He J. (2025) RESEARCH ON EMOTION RECOGNITION OF STUDENTS IN COLLEGE PHYSICAL EDUCATION ONLINE TEACHING BASED ON NEURAL NETWORK. Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte vol. 25 (100) pp. 330-347. **DOI:** https://doi.org/10.15366/rimcafd2025.100.021

ORIGINAL

RESEARCH ON EMOTION RECOGNITION OF STUDENTS IN COLLEGE PHYSICAL EDUCATION ONLINE TEACHING BASED ON NEURAL NETWORK

Jia He

Department of Physical Education, Hetao College, Bayannur 015000, Inner Mongolia, China
E-mail: 17604785800@163.com

Recibido 21 de junio de 2024 **Received** June 21, 2024
Aceptado 21 de diciembre de 2024 **Accepted** December 21, 2024

ABSTRACT

With the rapid development of information technology and the widespread adoption of home broadband and high-speed mobile networks, online education has become essential, especially during the COVID-19 pandemic. However, the neglect of physical education online has led to a decline in students' physical fitness. While familiarity with online learning has increased, there are higher expectations for teachers to engage with students' learning states and adjust their methods accordingly. Current online systems often focus on knowledge interaction, overlooking emotional engagement, which results in a significant "emotional lack." This paper proposes an online teaching model with emotional feedback using expression recognition technology, allowing teachers to monitor students' emotional states and adapt their teaching strategies. By implementing a convolutional neural network-based emotion recognition module, we can accurately classify students' learning emotions. The effectiveness of this module was tested through a nationwide online teaching experiment, showing high accuracy and practical value in addressing the issue of emotional disconnect in online education.

KEYWORDS: Online Physical Education; Convolutional Neural Network.

1. INTRODUCTION

In recent years, the concept of "Internet + education" has promoted the rapid development of the online education industry. By 2020, it will basically modernize education and basically form a learning society to develop. At the same time, personalized teaching driven by artificial intelligence technology has also become an important direction in the development field of education. Students, parents, and teachers all devote themselves to online education to understand its convenience and irreplaceable nature. Online education using the convenient environment of the Internet lets learners use new ways to acquire knowledge anytime and anywhere, breaks the traditional teaching in the process of fixed teaching place, fixed teaching time limit, lets learners freely arrange learning time, flexible choice of learning place, promote the development of lifelong learning. With the rapid development of information technology in the new century, a new generation of information technology arises at a historic moment, and many new technologies have become familiar keywords: big data, cloud computing, Internet of things, artificial intelligence, etc.; this new technology innovation, breakthrough and perfect, changed all aspects of public life, also provides new ideas for the improvement of learning environment.9 in the Education Informatization 2.0 Action Plan launched by the Ministry of Education points out that contemporary education informatization should vigorously promote intelligent education. Education should be learner-centered, promote intelligent teaching, and comprehensively promote the wide application of artificial intelligence in all aspects of the whole teaching process, teaching environment, educational activities, teaching management, and other directions. The intelligent learning environment combining artificial intelligence and online education platforms has become an inevitable requirement of online education, and it is also a high-end presentation form of the digital learning environment. From previous online teaching activities, we find that online education is a one-way learning process, and there is little direct communication between teachers and learners. This one-way learning process leads to the separation of teaching activities and learning activities and the loss of communication between teachers and learners, which not only affects the quality of teaching but also makes teachers lack motivation. Students lose communication with teachers, and the change in students 'emotions cannot get the teacher's attention and timely intervention. Especially for young students, self-study ability and self-binding force are poor, which affects the learning effect (Medina et al., 2021).

Both teachers, students, and parents have put forward higher requirements on the teaching mode, teaching quality, and educational environment of online education. Experts and scholars have proposed many methods to solve this problem. For example, an intelligent knowledge point recommendation system is constructed to create personalized learning, submit in-class tests, and arrange after-class auxiliary tracking to test the teaching effect. As for the judgment of learners 'learning experience and data collection, most online education platforms get users' feedback through the form of questionnaires, but the results are often incomplete answers or even misleading information. Under such circumstances, we can see that the existing online education platforms and live broadcast courses lack access to feedback from learners' learning emotions, resulting in teachers' passive teaching state. Online education fails to provide teachers with teaching auxiliary information, such as learners 'emotional state, and cannot help teachers to teach students in accordance with their aptitude, so teachers cannot adjust their learning strategies according to learners' learning experience, which is not conducive to the improvement of teaching quality. Therefore, we introduce expression recognition into the online learning emotion feedback, obtain the expression changes in learning, and feedback on the expression as an important learning state. Make the instructors accurately grasp the emotional ups and downs of the learners' learning, and guide the educators to appropriately respond to the emotional feedback. At the same time, emotion data can be applied to the intelligent recommendation system to assist the system development of the live broadcasting platform, which can help to make up for the lack of emotion in online education and create a better human-computer interaction environment.

