

Evaluation Method of Higher Education Teaching Reform Based on Deep Learning Analysis Technology

Taolin Zhang, University of Sanya, China*
Shuwen Jia, University of Sanya, China
Charoula Angeli, University of Cyprus, Cyprus

ABSTRACT

Considering the shortcomings of large evaluation errors, long time, human, and material resources in the evaluation process of the current college teaching mode to improve the accuracy of the evaluation of college teaching mode and reduce the cost of the evaluation, this study proposes an evaluation method for college teaching methods based on deep learning algorithms. Firstly, the research status of the evaluation of college teaching mode is analyzed, and the reasons for the poor evaluation results of the current college teaching mode are found; then the existing deep learning algorithm is improved, and the effectiveness and speed of the method are verified by comparing with other models. Then, the evaluation model of college teaching mode is established, and machine learning is performed on the evaluation data of college teaching mode; finally, the evaluation data of college teaching mode is collected, and the application example test of college teaching mode evaluation is performed.

KEYWORDS

data collection, deep learning algorithm, evaluation method, University teaching model

INTRODUCTION

Cultivating innovative talents is very important and meaningful for building an innovative society, which has become the consensus of countries all over the world (Hu et al., 2021). Universities are important bases for cultivating innovative talents, and undergraduate education in universities is a key factor in cultivating innovative talent (Zhang, 2020). To improve the quality of undergraduate education, it is necessary to reform and innovate the training mode of undergraduates (Chen et al., 2019).

In recent years, the attempt of improvement of teaching has been continuous, and many innovative educational theories have emerged (Hanandeh et al., 2021). These theories have injected great impetus into teaching reform. Great achievements have been made in the theory and practice of teaching reform, but there is a lack of corresponding research on the quantitative evaluation of reform results. In the whole teaching and learning process, classroom status is an important reference factor for evaluating students' acceptance of courses and teaching quality. In fact, teachers' judgment is the

DOI: 10.4018/IJWLTT.337604

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

main means of classroom state analysis, but it can distract teachers' attention (Guo, 2021). Therefore, it is important to find a method that can replace teachers in conducting classroom state analysis. In addition, the evaluation of the effectiveness of teaching reform also needs to consider factors such as students' test scores and teaching equipment, and these factors also need to be reasonably evaluated.

In response to the evaluation of the college teaching mode, scholars, and experts have invested a lot of time and energy to carry out some research work and put forward many effective methods of college teaching mode evaluation, which can be roughly divided into: qualitative college teaching mode evaluation method, quantitative college teaching mode evaluation method and the evaluation method of college teaching mode, and the combination of qualitative and quantitative evaluation method of college teaching mode (Harun et al., 2020). However, the evaluation results of the college teaching mode by the qualitative method are highly subjective, and the evaluation results are related to the knowledge background or learning experience (Esson & Wang, 2018).

Different evaluators have different evaluation results in the college teaching mode, and it is easy to add their subjective understanding and assumptions to make the evaluation results not credible (Horne, 2020). Quantitative methods mainly include the evaluation method of the college teaching model of the AHP (Analytic Hierarchy Process), the evaluation method of the college teaching mode of weighted fusion, and the evaluation method of the college teaching model of artificial neural network (McCormack, 2020). Among them, the analytic hierarchy process and weighting method can only describe the fixed change law of the college teaching mode, but the actual college teaching mode has a random change law, so using these methods to evaluate the college teaching mode has a large error (Chen & Lu, 2022). Deep learning is an extension of machine learning research and an effective way to implement artificial neural networks. In recent years, deep learning has become the focus of the field of artificial intelligence. Deep learning theory has achieved excellent results in image recognition, data mining. (Zhang, 2021). There are many implementations of deep learning algorithms, such as multilayer neuron self-encoding neural networks, convolutional neural networks, and deep belief networks. The evaluation results of an artificial neural network in the college teaching mode are better than other methods because it belongs to artificial intelligence technology, has good self-learning ability, and can accurately fit the fixed and random change laws (Zhu et al., 2021). Therefore, the use of deep optimization algorithms to evaluate the effectiveness of teaching reform can overcome the shortcomings of the above methods.

Evaluation teaching quality can provide a favorable analysis basis for improving teachers' ability and quality (Lu et al., 2021). It covers a wide range of evaluation indexes and forms large-scale and complex data (Li, 2021). Most teaching quality evaluation systems can only display the teaching quality of teachers and do not have the function of intelligently analyzing teachers' problems. What's more, they can't formulate reasonable suggestions, or these functions are required to be done manually (Fu & Li, 2022). Past case studies have indicated that deep learning algorithms can accurately extract the implicit laws of data and deal with the diversity of complex data (Yu et al., 2021).

The support vector machine method is a deep learning algorithm. There is no limit to the number of training samples, that is, for small samples, it can also accurately model and analyze (Alp et al., 2022). Therefore, it is reasonable for an application deep learning algorithm to design a teaching quality system (Sun et al., 2019). To solve some problems in the evaluation process of the college teaching mode, an evaluation method of the college teaching mode based on a deep learning algorithm and a test of its superiority by comparing it with other college teaching model evaluation methods are proposed in this study.

Finally, by scoring a case of teaching reform, the scores of the intelligent evaluation system proposed in this research case are compared with those of experts. The results show that the method is high-precision and is less time-consuming. The error of the evaluation method is lower than that of other college teaching model evaluation methods, which improves the evaluation efficiency (Chen, 2021).

IMPROVEMENTS IN DEEP LEARNING ALGORITHMS

Deep learning algorithms have become increasingly popular due to their remarkable performance in a variety of tasks. However, deep learning algorithms are still not perfect and face certain challenges. One approach that has gained significant attention is support vector machines (SVMs), which have been shown to be effective in addressing some of the limitations of deep learning algorithms. In this section, we will delve into the relationship between SVMs and deep learning algorithms and explore recent advancements in algorithm improvement and validation.

Support Vector Machines and Deep Learning Algorithms

Deep learning algorithms are currently a popular artificial intelligence technology (Hoch, 2019). Given training samples, deep learning algorithms can build predictive models of problem-solving. Then, the input and output relationship of the problem is effectively fitted by the prediction model, and a high-precision prediction is made. The specific principle is shown in Figure 1.

Support vector machine is one of the most widely used deep learning algorithms (Li et al., 2021). It is modeled based on the principle of structural risk minimization, and the modeling effect is excellent. Therefore, this paper adopted the support vector machine to build the teaching quantity mode.

