotion recognition for speakers:

  This research uses the YOLO method to directly detect emotions. It is worth adding a facial emotion recognition model for speakers, thereby improving the accuracy for identifying the teacher's emotional state.

- Promote teaching evaluation in various fields:

  Different student characteristics, subjects, and other factors that may affect teaching effectiveness should be considered, and satisfaction evaluation data sets in other fields should be collected to expand the richness of the data so that the model can be more consistent with reality.

- Address emergencies in the classroom:

  In the future, the real-time teaching evaluation system can be used to simulating a variety of emergency events that may occur in the teaching process, such as the impact of a teacher's anger on emotional communication in class can be analyzed post hoc to facilitate subsequent adjustments and improvements.

References

[1] D. Cavaliere, G. Fenza, V. Loia, F. Nota. "Emotion-Aware Monitoring of Users' Reaction With a Multi-Perspective Analysis of Long- and Short-Term Topics on Twitter," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 4, pp. 166-175, 2023, doi:10.9781/ijimai.2023.02.003.

[2] L. Tang, Y.S. Su. "Ethical Implications and Principles of Using Artificial Intelligence Models in the Classroom: A Systematic Literature Review," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 5, pp. 25-36, 2024, doi:10.9781/ijimai.2024.02.010.

[3] E. Hatfield, J. T. Cacioppo, and R. L. Rapson, Emotional contagion, New York: Cambridge University, 1994.

[4] J. Kim. "The quality of social relationships in schools and adult health: Differential effects of student–student versus student–teacher relationships," *School Psychology*, vol. 36, no. 1, pp. 6–16, 2021, doi: 10.1037/spq0000373.

[5] B. L. Fredrickson. "The role of positive emotions in positive psychology: the broaden-and-build theory of positive emotions," *American psychologist*, vol. 56, no.3, pp. 218–226, 2001, doi: 10.1037/0003-066X.56.3.218.

[6] J. E. Nurmi, and N. Kiuru, "Students' evocative impact on teacher instruction and teacher–child relationships: Theoretical background and an overview of previous research, " *International Journal of Behavioral Development*, vol.39, no.5, pp. 445-457, 2015, doi: 10.1177/0165025415592514.

[7] A. C. Frenzel, D. Fiedler, A. K. Marx, C. Reck, and R. Pekrun, "Who Enjoys Teaching, and When? Between- and Within-Person Evidence on Teachers' Appraisal-Emotion Links," *Frontiers in Psychology*, vol.11, 1092, 2020, doi: 10.3389/fpsyg.2020.01092.

[8] D. Keltner, J. L. Tracy, D. Sauter, and A. Cowen, "What basic emotion theory really says for the twenty-first century study of emotion," *Journal of nonverbal behavior*, vol. 43, no.2, pp. 195-201, 2019, doi: 10.1007/s10919-019-00298-y.

[9] M. L. Chang, "Emotion display rules, emotion regulation, and teacher burnout," *Frontiers in Education*, vol. 5, pp. 90, 2020, doi: 10.3389/feduc.2020.00090.

[10] K. R. Scherer, "What are emotions? And how can they be measured?," *Social science information*, vol. 44, no. 4, pp. 695-729, 2005, doi: 10.1177/0539018405058216.

[11] P. Ekman, and W. V. Friesen, "Facial action coding system", *Environmental Psychology & Nonverbal Behavior*, 1978.

[12] S. D'Mello, "A selective meta-analysis on the relative incidence of discrete affective states during learning with technology," *Journal of Educational Psychology*, vol. 105, no. 4, pp.1082–1099, 2013. doi: 10.1037/A0032674.

[13] S. Poria, E. Cambria, R. Bajpai, and A. Hussain, "A review of affective computing: From unimodal analysis to multimodal fusion," *Information Fusion*, vol. 37, pp. 98-125, 2017. doi: 10.1016/J.INFFUS.2017.02.003.

[14] E. Yadegaridehkordi, N.F.B. Mohd Noor, M.N.B. Ayub, H.B. Affal, and N.B. Hussin, "Affective computing in education: A systematic review and future research", Computers & Education, vol. 142, 103649, 2019.

[15] G. E. Lancia, "Instructor's Emotional Responses during Online Asynchronous Teaching," Ph.D. dissertation, Department of Applied Psychology and Human Development, University of Toronto, Toronto, Canada, 2021.