2. Review of Research at Home and Abroad

2.1. Research on Expression Recognition Technology

The exploration of facial expression recognition technology can be traced back to the 1970s of the last centuries. In previous studies, most studies of facial expression recognition technology did not expand to the practical application direction but mainly focused on the theoretical research of biology and psychology. According to the trend change of facial muscle movement generated when face expression changes, literature (Liu, 2022) proposed a facial movement coding system to divide the face into 44 independent units to detect changes in facial expression. Literature (Pappaterra, 2022) proposed the optical flow method to perform facial feature extraction, which greatly promoted the development of expression recognition. Literature (Raizer, 2021) emotion computing theory takes expression recognition as a current research hotspot; calculating the dynamics of expression to identify emotions has attracted the attention of many researchers. Professor (Fan et al., 2021) Gao Wen led the research team of Harbin Institute of Technology to introduce facial expression recognition technology and foreign advanced achievements into China. In 2003, Professor Wang Zhiliang summarized and summarized the development of facial recognition expressions in China and led the team of the University of Science and Technology Beijing to study the direction of emotional control. In 2006, the National Natural Science Foundation of China approved a research study on facial expression recognition. Literature (Hu et al., 2021) uses the slope and distance of eyebrows, eyes, lips, and mouth to depict the expression of the human face. Literature (Hu et al., 2021) proposes a non-rigid body motion light flow algorithm to form a facial expression image feature flow for facial expression recognition. Literature (Zhou, 2018) proposes an automatic classification method of facial expressions in the neural network tree with a good recognition effect. Literature (Zhou, 2018) proposed the local Gabor transformation method for automatic segmentation to extract the local feature information of faces so that the expression recognition rate reaches 85.714%. Literature (Ge et al., 2023) uses the spatiotemporal information of facial expressions to transform facial expression images into motion trajectory and uses multi-scale analysis to complete the dimension reduction of motion trajectory.

With the continuous development of computer technology, expression recognition technology has also attracted the wide attention of researchers. Through the research of scientific research experts in various countries, the expression recognition technology has been developed rapidly. The research basis is the establishment of a facial expression database, so various research institutions have also developed a facial expression database (Hopfield, 1982). For example, the early JAFFE Japanese Women Expression database was established by Japan, the ORL face database was created by the Olivetti Laboratory of Cambridge University, the Yale face database was created by the Computational Vision and Control Center of Yale University, the MIT face database created by the Media Laboratory of MIT University, and Ryerson Face Expression database of Canada, etc. In China, a documentary (Wang et al., 2020) has established a small-scale facial expression video database. Document (Lv et al., 2018) has established a large-scale Chinese facial expression video database, which is a more comprehensive facial expression database in this field in China (He, 2021).

2.2. Current Status of Emotion Research in the Online Education Field

Throughout the educational experience, students exhibit a range of emotional responses to varying subject matter. When students find the material comprehensible and aligned with their interests, they display positive emotions such as happiness or excitement, reflecting a heightened state of enthusiasm. Conversely, when the content is challenging to grasp or does not resonate with their preferences, students may exhibit negative reactions like furrowing their brows and showing a lack of engagement, indicative of a subdued mood. The significance of these facial cues is paramount, as they offer critical instructional insights to e-learning platforms and instructors (Mehta et al., 2018). Reference (Zhang et al., 2025) developed an internet-based educational platform that employs affective computing principles to detect six fundamental emotional states: joy, anxiety, sorrow, astonishment, irritation, and disgust. Reference (Cheng et al., 2022) integrates additional physiological indicators, including heart rate variability, blood pressure fluctuations, and skin conductance, to gauge the emotional shifts among students and to investigate how to leverage these emotional responses to enhance the educational experience. Reference

(Li et al., 2013) introduces the FILTWAM framework, which incorporates both facial and vocal expression analysis, utilizing peripherals like cameras and microphones to capture students' expressions and vocal cues. The FACS system, as put forth by Reference (Pal & Singh, 2010), has led to the creation of the SLE-FER framework for emotion analysis through facial recognition. Reference (Stahlberg, 2020) presents an intelligent-agent-based model for emotion and cognition recognition, incorporating eye-tracking and emotion-sensing technologies. Drawing on emotion detection, Reference discusses methods for detecting student fatigue during learning and suggests interventions to address it (Haarsa, 2024).Reviewing the aforementioned research, it is evident that learners' facial expressions are a more natural and authentic indicator of their emotional responses to learning, as opposed to other physiological signals. These expressions provide an objective reflection of their emotional states, garnering significant attention from researchers globally. According to Reference (Dashtipour et al., 2020), the field of education has seen a focus on emotion identification, with facial expression recognition proving to be more effective than alternative methods. Yet, the application of expression recognition in education has primarily aimed at addressing the emotional deficit in online learning environments rather than directly informing teaching strategies (Gholami et al., 2022). Initially, facial expression recognition in educational research relied on conventional machine learning techniques, including facial detection, feature analysis, and classifier development. More recently, the exploration of deep learning techniques, particularly convolutional neural networks, has emerged as a pivotal area of study.

3. Research on Emotion Recognition of Students in College Physical Education Online Teaching Based on Convolutional Neural Network

3.1. Expression Recognition System Based on Emotion Recognition

Face expression is important non-verbal information that reflects the emotional changes of users. Basic human expressions are divided into six categories: joy, anger, fear, sadness, surprise, and disgust, while emotional information includes 7% language, 38% voice, and 55% facial expressions. People make real responses to emotional experiences or show real physiological signals to judge the subject's emotional emotions through facial expressions, which can reach 88% accuracy. Therefore, the research in this paper is based on this theory. Facial expression analysis represents the utilization of artificial intelligence to detect and interpret changes in facial expressions. As computer technology advances swiftly, research into facial expression analysis has become increasingly sophisticated. This process typically encompasses three key stages: initial image processing, subsequent feature identification, and final categorization, as illustrated in Figure 1.