For training samples $\{x_i, y_i\}, i = 1, 2, \dots, L$, the support vector machine can establish the following linear regression function to estimate the relationship between input and output.

$$f(x) = w \cdot x + b \tag{1}$$

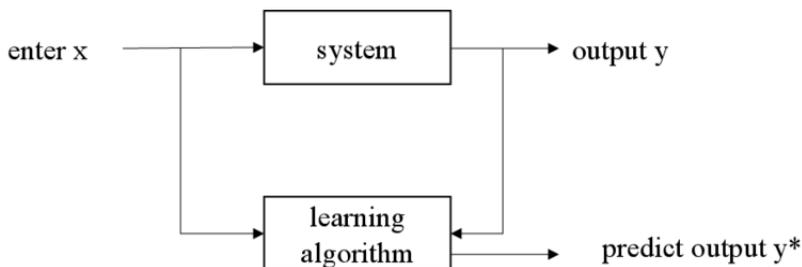
To get the optimal linear regression function, the smallest w must be found, so under the condition of fitting accuracy ϵ , the problem of solving the smallest w is characterized by a optimization convex, namely:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 \\ \text{s.t.} & y_i - w \cdot x_i - b \leq \epsilon \end{aligned} \tag{2}$$

In the modeling process of a support vector machine, relaxation factors ξ_i and ξ_i^* are introduced:

$$\begin{aligned} \min & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L (\xi_i + \xi_i^*) \\ \text{s.t.} & \begin{cases} y_i - w \cdot x_i - b \leq \epsilon + \xi_i \\ y_i - w \cdot x_i - b \geq \epsilon + \xi_i^* \end{cases} \end{aligned} \tag{3}$$

Figure 1. The working principle of the deep learning algorithm



According to dual theory, Formula (3) can be transformed into a quadratic programming problem, and a Lagrangian function is established, which is specifically:

$$L(w, \xi_i, \xi_i^*) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^L (\xi_i + \xi_i^*) - \sum_{i=1}^L \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^L \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle w, x_i \rangle + b) \quad (4)$$

where, α_i and α_i^* represent Lagrange multipliers. Take the partial derivatives for w , b , ξ_i , and ξ_i^* , and set them to 0, that is:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^L (\alpha_i - \alpha_i^*) x_i = 0 \quad (5)$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^L (\alpha_i - \alpha_i^*) = 0 \quad (6)$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i = 0 \quad (7)$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* = 0 \quad (8)$$

Transform the primal problem into a dual problem using the dual of the classical Lagrangian function:

$$\min \frac{1}{2} \sum_{i,j=1}^L (\alpha_i + \alpha_i^*)(\alpha_j + \alpha_j^*) \langle \varphi(x_i), \varphi(x_j) \rangle + \sum_{i=1}^L \alpha_i (\varepsilon + y_i) + \sum_{i=1}^L \alpha_i (\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) - \sum_{i=1}^L \alpha_i^* (\varepsilon + \xi_i^* - y_i + \langle w, x_i \rangle + b) \quad (9)$$

According to Formula (9), a set of optimal values of Lagrange multipliers α_i and α_i^* can be obtained:

$$w = \sum_{i \in SV} (\alpha_i - \alpha_i^*) \cdot x_i \quad (10)$$

where SV represents the samples corresponding to $\alpha_i^* \neq 0$ and $\alpha_i \neq 0$. The above process is to solve the linear regression problem. For the nonlinear regression problem, the nonlinear function is first used to map the samples to the high-dimensional feature space, and the linear regression is performed, so that Formula (10) becomes:

$$w = \sum_{i \in SV} (\alpha_i - \alpha_i^*) \cdot \varphi(x_i) \quad (11)$$

The kernel function is used to replace the inner product operation: $K(x_i, x_j) = (\varphi(x_i), \varphi(x_j))$, and finally, the regression form of the support vector machine is obtained as:

$$\begin{aligned} f(x) &= \sum_{i=1}^L (\alpha_i - \alpha_i^*) \langle \varphi(x_i), \varphi(x) \rangle + b \\ &= \sum_{i=1}^L (\alpha_i - \alpha_i^*) K(x_i, x) + b \end{aligned} \quad (12)$$

Different kernel functions can be selected to establish different support vector machine regression forms. The current kernel functions mainly have the following forms:

$$K(x_i, x) = \left(\langle x_i, x \rangle + C \right)^p \quad (13)$$

$$K(x_i, x) = \exp \left(-\frac{\|x_i - x\|^2}{2\sigma^2} \right) \quad (14)$$

$$K(x_i, x) = \tanh \left(v(x_i, x) + C \right) \quad (15)$$

A large number of research results show that these functions have their own advantages, but there are also obvious defects which affect the modeling performance of SVMs. Therefore, this paper chose the wavelet function as the kernel function of SVMs. For the wavelet base function $\psi(x)$, after performing m translation and a scale transformation operations on it, the wavelet function is obtained as:

$$\psi_{a,m}(x) = |a|^{\frac{1}{2}} \psi \left(\frac{x - m}{a} \right) \quad (16)$$

For any function, the continuous wavelet transform should be described as:

$$W_f(a, m) = \langle f, \psi_{a,m} \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(x) \left(\frac{x-m}{a} \right) dt \quad (17)$$

In practical application, since the signal processed by the computer is a digital signal, the continuous wavelet should be discretized, then the discrete wavelet coefficient is:

$$C_{j,k} = \int f(t) \psi_{j,k}^*(t) dt \quad (18)$$

According to the Mercer condition, the wavelet function of the support vector machine kernel function is obtained as:

$$\psi(x) = \cos(1.75x) \exp\left(-\frac{x^2}{2}\right) \quad (19)$$

The wavelet support vector machine is:

$$f(x) = \sum_{i=1}^n (a_i - a_i^*) \psi(x) + b \quad (20)$$

Algorithm Improvement and Validation

Assume a multiclass problem with K classes, the total number of samples equal to n , and each class contains the same number of samples. A SVM is actually a quadratic programming problem, and its training time T has a nonlinear growth relationship with the number of training samples n . The specific mathematical description is:

$$T = cn^2 \quad (21)$$

where c is a constant. Since the OVR(One-vs-Rest) method needs to construct a classifier, and all training samples must participate in the construction process of the optimal hyperplane of each classifier, the training time is:

$$T = ckn^2 \quad (22)$$

For the OVO(One-vs-One) method, the number of SVM classifiers to be constructed is $k(k-1)$, and the number of training samples participating in each classifier is $2n/k$, so the training time is:

$$T = c \left(\frac{k(k-1)}{2} \right) \left(\frac{2n}{k} \right)^2 = \frac{2cn^2(k-1)}{k} \approx 2cn^2 \quad (23)$$

