



OPEN ACCESS

EDITED BY

Ruey-Shun Chen,
National Chiao Tung University, Taiwan

REVIEWED BY

Dairi Abdelkader,
Oran University of Science and
Technology – Mohamed
Boudiaf,
Algeria

Sindhu C.,
SRM Institute of Science and Technology,
India
Alladoubaye Ngueilbaye,
Shenzhen University,
China

*CORRESPONDENCE

Lele Ye
1501269925@qq.com
Min Chen
xuancm@gmail.com

SPECIALTY SECTION

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 08 May 2022

ACCEPTED 28 June 2022

PUBLISHED 02 September 2022

CITATION

Zhang R, Yao X, Ye L and Chen M (2022)
Students' adaptive deep learning path and
teaching strategy of contemporary ceramic
art under the background of Internet +.
Front. Psychol. 13:938840.
doi: 10.3389/fpsyg.2022.938840

COPYRIGHT

© 2022 Zhang, Yao, Ye and Chen. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Students' adaptive deep learning path and teaching strategy of contemporary ceramic art under the background of Internet +

Rui Zhang¹, Xianjing Yao², Lele Ye^{3*} and Min Chen^{4*}

¹School of Art and Design, Xinyang Normal University, Xinyang, China, ²College of Cultural Relics and Art, Hebei Oriental University, Langfang, China, ³Zhejiang College of Zhejiang University of Technology, Shaoxing, China, ⁴School of Business, Wenzhou University, Wenzhou, China

With the rapid expansion of Internet technology, this research aims to explore the teaching strategies of ceramic art for contemporary students. Based on deep learning (DL), an automatic question answering (QA) system is established, new teaching strategies are analyzed, and the Internet is combined with the automatic QA system to help students solve problems encountered in the process of learning. Firstly, the related theories of DL and personalized learning are analyzed. Among DL-related theories, Back Propagation Neural Network (BPNN), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) are compared to implement a single model and a mixed model. Secondly, the collected student questions are selected and processed, and experimental parameters in different models are set for comparative experiments. Experiments reveal that the average accuracy and Mean Reciprocal Rank (MRR) of traditional retrieval methods can only reach about 0.5. In the basic neural network, the average accuracy of LSTM and GRU structural models is about 0.81, which can achieve better results. Finally, the accuracy of the hybrid model can reach about 0.82, and the accuracy and MRR of the Bidirectional Gated Recurrent Unit Network-Attention (BiGRU-Attention) model are 0.87 and 0.89, respectively, achieving the best results. The established DL model meets the requirements of the online automatic QA system, improves the teaching system, and helps students better understand and solve problems in the ceramic art courses.

KEYWORDS

Internet +, ceramic art, deep learning, teaching strategy, personalized learning

Introduction

With the rapid development of the Internet, there is more and more integration of different fields with the Internet. Ceramic art has a profound cultural background in China, and most of the ceramic art courses in schools are about related knowledge and skills. Teachers unilaterally instill knowledge to students, which is difficult to mobilize students' enthusiasm and learning autonomy (Zdorovets et al., 2020). The combination of the Internet and education can make use of the advantages of the Internet, establish a

personalized interactive platform, enhance teaching interaction, and analyze student problems from the students' learning situation and student psychology (Safdar et al., 2020). To help students solve learning problems in time, improve learning quality and efficiency, and help teachers explore new teaching strategies (Chang et al., 2021), an automatic question answering (QA) system is established based on deep learning (DL) by combining the Internet with an automatic QA system to improve teaching effects.