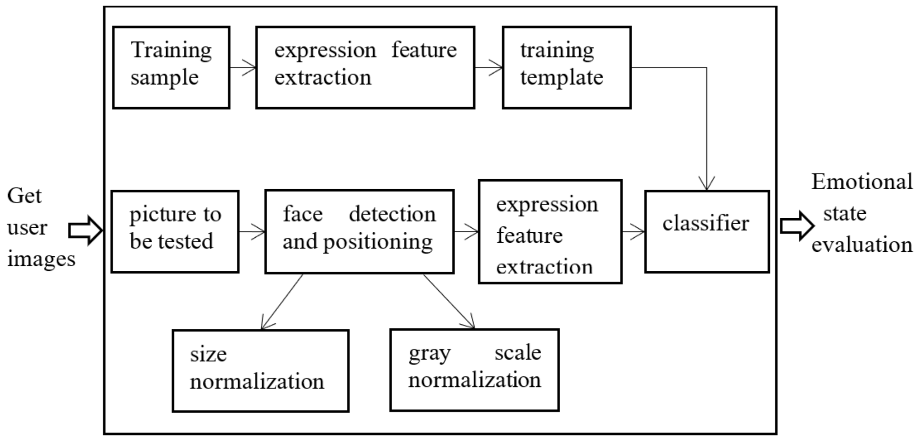


Figure 1: The Process of Facial Expression Recognition

After years of exploration and development of expression recognition technology, the market potential has been continuously developed, attracting more and more professional researchers. Different research institutions at home and abroad have also developed a rich database of facial expressions to meet the research needs of more fields. The JAFFE database has 213 images acquired in an experimental environment. It contains seven expressions: joy, sadness, anger, disgust, surprise, fear, and neutral; it belongs to a relatively early database, KDEF, and the AKDEF dataset; the dataset has 4,900 images and contains seven expressions from 70 people aged 20 to 30 years, Each male and female, The collector tor is the same color, No makeup, beard, glasses, earrings, etc., Light exposure is uniform, Each person has five different angles for each expression, Mainly used in the scientific research of medicine and psychology; GENKI data set, This dataset collected images of different regions, different light, different ethnic groups, and backgrounds. Therefore, each picture size, size, light, posture, etc., are not the same; images are divided into "laughing" and "not laughing" " mainly used for smiling face recognition; in the RaFD data set, A total of 8,040 images were, DE contempt is added to the seven basic expressions. The images contain 49 white models (20 adult men, 19 adult women, four boys, and six girls) and 18 Moroccan adult men. The collection of each expression requires models to change three different directions and shoot through 5 cameras from different angles. The data quality is relatively high; the CK + data set, released in 2010, is expanded on the original Cohn-Kan ade data set, including 593 images of 123 participants and emotional tags with seven expressions, indicating the expression of the experimenter. This database is a highly popular database widely used in facial expression recognition tests; the Celeb A dataset contains 202599 facial images, 40 attribute tags covering complex backgrounds and poses, 10177 identities, and five geographic coordinates dedicated to studying facial attributes; the Fer2013 dataset contains 358864848 grayscale images into a test set and validation set, seven expressions, each with corresponding digital tags.

3.2. Convolutional Neural Network

The advent of deep learning has brought the representational capabilities of convolutional neural networks (CNNs) to the forefront, especially with advancements in numerical computation hardware. Over the past few years, a significant area of research has been the exploration of CNNs and their relationship with other types of neural networks. These networks are structured with an input layer, an output layer, and several intermediate hidden layers, with each layer comprising numerous neurons, as depicted in Figure 2.

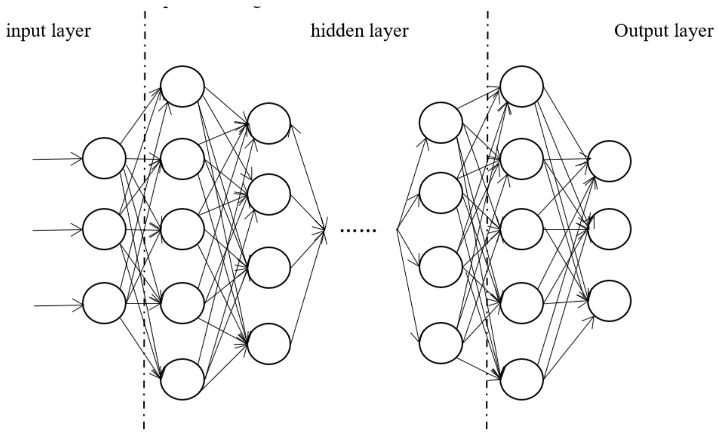


Figure 2: Schematic Representation of the Neural Network

As illustrated in Figure 2, in conventional fully connected deep neural networks (DNNs), neurons in the lower layers are connected to those in all higher layers. Such a comprehensive, multi-tiered connectivity is prone to cause an increase in the number of parameters, potentially leading to overfitting and entrapment in local optima. Furthermore, the distinct local features, such as those found in the eyes, nose, and mouth, are not adequately captured by this architecture.