It is worth noting that the training time of the DAG(Directed Acyclic Graph) method has the same form as the OVO. For the binary tree method, regardless of its structure, the number of SVM classifiers

to be constructed is $k - 1$, but the number of training samples for each classifier is not necessarily the same. In the complete binary tree method, there are 2^{i-1} classifiers on the i layer. Assuming the same number of classifier samples on the same layer, the training time of this method is:

$$T = c \sum_{i=1}^h \left(2^{i-1} \left(\frac{n}{2^{i-1}} \right)^2 \right) = cn^2 \left(1 + \frac{1}{2} + \dots + \frac{1}{2^{h-1}} \right) \approx 2cn^2 \quad (24)$$

For the partial binary tree method, the number of categories contained in the classifier of the i -th layer is $(k - i + 1)$. Assuming that each category contains the same number of samples, the total training time is:

$$T = \sum_{i=1}^{k-1} c \left(\frac{n(k-i+1)}{k} \right)^2 \approx \frac{ckn^2}{3} \quad (25)$$

The incomplete binary tree method proposed in this paper has a depth between the complete binary tree and the partial binary tree, so the complexity range of the algorithm is $[2cn^2, ckn^2 / 3]$. It can be seen from the above analysis and comparison that when k is relatively large, the time cost of the incomplete SVM multiclassification algorithm in the training phase is smaller than that of the OVR method but greater than that of the OVO method. In the testing phase, the time cost is the smallest because the number of classifiers that the test sample needs to go through is the least, at most, $k - 1$.

In order to verify the comprehensive performance of the improved incomplete multiclassification algorithm, this paper selected the Glass Identification and Vowel data sets in the UCI(University of California, Irvine) database and the Letter, Satimage and Vehicle data sets in the Statlog database, a total of five sets of data for experiments (Fu & Li, 2022). The detailed information is listed:

- 1) Glass data set: Contains six categories, 114 training samples, 100 test samples, and each sample has nine attributes.
- 2) Vehicle data set: A data set about cars, the number of categories, training samples, test samples, and dimensions are 4, 592, 254, and 8, respectively.
- 3) Vowel data set: It has 11 categories, a large number, 13 attributes, and a total of 846 samples, of which 592 are training samples.
- 4) Satimage data set: The number of categories included is the same as that of the Glass data set, both of which are six, but it has 36 attributes and 6,435 samples (2,000 test samples, much larger than the Glass data set).
- 5) Letter data set: It contains the largest number of categories and samples, 26 and 20,000, respectively, and the number of attributes is 16.

Table 1 summarizes the basic information of these five groups of resupply data sets.

In general, the SVM classifier constructed with radial basis (BRF) kernel function can achieve the highest classification accuracy. In addition, it has only two parameters, C and r . In the training stage, the grid search method can be used to select parameters, and then use k -fold cross-validation to find the optimal combination of these two parameters. The structure is convenient to adjust and operate and is simple. Therefore, the algorithm in this paper also selected the kernel function. This research used the OVR method, OVO method, and the improved SVM algorithm, selected five data

Table 1. Basic information of the simulation experiment data set (Jian, 2019)

Data set	Categories	Attributes	Training samples	Test samples
Glass	6	9	114	100
Vehicle	4	18	592	254
Vowel	11	13	693	297
Satimage	6	36	4,435	2,000
Letter	26	16	15,000	5,000

sets and conducted experiments 10 times, respectively, and recorded the classification time and accuracy of them. Figures 2, 3, and 4 show the average value of 10 experiments. When training the binary tree classification model, the range of parameters C and r is set to be $(2^{-5}, 2^{10})$, and the preorder traversal method is used to call the grid optimization method at the nonleaf nodes of the binary tree to automatically select the optimal parameter combination. The final accuracy of the classification results is shown in Figure 2.

From Figure 2, we can conclude that:

- 1) Since the number of categories in the data set Vehicle is small and the number of corresponding combinations is small, the improved algorithm cannot give full play to its advantages. Therefore, the accuracy rates of these three algorithms are almost the same.
- 2) For the Glass and Satimage data sets, the number of categories is six, and the difference in the accuracy of the three algorithms becomes larger.
- 3) For the Vowel and Letter data sets, which contain more than 10 categories, it can be seen from the figure that the improved algorithm in this paper can outperform the best OVR and OVO algorithms by two percentage points.

Figure 2. Accuracy of classification by the three multiclassification algorithms

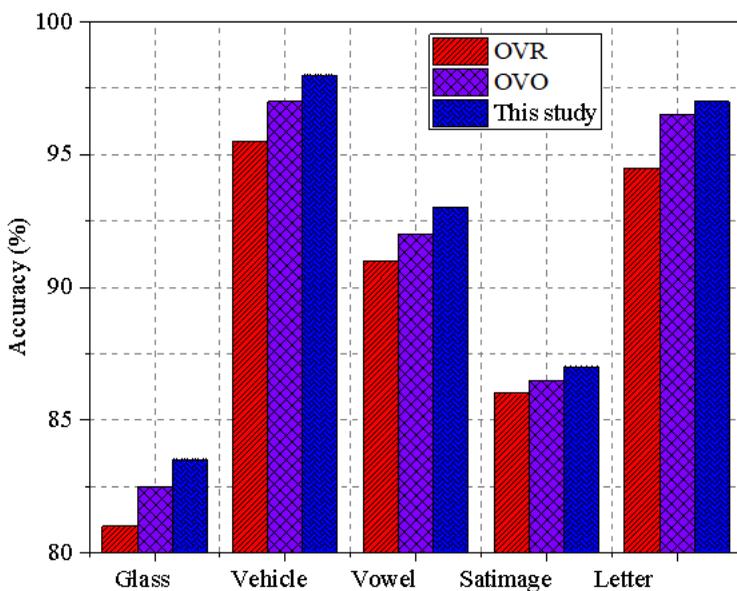
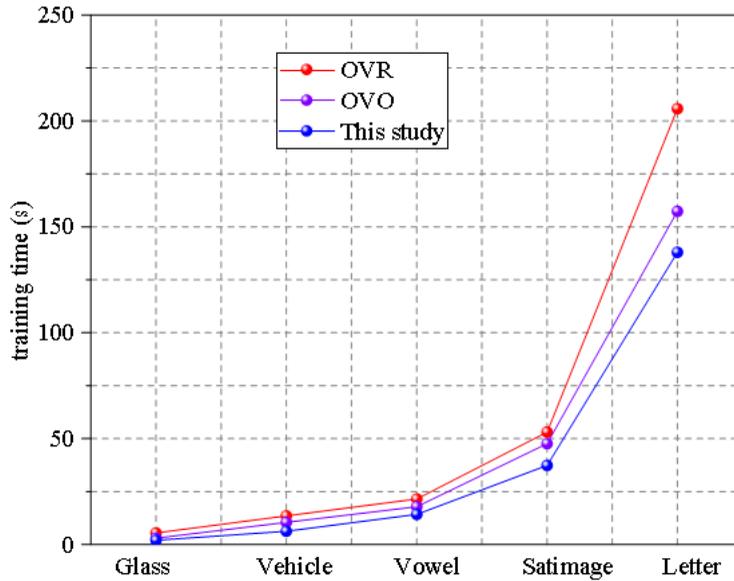
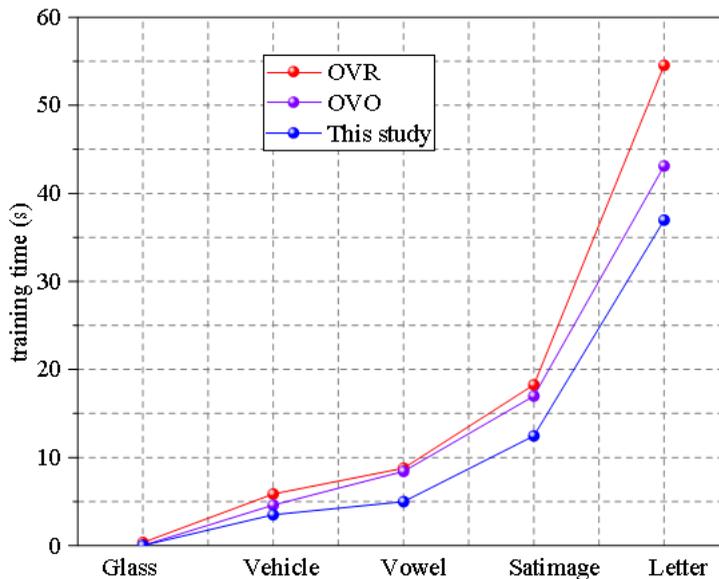