[16] J. Liu, and X. Wu, "Prototype of Educational Affective Arousal Evaluation System Based on Facial and Speech Emotion Recognition," *International Journal of Information and Education Technology*, vol. 9, no. 9, pp. 645-651, 2019, doi: 10.18178/ijiet.2019.9.9.1282.

[17] J. Glenn, yolov5. Accessed: Jun. 14, 2024. [Online]. Available: GitHub https://github.com/ultralytics/yolov5.

[18] C. Darwin, "The expression of the emotions in man and animals," *The Journal of Nervous and Mental Disease*, vol. 123, no. 1, pp. 90, 1956. doi: 10.1097/00005053-195601000-00016.

[19] P. Riley, "An adult attachment perspective on the student-teacher relationship & classroom management difficulties," *Teaching and Teacher Education*, vol. 25, pp. 626–635, 2009. doi: 10.1016/J.TATE.2008.11.018.

[20] P. Riley, "Attachment theory and the teacher-student relationship," London: Routledge, 2011.

[21] A. C. Frenzel, T. Goetz, O. Lüdtke, R. Pekrun, and R. E. Sutton, "Emotional transmission in the classroom: Exploring the relationship between teacher and student enjoyment," *Journal of educational psychology*, vol. 101, no. 3, pp. 705-716, 2009. doi: 10.1037/A0014695.

[22] A. B. Bakker, "Flow among music teachers and their students: The crossover of peak experiences," *Journal of vocational behavior*, vol. 66, no. 1, pp.26-44, 2005. doi: 10.1016/J.JVB.2003.11.001.

[23] P. M. Podsakoff, S. B. MacKenzie, J. Y. Lee, and N. P. Podsakoff, "Common method biases in behavioral research: a critical review of the literature and recommended remedies," *Journal of applied psychology*, vol. 88, no.5, pp. 879-903, 2003. doi: 10.1037/0021-9010.88.5.879.

[24] E. S. Becker, M. M. Keller, T. Goetz, A. C. Frenzel, and J. L. Taxer, "Antecedents of teachers' emotions in the classroom: an intraindividual approach," *Frontiers in psychology*, vol. 6, 635, 2015. doi: 10.3389/fpsyg.2015.00635.

[25] T. Goetz, M. Bieg, O. Lüdtke, R. Pekrun, and N. C. Hall, "Do girls really experience more anxiety in mathematics?," *Psychological science*, vol. 24, no. 10, pp. 2079- 2087, 2013. doi: 10.1177/0956797613486989.

[26] R. Akbari, R. G. Samar, G. R. Kiany, and M. Tahernia, "A qualitative study of EFL teachers' emotion regulation behavior in the classroom," *Theory and practice in Language Studies*, vol. 7, no. 4, pp. 311-321, 2017. doi: 10.17507/TPLS.0704.10.

[27] R. E. Sutton, and K. F. Wheatley, "Teachers' emotions and teaching: A review of the literature and directions for future research," *Educational psychology review*, vol. 15, no. 4, pp. 327-358, 2003. doi: 10.1023/A:1026131715856.

[28] O. L. Siu, C. L. Cooper, and D. R. Phillips, "Intervention studies on enhancing work well-being, reducing burnout, and improving recovery experiences among Hong Kong health care workers and teachers," *International Journal of Stress Management*, vol. 21, no. 1, pp. 69-84, 2014. doi: 10.1037/A0033291.

[29] H. Yin, S. Huang, and W. Wang, "Work environment characteristics and teacher well-being: The mediation of emotion regulation strategies," *International journal of environmental research and public health*, vol. 13, no. 9, 2016. doi: 10.3390/ijerph13090907

[30] T. Heydarnejad, A. Hosseini Fatemi, and B. Ghonsooly, "Emotions and self-regulation: A case of Iranian EFL high school and private language institute teachers," *International Journal of Educational Investigations*, vol. 4, no. 4, pp. 82-100, 2017.

[31] N. D. Dobransky, and A. B. Frymier, "Developing teacher-student relationships through out of class communication," *Communication Quarterly*, vol. 52, no.3, pp. 211-223, 2004. doi: 10.1080/01463370409370193.

[32] W. K. Berg, and M. Davis, "Associative learning modifies startle reflexes at the lateral lemniscus," *Behavioral Neuroscience*, vol. 99, no. 2, pp. 191-199. 1985. doi: 10.1037/0735-7044.99.2.191.

[33] J. D. Laird, and C. Bresler, "The process of emotional experience: A self-perception theory," In *Emotion*, M. S. Clark Ed., Thousand Oaks, CA, USA: Sage Publications, Inc., 1992. pp. 213–234.