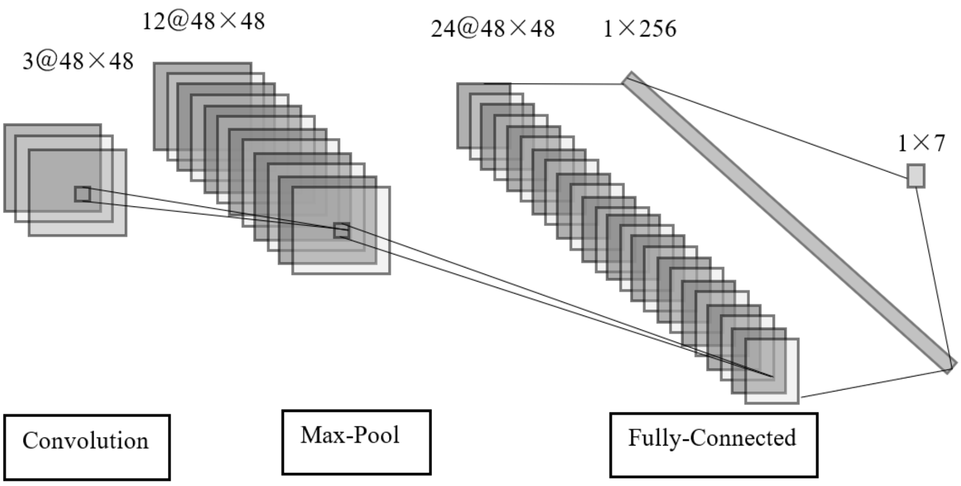


Figure 3: Schematic Representation of the Convolutional Neural Network

Figure 3 delineates the foundational framework of CNNs, which incorporates hidden layers that include convolutional, pooling, and dense (fully connected) layers. In CNNs, connectivity is not established between every neuron across adjacent layers. These networks are defined by their local connectivity, weight sharing, and sampling reduction. They are superior to traditional neural networks in effectively capturing local image features and maintaining robustness against image distortions. The convolutional layer plays a crucial role in the network for feature extraction. Post-convolution, the pooling layer is tasked with diminishing the computational load by decreasing the resolution of the subsequent layer's output and the number of parameters for network training. Preceding the output layer, the dense layer transforms the high-dimensional data from the prior layers into a one-dimensional format for classification.

3.2.1. Convolutional Layer, Pooling Layer, Full Connection Layer

Convolutional layer: A key component within convolutional neural networks (CNNs) is the convolutional layer itself. Within this layer, the convolution kernel—alternatively termed the filter—progresses across the input matrix with a defined stride, moving horizontally and vertically to execute the convolution operation, ultimately generating the feature map. The resulting output from this layer is detailed in Equation 3-1.

$$feature_j^l = \phi\left(\sum_{i \in M_j} w_{ij}^l \otimes x_i^{l-1} + b_j^l\right) \quad (3\text{-}1)$$

In Equation 3-1, the convolution operator represents the j th output feature graph in layer l, M_j represents the set of input feature graphs, w_{ij}^l represents the convolution kernel of the j th feature graph, and the I th input data, x_i^{l-1} represents the I th feature graph in layer l-1, b_j^l represents the bias of the j th feature graph in layer 1, and represents the activation function. Commonly used activation functions are ReLU, sigmoid, tanh, etc. （2）Activation function. Activation functions endow convolutional neural networks (CNNs) with the capacity for nonlinear representation. In the absence of an activation function, CNNs would be confined to linear transformations, rendering the distinction between complex, multi-layered CNNs and simpler, single or dual-layer architectures minimal. The following section outlines several prevalent activation functions. The Sigmoid function is accessible everywhere in the domain, with the output between 0 and 1. Since the Sigmoid function tends to 0 at both ends, the function value change range is very small, which easily leads to the gradient disappearance and is not conducive to the reverse transmission of deep neural networks. The expression for the Sigmoid function is shown in Equation 3-2:

$$sigmoid(x) = \frac{1}{1+e^{-x}} \tag{3-2}$$

The Hyperbolic Tangent (Tanh) function, an odd function, outputs values within the interval of -1 to 1 and is characterized by its gradient saturation properties. Compared to the Sigmoid function, the Tanh function has a zero-centered output, which tends to lead to faster convergence during the training phase of neural networks. The formula for the Tanh function is presented subsequently.

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{3-3}$$

The ReLU function is a segment function with derivatives of 1 when x is greater than 0 and 0 when x is less than 0. The ReLU function solves the gradient vanishing problem in the positive interval, and compared to the Sigmoid and Tanh functions, the ReLU function grows linearly and has much faster convergence than the Sigmoid and Tanh functions. To solve the gradient vanishing problem, most scholars use the modified linear units as the activation function. The expression for the ReLU function is shown below.

$$relu(x) = \begin{Bmatrix} x \; if \; x \geq 0 \\ 0 \; if \; x < 0 \end{Bmatrix} \tag{3-4}$$

Pooling layer: The role of the pooling layer is to diminish the spatial dimensions of feature maps while retaining the essential information. Typically, these layers are interspersed at regular intervals within the architecture of convolutional neural networks (CNNs).

Primarily, CNNs employ two types of pooling layers: max pooling and average pooling. The pooling of the size 5 * 5 feature map was completed by defining a pooling kernel of size 2 * 2 and step size 1 for 16 calculations. The pooling process and convolution calculation processes are similar, using the specified size pooling kernel to slide from the left to the right and from the top down on the feature graph. The pooling layer places more emphasis on the relative location between the features rather than the exact location of the features themselves.