Figure 3. Comparison of training time of three multiclassification algorithms



To sum up, in terms of accuracy, the improved algorithm is better than the OVR and OVO algorithms. In the classification problem, the classification time is another important criterion for evaluating the comprehensive performance of the algorithm, and the most important indicator is the training time. In order to better compare these three algorithms, this simulation experiment counted their training time and test time, respectively. Figures 3 and 4 show the time consumed by the three algorithms in the training phase and the testing phase, respectively.

Figure 4. Comparison of test times for three multiclassification algorithms



From the experimental results in Figures 5 and 6:

- 1) At present, there is no theory for determining some parameters in the support vector machine-based multiclassification algorithm. In order to find a more reasonable parameter, the method that can be used is to conduct multiple trials within a certain value range and select the better one.
- 2) When training the classifier, although the number of classifiers in the method is small, each classifier requires all samples to participate in the construction of the optimal hyperplane, and the solution process takes a lot of time. Although the number of classifiers the OVO method

Figure 5. Teaching quality evaluation system

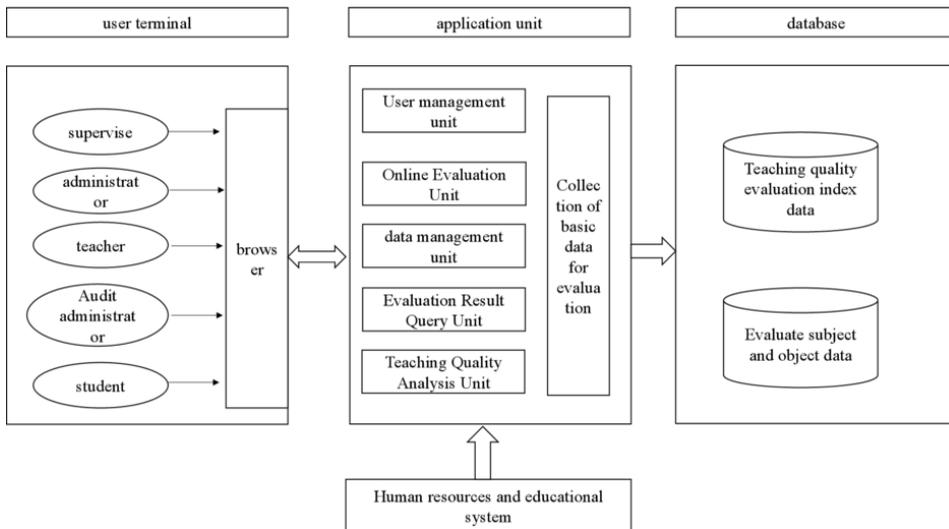
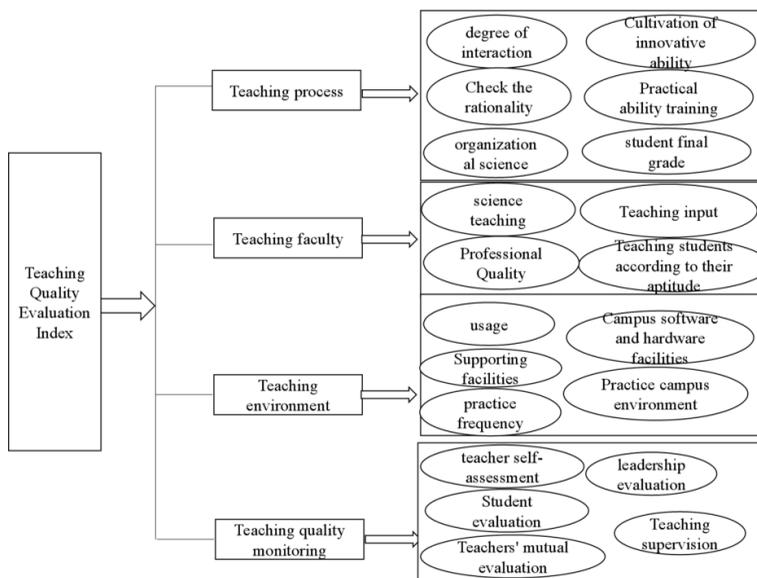


Figure 6. Teaching quality evaluation index system



needs to construct is large, each classifier only needs samples of one class to participate in the solution process, which makes the final overall time less time. In addition, since the relative distance-based incomplete binary tree support vector machine multiclassification method needs to calculate the relative distance between each two categories, it takes longer than the OVO method in the training phase but less than the OVR method. In the testing phase, the binary classification tree constructed by the algorithm in this paper is improved in both the depth of the tree and the arrangement of the structure. The number of classifiers that the test sample needs to go through is the least, so the test time used shortest.