[34] P. Ekman, "Facial expression and emotion," *American psychologist*, vol. 48, no. 4, pp. 384–392, 1993. doi: 10.1037/0003-066X.48.4.384.

[35] T. Hennig-Thurau, M. Groth, M. Paul, and D. D. Gremler, "Are all smiles created equal? How emotional contagion and emotional labor affect service relationships," *Journal of Marketing*, vol. 70, no. 3, pp. 58-73. 2006. doi: 10.1509/JMKG.70.3.58.

[36] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, pp. 169- 200, 1992. doi: 10.1080/02699939208411068.

[37] R. W. Picard, "Affective computing" Cambridge, MA, USA: MIT press, 1997.

[38] L. Shen, M. Wang, and R. Shen, "Affective e-learning: Using "emotional" data to improve learning in pervasive learning environment," *Journal of Educational Technology & Society*, vol. 12, no. 2, pp. 176-189.

[39] S. Yang, P. Luo, C.C. Loy, and X. Tang, X. "WIDER FACE: A Face Detection Benchmark," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5525-5533, 2016.

[40] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, "Time series analysis: forecasting and control", Hoboken, NY, USA: Wiley, 2015.

[41] K.C. Lin, N.T. Li, and M.Y. Chen, "Sustainable education on improving the quality of peer assessment: design and implementation of an online deep learning-based peer assessment system," *Library Hi Tech*, Vol. ahead-of-print No. ahead-of-print, 2024. doi:10.1108/LHT-11-2023-0569.

[42] T.C. Huang, M.Y. Chen, and W.P. Hsu, "Do Learning Styles Matter? Motivating Learners in an Augmented GeoPark," *Journal of Educational Technology & Society*, vol. 22, no. 1, pp. 70-81, 2019.

[43] K.L. Hsiao, T.C. Huang, M.Y. Chen, and N.T. Chiang, "Understanding the behavioral intention to play Austronesian learning games: from the perspectives of learning outcome, service quality, and hedonic value," *Interactive Learning Environments*, vol. 26, no.3, pp. 372-385, 2018.

[44] K.M. La Paro, and R.C. Pianta, "Classroom Assessment Scoring System (CLASS)," 2003. APA PsycTests.

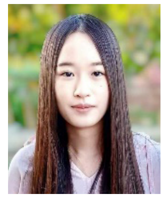
[45] S.M. Gamlem, "Mapping Teaching Through Interactions and Pupils' Learning in Mathematics," SAGE Open, July-September 2019, pp. 13. doi:10.1177/2158244019861485.

[46] C. Evertsen, I. Størksen, and N. Kucirkova, "Professionals' Perceptions of the Classroom Assessment Scoring System as a structure for professional community and development", European Early Childhood Education Research Journal, vol. 30, no. 5, pp. 701-714, 2022.

[47] C. Evertsen, I. Størksen, K.B. Tharaldsen, and N. Kucirkova, "Gains and challenges with the Classroom Assessment Scoring System in a social pedagogical tradition", Frontiers in Education, vol. 7, 965174, 2023. doi: 10.3389/feduc.2022.965174

Kuan-Cheng Lin

Dr. Kuan-Cheng Lin was born on September 13, 1964, in Taiwan. He received his Ph.D. in Applied Mathematics from National Chung Hsing University, Taichung, Taiwan, in January 2000. Since August 2015, he has been serving as a professor in the Department of Management Information Systems at National Chung Hsing University. His primary research interests include affective computing, text sentiment analysis, and big data analysis in meteorology and finance.



Ya-Hsuan Lin

Ya-Hsuan Lin was born on March 13, 1997, in Taiwan. She graduated from the Department of Management Information Systems, National Chung Hsing University, in 2022. During her master's degree, she researched the use of deep learning techniques in affective computing.

Mu-Yen Chen

Dr. Mu-Yen Chen is a Professor of Engineering Science at National Cheng Kung University, Taiwan. His current research interests include artificial intelligent, soft computing, data mining, deep learning, context-awareness, machine learning, and social network mining with more than 100 publications in these areas. He has served as Editor in Chief and Associate Editor of international journals [e.g. International Journal of Big Data and Analytics in Healthcare, IEEE Transactions on Engineering Management, IEEE Access, Applied Soft Computing, Granular Computing, Human-centric Computing and Information Sciences, Journal of Information Processing Systems, International Journal of Social and Humanistic Computing] while he is an editorial board member on several SCI journals.