（3）Full link layer: The fully connected layer maps the high-dimensional feature maps obtained after convolution and pooling to the low-dimensional space maps the facial expression features learned by the convolutional neural network to the marker space of the dataset, and outputs n data, representing the probability of each type in the n category. In general, the highest probability of the n species is selected as the final output.

3.2.2. Forward-Propagation and Back-Propagation of Convolutional Neural Networks

Convolutional neural networks can also be divided into forward propagation mechanism and backward propagation mechanism, which will be described specifically below: Convolutional neural networks propagate forward. Convolution layer operation: In a convolution layer, the feature graph of the previous layer is convolved by the learned convolution kernel. Then, these convolution operations are combined, and an incentive function can get an output feature graph. The formula is expressed as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (3\text{-}5)$$

Subsampling layer operation: for the subsampling layer, there are N input maps there are N output maps, but each output graph is smaller; the formula is expressed as:

$$x_j^l = f\left(\beta_j^l down\left(x_j^{l-1}\right) + b_j^l\right) \quad (3\text{-}6)$$

Convolutional neural network backpropagation. Convolutional layer gradient calculation: assuming that each convolutional layer l will receive a subsampling layer l + 1, according to the classical neural network backpropagation algorithm, to get each neuron layer one weight gradient, you need to find the residual of each neural node in layer 1, and then multiplied by the weight of the connection, multiplied by the current layer l the neuron node input of the incentive function, so you can get the current layer l each neural node. Multiplying the result completes the calculation of the l-layer residue. This process is described by the formula as:

$$\delta_j^l = \beta_j^{l+1}\left(f'\left(u_j^l\right) \cdot * \, up\left(\delta_j^{l+1}\right)\right) \quad (3\text{-}7)$$

$$up(x) = x \otimes l_{mn} \quad (3\text{-}8)$$

$$\frac{\partial J}{\partial b_j} = \sum_{u \cdot v}\left(\delta_j^l\right)_{uv} \quad (3\text{-}9)$$

$$\frac{\partial J}{\partial k_{ij}^t} = \sum_{u \cdot v}\left(\delta_j^l\right)_{uv}\left(p_i^{l-1}\right)_{uv} \quad (3\text{-}10)$$

$$\frac{\partial J}{\partial k_{ij}^t} = rot180\left(conv2\left(x_i^{l-1}, rot180\left(\delta_j^l\right), 'valid'\right)\right) \quad (3\text{-}11)$$

Gradient calculation of down sampling layer: The forward process of down sampling is a multiplicative parameter and a bias term b corresponding to each feature graph. If the residual map of this layer is obtained, the gradient of these two parameters is easily obtained. When calculating the gradient of the convolution core, we need to find which input area corresponds to output pixels; we need to find the current layer residual area corresponding to the next layer of the given pixels so you can see the residual reverse propagation back in addition, need to be multiplied by the weights between input area and output pixels, formula expressed as:

$$\delta_j^l = f'(u_j^l) \cdot * \, conv2(\delta_j^{l+1}, rot180(k_j^{l+1}), 'full') \quad (3\text{-}12)$$

$$\frac{\partial J}{\partial k_j} = \sum_{u \cdot v} (\delta_j^l)_{uv} \quad (3\text{-}13)$$

$$\frac{\partial J}{\partial \beta_j} = \sum_{u \cdot v} (\delta_j^l \cdot * \, d_j^l)_{uv} \quad (3\text{-}14)$$

3.3. Emotion Recognition Based on Convolutional Neural Networks

3.3.1 Emotion Recognition Module Design

The architecture of the emotion recognition system is primarily tasked with detecting fluctuations in facial expressions as students engage in learning activities. Figure 4 depicts the schematic structure of this module. This system is broadly segmented into two components. The first component leverages a pre-existing dataset of expressions to train a convolutional neural network, thereby creating an expression feature model. The second component processes the facial images captured from users and subsequently delivers the emotion recognition outcomes.

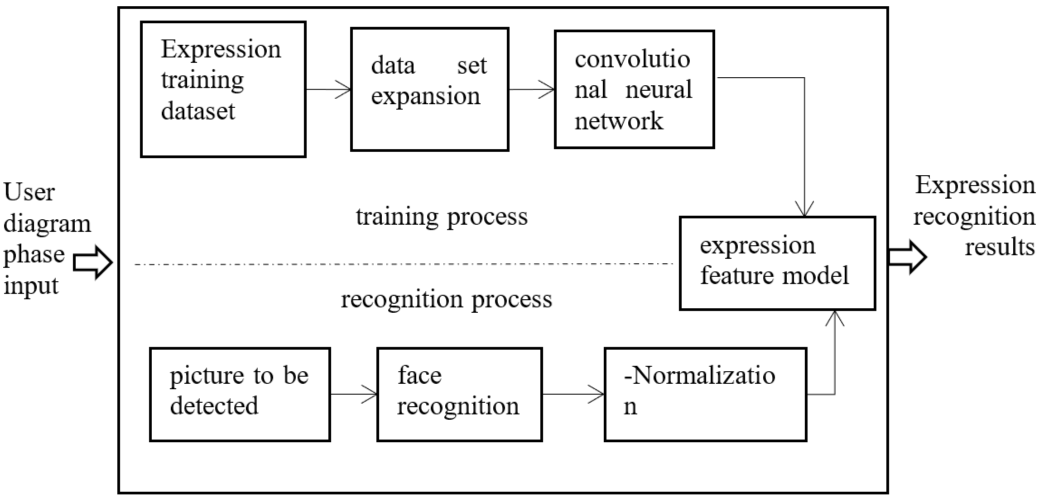


Figure 4: Schematic Diagram of the Emotion Recognition Module

3.3.1.1 Training Process

Step1: Obtain the facial expression training data set;

Step2: Normalized the images in the facial expression training dataset;

Step3: The normalized expression training dataset was augmented;

Step4: Send the newly generated expression training data set to the defined convolutional neural network for training;

Step 5: Save the trained model and end the exit.