- 3) The increase in the number of training texts will increase the computational complexity of the classification process, making the training time and testing time of the classification algorithm longer.

In general, the incomplete SVM algorithm based on relative distance improves both the time consumption of the classification process and the accuracy of the classification results, which shows that the algorithm is effective and feasible.

EVALUATION OF THE COLLEGE TEACHING MODE BY THE DEEP LEARNING ALGORITHM

The Architecture of an Evaluation System

The architecture of an evaluation system is designed based on the B/S mode, which is a deep learning algorithm, as shown in Figure 5. The advantages of designing a teaching quality evaluation system based on the B/S model are: it is convenient for different types of users to operate, an online evaluation can be completed in a short period of time, and system maintenance is convenient. Client, application unit, and database are three important components of the system. The client side includes five types of users: supervisors, administrators, teachers, review administrators, and students. Different types of user interfaces are combined with browsers to display content, such as page operations. The system application unit covers five aspects: user management, online evaluation, data management, evaluation result query, and teaching quality analysis.

Construction of an Evaluation Index

Compared with other evaluation problems, the evaluation of the college teaching mode is more complicated (Feng, 2020). In order to carry out a high-precision evaluation of the teaching model of colleges and universities, the evaluation index system of the college teaching mode should be established following the principles of consistency, scientificity, objectivity, systematicness, and operability. The teaching process, environment, teachers, and quality monitoring are the first-level indexes of the index system. There are 20 second-level indexes under the first-level indexes. Figure 6 shows the detailed information.

The Evaluation Method of the College Teaching Mode Based on a Deep Learning Algorithm

Learning the inherent laws and representation levels of sample data is the function of deep learning (Ma et al., 2020). When designing the teaching quality evaluation system, the support vector machine model in the deep learning algorithm is used to evaluate the teaching quality. The first step is to build a teaching quality evaluation index system; secondly, the support vector machine model is applied to learn training samples, and the learning samples are samples for expert teaching quality evaluation. After learning the training samples, a teaching quality evaluation model is constructed. Finally, the test samples are input to start evaluation and analysis. The specific process is as follows:

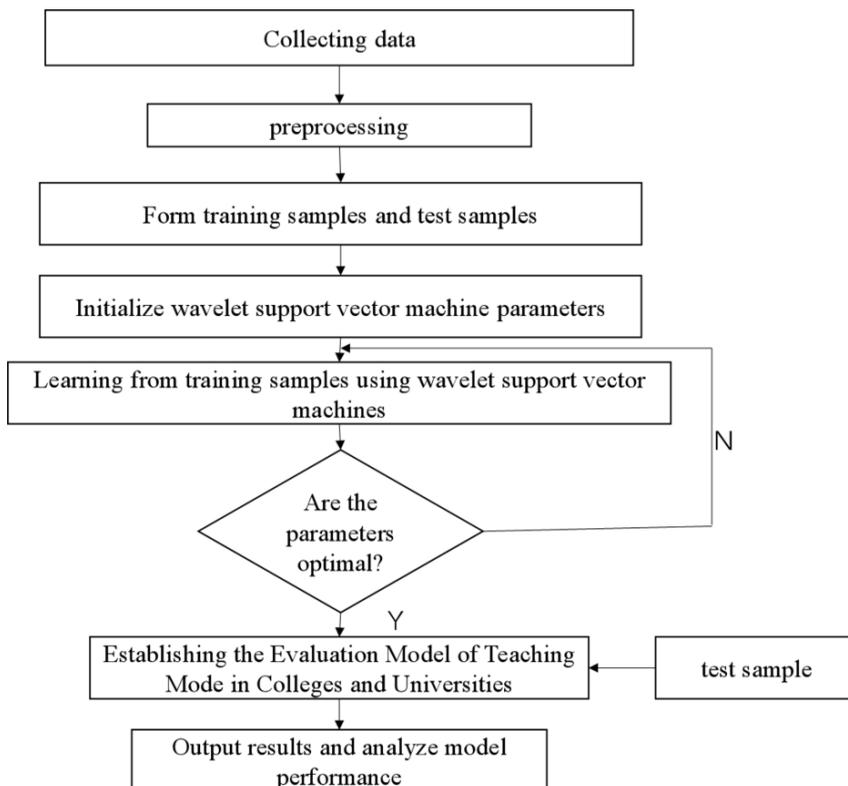
- 1) Collect historical data of the college teaching model evaluation and preprocess them, such as removing singular values, normalizing operations, to prevent them from adversely affecting the learning process of support vector machines. The normalization formula is specifically expressed as:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (26)$$

- 2) Randomly select part of the data from the preprocessed historical data of the college teaching mode evaluation to form a training sample set and establish a corresponding test sample.
- 3) Set the relevant parameters of the wavelet support vector machine, such as the initial values of α_i, α_i^*, b , and randomly initialize the translation factor and scale transformation factor.
- 4) The evaluation accuracy of the university teaching mode is used as the objective function of the support vector machine training, and the optimal α_i, α_i^*, b are obtained through continuous learning and training.
- 5) According to the optimal α_i, α_i^*, b , a support vector machine evaluation model of the college teaching mode is established.
- 6) Use the evaluation model to analyze the test samples and output the results.

To sum up, the evaluation process of college teaching mode is shown in Figure 7.

Figure 7. Evaluation process of the college teaching mode based on a deep learning algorithm



SIMULATION TEST OF THE TEACHING EFFECT IN COLLEGES

The evaluation of teaching reform is divided into two parts. First, we need mobile phone data, and second, we need to analyze these data in order to understand the effectiveness and practicability of the teaching evaluation method based on deep learning, as well as the shortcomings of the evaluation method (Ahmadi & Khashei, 2021). This section first takes the feedback of students in class as an example to carry out experiments to illustrate the application of deep learning algorithms in data collection. Then, the evaluation of the teaching results by the deep learning algorithm is compared with the memory of several experts to verify the reliability of the method.

Classroom Trials and Data Collection

In the experiment, the face recognition technology in the deep learning algorithm was used in the evaluation of students' engagement in class, and interviews were conducted with students after class for comparison. Through the quasiexperimental method, it is verified that the evaluation system can help teachers to realize the sentiment analysis of individual students, and at the same time, it can also realize the analysis of the learning status of students, including students' attention, participation, and difficulty, and the recognition results have high accuracy. After the experiment, we used the form of an interview to further understand the feedback of teachers and students on the use of the intelligent, emotional evaluation system. It is found the intelligent, emotional evaluation system can save manpower and material resources while helping teachers to better understand students' learning status. In order to provide timely help for students, it is meaningful to improve students' academic performance and improve classroom teaching.