The expression feature model plays a role in classifying the input expression in the whole emotion recognition module. The training time of the expression feature model is related to the size of the training data set and the device used for the training. In general, the model update is conducted offline.

3.3.1.2 Recognition Process

Step1: Get learner images to be detected through the learner's smart device camera;

Step2: Identify the facial contour in the image to be detected and intercept the facial information;

Step3: Size normalization and grayscale normalization of learners' facial information;

Step4: Send the normalized images to be detected to the trained expression feature model for detection;

Step 5: Output the detection results and end the exit.

3.3.2 User Expression Recognition

In this study, the training phase leverages the CK+ dataset, which encompasses 123 participants, comprising 593 image sequences in total. Among these, 327 sequences are annotated with labels for seven distinct emotional expressions: anger, contempt, disgust, fear, happiness, sadness, and surprise.

Upon analysis of the dataset's image sequences, it is observed that the frequency of each emotional expression is inconsistent. To maintain the integrity of the dataset, the detailed composition of the dataset utilized is presented in Table 1.

Table 1: The Expression Distribution Table of the CK + Dataset used in this Article

CATEGORY	ANGER	DESPISE	DETEST	FEAR	HAPPY	SAD	AMAZED	TOTAL
QUANTITY	135	54	177	75	207	84	249	981
PROPORTION	13.76%	5.50%	18.04%	7.65%	21.10%	8.56%	25.38%	100%

In order to test the actual use effect, the OpenCV call and the local camera call were used to load the trained model for real-time test, and the expression test effect is shown in Figure 5.

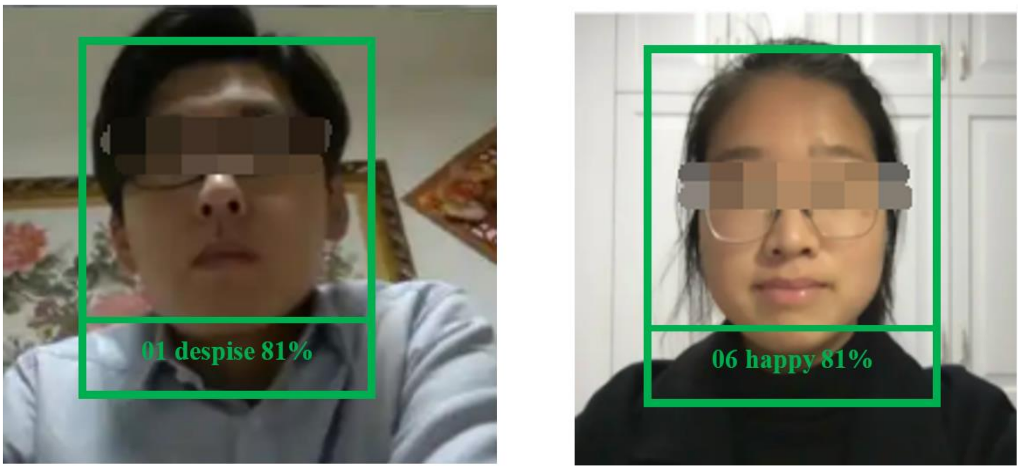


Figure 5: Facial Expression Test Effect

4. Experimental Analysis

4.1. Online Practice

Prior to the commencement of the online study, it is essential to deploy the emotion recognition system on the computers of all participants, with the online learning phase initiating once the system is operational. Participants capture their own images using the computer's webcam at 10-second intervals for the analysis and recognition of their facial expressions. Twenty-two of the 25 subjects could not participate in the test, so the online experiment was completed by 22 subjects. Between 262 and 287 images were acquired per learner. In the overall learning process, some subjects may have bowed their heads to take notes, or have never collected their facial expressions due to external factors. For this case, this experiment set them to a neutral learning state. For relatively simple and effective statistics and recording of learners' learning status, 230 learning images of each subject after 4 minutes and 50 seconds from the start of the class session were used. By identifying the collected image data, the learner's facial expression in the image as a positive learning state is 1, the learner faces a negative learning expression is 1, and the learner faces a neutral learning state is 0. The learning status transcript table for the 22 subjects is shown in Table 2.

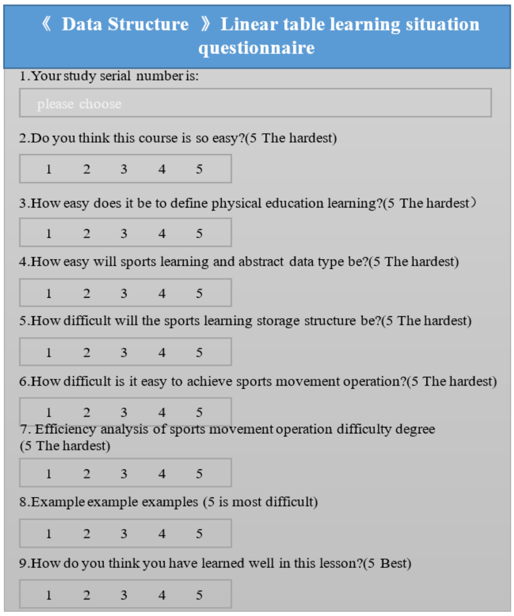
Table 2: Subject Learning Status Record Table

NUMBER	POSITIVE	NEUTRAL	NEGATIVE
1	3	201	26
2	0	172	58
3	0	203	27
4	4	141	85
5	1	218	11
6	1	175	54
7	0	166	64
8	5	221	4
9	0	143	87
10	0	153	77
11	1	181	48
12	2	209	19
13	0	194	36
14	4	185	41
15	1	195	34
16	0	170	60
17	2	187	41
18	0	162	68
19	0	134	96
20	0	191	39
21	0	185	45
22	1	175	54

Through the recording of subject learning status, the negative learning state statistics, and the temporal learning map of each learner, the learners are roughly divided into three categories, namely, active learners, mass learners, and burnout learners. (1) Active learner: Figure 5.3 shows the timing diagram of learning state changes represented by learner 8. This kind of learners have less fluctuation in their learning state. In most of their time, their learning expression is relatively calm and in a neutral learning state, with more attention in the learning process. These learners usually have a strong interest in learning, spending more than 85% of their classroom time. Such learners accounted for 22.7% of the total number of students in this experiment. (2) Mass-type learners: Figure 5.4 shows the timing diagram of learning state changes represented by learner No.22.The positive learning state of such learners is almost 0, and the fluctuations between neutral and negative learning states during the learning process cannot maintain learning attention for a long time. Such learners spend less time in class than active learners. Such learners accounted for 59.1% of the total number of participants in this experiment. (3) Burnout-type learners: Figure 5.5 shows the timing diagram of learning state changes represented by learner # 19. The positive learning state of such learners is almost 0, and the learning expression fluctuates greatly throughout the learning process. They frequently switch between neutral and negative learning states and are almost unable to maintain learning attention. Such learners have the least proportion of classroom investment time compared with the first two types of learners. From the data, it is impossible to know whether such learners study effectively in the learning process, which makes them seem uninterested in this course. Such learners accounted for 18.2% of the total number of this experiment. It can be seen from practice that the overall proportion of negative learning states in the classroom is high; second, some learners appear in a positive learning state at the beginning of the course and then gradually tend to be in a neutral learning state. Through the playback of a video course in the linear table of Data Structure, the class content is the last course at the beginning of 9 minutes and 20 seconds, the 21 minutes and 10 seconds is the implementation of the order table, and the 35 minutes is the efficiency analysis of the order table near 35 minutes and 50 seconds. It can be seen that the class as a whole show more difficulties in learning the two knowledge points when realizing the sequential table operation and the efficiency analysis of the sequential table operation.

4.2. Offline Practice

Offline experiments investigated the subjects in the form of questionnaires. For further study, the questionnaire is a registered questionnaire. The first is to evaluate the overall effect of the subject from the subjective perspective; the other is to let the subjects evaluate the overall difficulty of each knowledge point in the course; There are 9 questions in this questionnaire, and the questionnaire is made using "MAC form". The specific content is shown in Figure 6.


《 Data Structure 》Linear table learning situation questionnaire

1.Your study serial number is:

please choose

2.Do you think this course is so easy?(5 The hardest)

1 2 3 4 5

3.How easy does it be to define physical education learning?(5 The hardest)

1 2 3 4 5

4.How easy will sports learning and abstract data type be?(5 The hardest)

1 2 3 4 5

5.How difficult will the sports learning storage structure be?(5 The hardest)

1 2 3 4 5

6.How difficult is it easy to achieve sports movement operation?(5 The hardest)

1 2 3 4 5

7. Efficiency analysis of sports movement operation difficulty degree (5 The hardest)

1 2 3 4 5

8.Example example examples (5 is most difficult)

1 2 3 4 5

9.How do you think you have learned well in this lesson?(5 Best)

1 2 3 4 5

Figure 6: Facial Expression Test Effect

Question 1 in the questionnaire is identity identification in order to facilitate the correspondence between offline experiment results and online experiment results; question 9 is a subjective comprehensive evaluation of the course, and questions 3 to 8 are subjective evaluations of the difficulty of this video course. In total, 22 questionnaires were distributed, 22 were collected, and 22 were valid questionnaires. Grades 1 to 5 represent that the overall difficulty of this course ranged from easy to difficult. In the questionnaire, 40.91 percent of the subjects subjectively thought that the course difficulty was moderate, and only 4.55% thought that the course was relatively easy to learn. At the sixth knowledge point, more than 60% of subjects found the knowledge difficult (the number of scores more than or equal to 4), and nearly 50% of subjects found the knowledge difficult (the number of scores greater than or equal to 4). Most subjects gave a neutral evaluation of the learning effect of this course, while 36.36% considered it poor and poor, 13.64% considered it a good learning effect, and only 9.09% considered it a good learning effect.