The subjects of this study are 36 freshman students from a university. The reason for choosing freshman students is that students at this stage face a transition from high school to university, which seriously affects their physical and mental development. Students at this stage need the most attention from teachers, especially emotional support. However, teachers also face the problem of high pressure and heavy tasks, resulting in them not paying too much attention to the emotional state of students. To this end, we introduce deep learning algorithms into college classrooms to help teachers obtain emotional information from students, then master their learning status and students' inner emotional changes, and provide students with cognitive and emotional help in a timely manner so that students can study healthily and happily. The experimental arrangement is shown in Table 2.

Table 2. The experimental arrangement

Before experiment	<ol style="list-style-type: none"> 1. Check the hardware and software facilities of the intelligent, emotional teaching evaluation system. We needed to check the various software and hardware devices that constitute the intelligent, emotional teaching evaluation system, including cameras, monitors, hosts, switches, and other hardware devices. 2. Check whether the system is easy to use. Before starting the experiment, several simulation experiments were conducted to compare the system identification results with the manual monitoring results to ensure that the similarity between the two was above 80%. Otherwise, the system needs to be debugged repeatedly until the accuracy rate reaches the standard. 3. Teach users how to use the intelligent, emotional teaching evaluation system and introduce the relevant functions of the system.
Under experiment	<ol style="list-style-type: none"> 1. Teachers will teach as usual, without any interference with the teaching process, and ensure that experiments are carried out in a natural environment. 2. Teachers use the system to achieve two functions in the process of teaching, including the evaluation of the learning status of the designated students and the evaluation and analysis of the student group, and pay attention to the prompts of the system and adjust the teaching strategy appropriately.
After experiment	Analyze the data of the system evaluation results. Interviews were used to communicate with teachers and students to obtain their feedback on the emotional evaluation system.

To demonstrate the accuracy and correctness of an intelligent evaluation system, we played back the video of this course after class and used the structured observation method to count the number of participants, followers, and doubts at each time point and then calculated the number of participants at each time point. The degree of participation, attention, and doubt of the points is compared and then compared with the system detection results, and the accuracy of the system detection was calculated with reference to the manual statistical results.

Table 3 shows the comparison results of attention every 5 minutes between the start of class at 10:00 and the end of class at 10:45a.m. Table 4 below is the comparison result of participation, and Table 5 is the comparison result of difficulty. Figures 8, 9, and 10 are the line charts for the comparison of the three parameters of attention, participation, and difficulty. From the figures, we can see that the system is basically consistent with the actual test results. For further verification, we calculated the accuracy rates of the three parameters: 94%, 94%, and 91%, and the accuracy rates are all higher than 90%, indicating that the detection results of this system are highly accurate and have strong practical value.

Totally speaking, an intelligent, emotional teaching evaluation system can realize the emotional tracking of individual students and understand the learning status of individual students; it can calculate the attention, participation, and difficulty of the student group and output the statistical results in a visual form, which is convenient for teachers to refer to. Through feedback from teachers and students, it is verified that the system is easy to operate and has a high accuracy of recognition results. It can provide feedback on students' learning status in real-time, prompt teachers to grasp classroom teaching content and students' learning situation in time, save manpower and material resources, and realize the strong practical value of classroom teaching.

Analysis of Teaching Effectiveness

In the process of evaluating teaching effectiveness, it is not only necessary to collect the data on the teaching situation in the classroom but also to evaluate the teaching from the whole process and

Table 3. Contrast results of attention

Time	10:05	10:10	10:15	10:20	10:25	10:30	10:35	10:40	10:45
Actual number	26	25	23	10	12	30	32	30	33
Attention (real)	72.22%	69.44%	63.89%	27.78%	33.33%	83.33%	88.89%	83.33%	91.67%
Detected number	24	24	22	9	11	29	30	32	32
Attention (department)	66.67%	66.67%	61.11%	25.00%	30.56%	80.56%	83.33%	88.89%	88.89%
Error rate	0.08	0.04	0.04	0.10	0.08	0.03	0.06	0.07	0.03

Table 4. A comparison of engagement

Time	10:05	10:10	10:15	10:20	10:25	10:30	10:35	10:40	10:45
Actual number	20	19	18	15	10	25	27	26	28
Participate (real)	55.56%	52.78%	50.00%	41.67%	27.78%	69.44%	75.00%	72.22%	77.78%
Detected number	19	17	16	14	9	26	27	27	28
Participate (department)	52.78%	47.22%	44.44%	38.89%	25.00%	72.22%	75.00%	75.00%	77.78%
Error rate	0.05	0.11	0.11	0.07	0.10	0.04	0.00	0.04	0.00

Figure 8. Attention contrast diagram

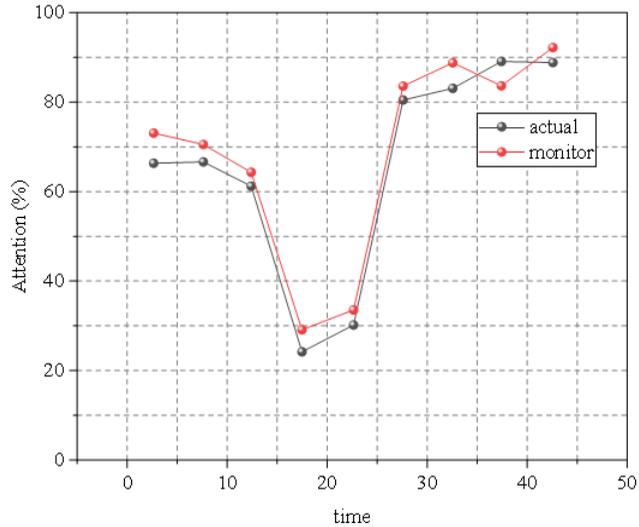
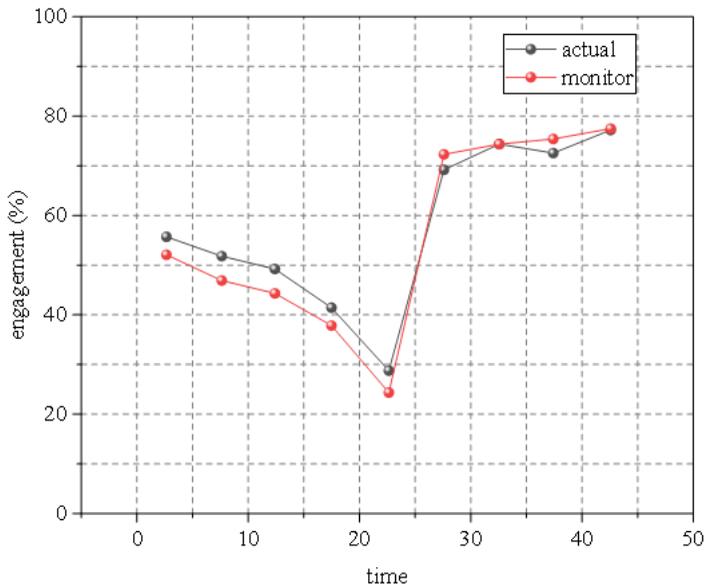


Figure 9. Engagement comparison graph



multiple aspects. This section evaluates teaching from the aspects of the process, teachers, environment, and monitoring of teaching quality.

Scoring by experts is highly subjective and has a strong relationship with the knowledge of experts. To verify the reliability of the design method, 10 experts scored the same course from different angles and compared these with the scores given by the program designed by the deep learning algorithm. Table 6 lists the information.

Table 5. The comparison of the difficulty

Time	10:05	10:10	10:15	10:20	10:25	10:30	10:35	10:40	10:45
Actual number	5	1	4	4	5	3	3	4	3
Doubtful (real)	13.89%	2.78%	11.11%	11.11%	13.89%	8.33%	8.33%	11.11%	8.33%
Detected number	5	1	3	4	4	3	3	4	2
Doubtful (department)	13.89%	2.78%	8.33%	11.11%	11.11%	8.33%	8.33%	11.11%	5.56%
Error rate	0.00	0.00	0.25	0.00	0.20	0.00	0.00	0.00	0.33

Figure 10. Difficulty comparison diagram

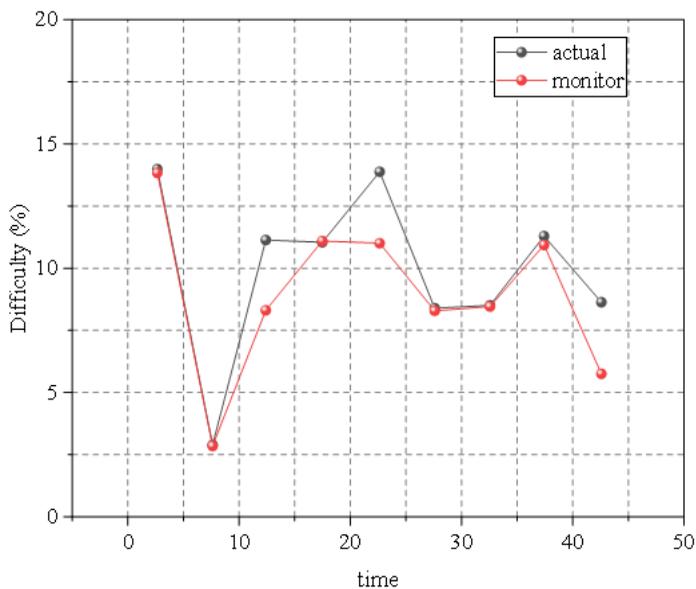


Table 6. Comparison expert scoring and algorithmic program scoring

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	ave	P
T1	86.57	94.71	94.57	93.00	94.16	89.85	86.41	92.92	94.59	89.21	22.90	22.71
T2	95.51	90.55	93.06	94.99	96.91	97.59	93.47	89.39	89.49	90.58	23.29	24.15
T3	91.41	85.54	85.44	92.29	91.14	86.50	84.97	85.51	89.16	87.73	21.99	22.24
T4	87.61	87.43	91.25	93.23	94.47	88.04	91.55	90.76	87.10	89.70	22.53	21.37
ave	90.27	89.56	91.08	93.38	94.17	90.50	89.10	89.64	90.08	89.30	90.71	90.47

In Table 6, T1–T4 represent different items for scoring, and E1–E10 represent 10 different experts. From Table 6, it can be seen that the subjectivity of the experts’ scores is strong, which will cause large deviations. For example, there is a difference of seven points between the Expert 1 score and the Expert 5 scores in the T4 project. In the total score, the difference between the highest score and the lowest score is also more than five points. Compared with the average score of 10 experts,

whether it is the score of a single item or the total score, the difference between the score of the deep learning method and the experts is within one point. This illustrates the reliability of the deep learning algorithm scoring.

CONCLUSION

Combining the current situation and needs of college education, this study adopted the method of combining theoretical research with practice. First, it analyzed the concept, type, evaluation method, and applied learning theory of teaching evaluation. This study summarized the development of teaching evaluation in colleges and demonstrates the theoretical significance of this research. It is found that the existing teaching evaluation methods are labor-intensive and time-consuming. Then, combined with the reality of teaching evaluation in colleges, a method of evaluating college teaching by using deep learning algorithm is proposed, and the existing methods are improved to make the evaluation faster and more accurate. And compared with other methods, it proved the rapidity and reliability of this method in modeling and calculation. Finally, in order to solve the urgent need for special evaluation methods in current college teaching, based on deep learning, a college teaching evaluation system was established, and the system will analyze the teaching effect in a timely, detailed, and specific manner. The reliability of the method was illustrated by comparing the results of the deep learning algorithm evaluation with the evaluation results of 10 experts. The proposal of this method is helpful to improve the teaching quality of colleges and promote the development of college education.

AUTHOR NOTE

TaoLin Zhang was born in Shanghai, P.R. China, in 1982. She received a master's degree from Shaanxi University of Science & Technology, P.R. China. Now, she works in the School of Teaching Management, University of Sanya. Her research interests include higher education management and teaching operation management.

ShuWen Jia was born in Shandong, China, in 1981. He received a BS degree and MS degree in educational technology from Qufu Normal University, Qufu, China, in 2006 and 2009, respectively. He is currently a senior engineer at Saxo Fintech Business School. His research interests include the application of multimedia technology and fog computing.

Charoula Angeli is a professor in the Department of Education, University of Cyprus, with research interests in the development and application of machine learning techniques for the evaluation of college teaching modes.

The figures and tables used to support the findings of this study are included in the article.

The authors declare that they have no conflicts of interest.

This work was not supported by any funds.

The authors would like to show sincere thanks to those techniques who have contributed to this research.