4.3. Summary of Practice

Through the analysis and summary of the online experiment and offline experimental data of 22 students, the emotion recognition module for learners learning state recognition agreement average of 72.67%; the experiment shows that the emotion recognition module can be real-time supervision and feedback, obtain the current learning status, and can record the current learners during the whole course, has certain practical value. The experiment uses the emotion recognition module to make the real-time emotion recognition of the online learning state timing map for each learner and draws the positive learning state timing map according to the learning state timing map of each learner. Through the timing diagram, the learning status of the whole class is determined during the whole learning process. After the end of the online experiment, the results of the questionnaire survey for each learner were analyzed and summarized, and the subjects were compared with the online experiment results. By analyzing the recognition results of the emotion recognition module, the questionnaire survey results, and the summary of the two data comparative analysis results, we prove that the emotion recognition module proposed in this paper has high accuracy and certain practical value.

5. Conclusion

Online learning has now been accepted by the public. In particular, as the epidemic sweeps across the country, all universities, middle schools, and primary schools are using various types of online learning systems to carry out normal teaching work, including "Rain Classroom," "Learning Pass," and "Tencent Classroom." In this case, it tests the stability of the current online teaching system and the stability of the system. Present online educational platforms lack the capability to discern students' emotional states, leading to an 'affective disconnect' within virtual classrooms. This shortfall often leads to disengagement and a lack of concentration among a significant number of students. For this problem, this paper investigates how to collect learners' learning emotions in real-time and integrate them into the online classroom. In the future, the emotion recognition module proposed in this paper can be used to provide feedback on the learning status of individuals and collectives to the teachers in real-time, and the teachers who assist online teaching can timely adjust the teaching methods according to the current state of learners, improve the learning efficiency of learners, and promote the overall teaching quality.

REFERENCE

Cheng, W.-H., Lu, J., & Cucchiara, R. (2022). Welcome Message from the General Chairs. In *ICMEW 2022-IEEE International Conference on Multimedia and Expo Workshops 2022, Proceedings*.

Dashtipour, K., Gogate, M., Li, J., Jiang, F., Kong, B., & Hussain, A. (2020). A hybrid Persian sentiment analysis framework: Integrating dependency grammar based rules and deep neural networks. *Neurocomputing*, *380*, 1-10.

Fan, D., Su, X., Weng, B., Wang, T., & Yang, F. (2021). Research progress on remote sensing classification methods for farmland vegetation. *AgriEngineering*, *3*(4), 971-989.

Ge, M., Mao, R., & Cambria, E. (2023). A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, *56*(Suppl 2), 1829-1895.

Gholami, V., Khaleghi, M., Pirasteh, S., & Booij, M. J. (2022). Comparison of self-organizing map, artificial neural network, and co-active neuro-fuzzy inference system methods in simulating groundwater quality: geospatial artificial intelligence. *Water Resources Management*, *36*(2), 451-469.

Haarsa, P. (2024). Solving Bernoulli’s Equations Using Python: Enhancing Student Understanding Through Inquiry-Based Learning. *Unnes Journal of Mathematics Education*, *13*(3), 210-219.

He, Z. (2021). English grammar error detection using recurrent neural networks. *Scientific Programming*, *2021*(1), 7058723.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, *79*(8), 2554-2558.

Hu, T., Xie, Q., Yuan, Q., Lv, J., & Xiong, Q. (2021). Design of ethnic patterns based on shape grammar and artificial neural network. *Alexandria Engineering Journal*, *60*(1), 1601-1625.

Li, J., Ouazzane, K., Kazemian, H. B., & Afzal, M. S. (2013). Neural network approaches for noisy language modeling. *IEEE transactions on neural networks and learning systems*, *24*(11), 1773-1784.

Liu, S. (2022). *Xin Fengxia and the Transformation of China's Ping Opera*. Cambridge University Press.

Lv, Z.-H., Yan, H.-B., & Mei, R. (2018). Automatic and accurate detection of webshell based on convolutional neural network. In *China Cyber Security Annual Conference* (pp. 73-85). Springer.

Medina, A. J., Lorente, M., San Jose, R. J., & Nuviala, N. A. (2021). Construction and Validation of an Observational Analysis Tool in Futsal (OAF-I). *REVISTA INTERNACIONAL DE MEDICINA Y CIENCIAS DE LA ACTIVIDAD FISICA Y DEL DEPORTE*, *21*(84), 639-651.

Mehta, D., Siddiqui, M. F. H., & Javaid, A. Y. (2018). Facial emotion recognition: A survey and real-world user experiences in mixed reality. *Sensors*, *18*(2), 416.

Pal, A., & Singh, D. (2010). Handwritten English character recognition using neural network. *International Journal of Computer Science & Communication*, *1*(2), 141-144.

Pappaterra, M. J. (2022). A Review of Literature and Public Datasets for the Application of Artificial Intelligence in the Railway Industry.

Raizer, V. (2021). *Remote Sensing of Turbulence*. CRC Press.

Stahlberg, F. (2020). Neural machine translation: A review. *Journal of Artificial Intelligence Research*, *69*, 343-418.

Wang, Y., Wang, Y., Liu, J., & Liu, Z. (2020). A comprehensive survey of grammar error correction. *arXiv preprint arXiv:2005.06600*.

Zhang, S., Gao, M., Wang, L., Xu, S., Shao, W., & Kuang, R. (2025). A Malware-Detection Method Using Deep Learning to Fully Extract API Sequence Features. *Electronics*, *14*(1), 167.

Zhou, D.-X. (2018). Deep distributed convolutional neural networks: Universality. *Analysis and applications*, *16*(06), 895-919.