REFERENCES

- Ahmadi, M., & Khashei, M. (2021). Generalized support vector machines (GSVMs) model for real-world time series forecasting. *Soft Computing*, 25(22), 14139–14154. <https://link.springer.com/article/10.1007/s00500-021-06189-z>. doi:10.1007/s00500-021-06189-z
- Alp, S., Yilmaz, F., & Geçici, E. (2022). Evaluation of the quality of health and safety services with SERVPERF and multi-attribute decision-making methods. *International Journal of Occupational Safety and Ergonomics*, 28(4), 2216–2226. doi:10.1080/10803548.2021.1984711 PMID:34556013
- Chen, J., & Lu, H. (2022). Evaluation method of classroom teaching effect under intelligent teaching mode. *Mobile Networks and Applications*, 27(3), 1262–1270. <https://link.springer.com/article/10.1007/s11036-022-01946-2>. doi:10.1007/s11036-022-01946-2
- Chen, S., Cao, J., & Huang, Z. (2019). Weighted linear loss projection twin support vector machine for pattern classification. *IEEE Access : Practical Innovations, Open Solutions*, 7, 57349–57360. doi:10.1109/ACCESS.2019.2914022
- Chen, Y. (2021). College English teaching quality evaluation system based on information fusion and optimized RBF neural network decision algorithm. *Journal of Sensors*, 2021, 1–9. doi:10.1155/2021/1345759
- Esson, J., & Wang, K. (2018). Reforming a university during political transformation: A case study of Yangon University in Myanmar. *Studies in Higher Education*, 43(7), 1184–1195. doi:10.1080/03075079.2016.1239250
- Feng, Y. (2020). An evaluation method of PE classroom teaching quality in colleges and universities based on grey system theory. *Journal of Intelligent & Fuzzy Systems*, 38(6), 6911–6915. doi:10.3233/JIFS-179769
- Fu, J. F., & Li, J. (2022). Teaching quality evaluation model of a flipped classroom in colleges and universities based on support vector machine. *Wireless Communications and Mobile Computing*, 2022, 1–12. doi:10.1155/2022/6889564
- Guo, H. (2021). Research on the influence of university education system reform on college students' innovation ability. *International Journal of Electrical Engineering Education*, 00207209211002082. doi:10.1177/00207209211002082
- Hanandeh, R., Alnajdawi, S. M., Almansour, A., & Elrehail, H. (2021). The impact of entrepreneurship education on innovative start-up intention: The mediating role of entrepreneurial mind-sets. *World Journal of Entrepreneurship, Management and Sustainable Development*, 17(4), 856–871. <https://www.emerald.com/insight/content/doi/10.1108/WJEMSD-02-2020-0016/full/html>. doi:10.1108/WJEMSD-02-2020-0016
- Harun, H., Wardhaningtyas, S., Khan, H. Z., An, Y., & Masdar, R. (2020). Understanding the institutional challenges and impacts of higher education reforms in Indonesia. *Public Money & Management*, 40(4), 307–315. doi:10.1080/09540962.2019.1627063
- Hoch, J. C. (2019). If machines can learn, who needs scientists? *Journal of Magnetic Resonance (San Diego, Calif.)*, 306, 162–166. doi:10.1016/j.jmr.2019.07.044 PMID:31362904
- Horne, J. (2020). Mass education and university reform in late twentieth century Australia. *British Journal of Educational Studies*, 68(5), 671–690. doi:10.1080/00071005.2020.1804526
- Hu, W., Hu, Y., Lyu, Y., & Chen, Y. (2021). Research on integrated innovation design education for cultivating the innovative and entrepreneurial ability of industrial design professionals. *Frontiers in Psychology*, 12, 693216. doi:10.3389/fpsyg.2021.693216 PMID:34484043
- Jian, Q. (2019). Multimedia teaching quality evaluation system in colleges based on genetic algorithm and social computing approach. *IEEE Access : Practical Innovations, Open Solutions*, 7, 183790–183799. doi:10.1109/ACCESS.2019.2957447
- Li, B., Fei, Y., & Liu, H. (2021). An artificial intelligence based model for evaluation of college students' ability and characteristics through teaching evaluation. *Journal of Intelligent & Fuzzy Systems*, 40(2), 3397–3407. doi:10.3233/JIFS-189378
- Li, N. (2021). A fuzzy evaluation model of college English teaching quality based on analytic hierarchy process. [IJET]. *International Journal of Emerging Technologies in Learning*, 16(2), 17–30. doi:10.3991/ijet.v16i02.19731

- Lu, C., He, B., & Zhang, R. (2021). Evaluation of English interpretation teaching quality based on GA optimized RBF neural network. *Journal of Intelligent & Fuzzy Systems*, 40(2), 3185–3192. doi:10.3233/JIFS-189357
- Ma, F., Zhu, Z., Zhou, M., & Pan, W. T. (2020). Fuzzy comprehensive evaluation of classroom teaching quality of college teachers. *International Journal of Electrical Engineering Education*, 0020720920928540. doi:10.1177/0020720920928540
- McCormack, C. F. (2020). Society, science and institutionally-embodied higher education reform in nineteenth-century Ireland: The role of mobile, professional elites in fashioning reform. *History of Education*, 49(1), 18–37. doi:10.1080/0046760X.2019.1659426
- Sun, S., Cao, Z., Zhu, H., & Zhao, J. (2019). A survey of optimization methods from a machine learning perspective. *IEEE Transactions on Cybernetics*, 50(8), 3668–3681. <https://arxiv.org/abs/1906.06821>. doi:10.1109/TCYB.2019.2950779 PMID:31751262
- Yu, D., Kou, G., Xu, Z., & Shi, S. (2021). Analysis of collaboration evolution in AHP research: 1982–2018. *International Journal of Information Technology & Decision Making*, 20(1), 7–36. doi:10.1142/S0219622020500406
- Zhang, J. (2021). Research on classroom teaching evaluation and instruction system based on GIS mobile terminal. *Mobile Information Systems*, 2021, 1–11. doi:10.1155/2021/6557937
- Zhang, J. L. (2020). The application of human comprehensive development theory and deep learning in innovation education in higher education. *Frontiers in Psychology*, 11, 1605. doi:10.3389/fpsyg.2020.01605 PMID:32793040
- Zhu, Y., Xu, J., & Zhang, S. (2021). Application of optimized GA-BPNN algorithm in English teaching quality evaluation system. *Computational Intelligence and Neuroscience*, 2021, 1–9. doi:10.1155/2021/4123254 PMID:35003243

TaoLin Zhang was born in Shanghai, P.R. China, in 1982. She received the Master's degree from Shaanxi University of Science & Technology, P.R. China. Now, she works in School of Teaching management, University of Sanya. Her research interests include Higher Education Management and teaching operation management .

ShuWen Jia, was born in ShanDong, China, in 1981. received the B.Sc degree and M.Sc degree in educational technology from Qufu Normal University, Qufu, China, in 2006 and 2009, respectively. He is currently a senior engineer at Saxo Fintech Business School. His research interests include the application of multimedia technology and fog computing.

Charoula Angeli is a professor at the Department of Education, University of Cyprus, with research interests in the development and application of machine learning techniques for the evaluation of college teaching modes.