A large number of studies have been conducted on the exploration of teaching strategies in domestic and overseas. Arnesen et al. (2019) created an online module based on blended learning pedagogy, which includes a personalized learning experience; in the part of participating in a personalized course, pre-service teachers are more willing to participate in personalized learning at the end of the course than at the beginning. As teachers learn and experience personalized learning in the curriculum, teachers' attitudes generally become more positive and confident and they pay more attention to the role of individualization in student growth (Arnesen et al., 2019). Alamri et al. (2020) used the self-determination theory as a framework to investigate the relationship between students' perceptions of the satisfaction of their psychological needs (such as ability, autonomy, and relevance) and their intrinsic motivation when participating online courses that implement individualized learning principles. The research results indicated that the implementation of personalized learning principles in online courses can help support students' psychological needs (such as autonomy and ability) and intrinsic motivation. In addition, students believed that personalized learning interventions were attractive and effective in meeting their learning needs and interests (Alamri et al., 2020). Hussain et al. (2020) adopted a student-centered approach and implemented a "flipped classroom" model in engineering courses based on constructivism (that is, experience-based learning) and student personalized learning. The performance of the flipped classroom teaching method and the traditional teaching method was compared, using four lenses: students' performance, three investigations of students at different stages of the semester, teacher's observation, and peer's observation (Hussain et al., 2020). Al-Razgan and Alotaibi (2019) designed two stages: the first stage was a focus group consisting of five language teachers to conduct a needs analysis. The next step is to analyze the data collected in the previous stage and apply it to the system design. It consists of two parts: one is the practice part, where learners can practice specific spelling rules, and the other is the game part, where learners can play spelling games and score points (Al-Razgan and Alotaibi, 2019). Xu (2021) proposed that the educational nature of multimedia technology is its fundamental attribute. The application of multimedia technology in the field of education is often carried out in combination with the content of the syllabus, teaching objectives, and goals. It is to use the audition function of multimedia in specific classroom teaching activities to transmit relevant knowledge to students in a more vivid and flexible form, which can help students better understand and master relevant knowledge and skills. The artistic

characteristics of multimedia technology provide an important guarantee for the smooth development of ceramic art course teaching (Xu, 2021). Luo (2021) pointed out that ceramic art teaching is often combined with other disciplines, and it will certainly be able to shine. Combined with tea culture, ceramic art teaching can introduce the development of ceramic craftsmanship from the perspective of utensils, and the influence of glaze color and texture on ceramic art from the perspective of color. And it can also form a new esthetic meaning under the influence of tea culture aesthetics such as tea ceremony, tea science, tea painting, and tea art. Combined with art, it can introduce the composition and emotional expression of ceramic art from the perspective of the similarities and differences between ancient and modern Chinese and foreign art. Combined with history, the characteristics of famous kilns of various generations can be introduced from the perspective of unearthed cultural relics. Combined with software programming, from the perspective of mobile games of ceramic art, the basic process and core keys of completing ceramic art can be introduced. By allowing students to dabble in different disciplines, it is beneficial for students to improve their knowledge system, broaden their horizons, and meanwhile, they can understand the history of ceramic art development and build a theoretical system of ceramic art (Luo, 2021).

Internet + education can make up for the shortcomings of the traditional teaching model, so that students can actively learn and consolidate knowledge and check for omissions according to their own learning situation in spare time. Internet + education can also analyze the psychological related problems of students according to their current status in terms of students' learning psychological problems. In terms of knowledge learning, when students encounter a problem, they often use commercial engines to search for answers, but it is often difficult to solve the problem quickly and accurately, which leads to the accumulation of problems and affects the learning effect (Uroková, 2020). The establishment of an automatic QA system based on the DL network, as a supplement to the new teaching model, can help students get help faster and more accurately, and enhance their interest in learning.

Personalized Q&A system for students in the context of Internet +

DL-related theories

Deep learning is a new research direction in the field of Machine Learning (ML). DL is the inherent law and representation level of sample data. The information obtained in the learning process is of great help to the interpretation of data such as text, images, and sounds (Wang et al., 2020). The ultimate goal of DL is to enable machines to have human-like learning and analysis capabilities. It is a complex ML algorithm that far surpasses previous related technologies in speech and image recognition (Pang et al., 2021). Most ML algorithms use supervised learning. A supervised

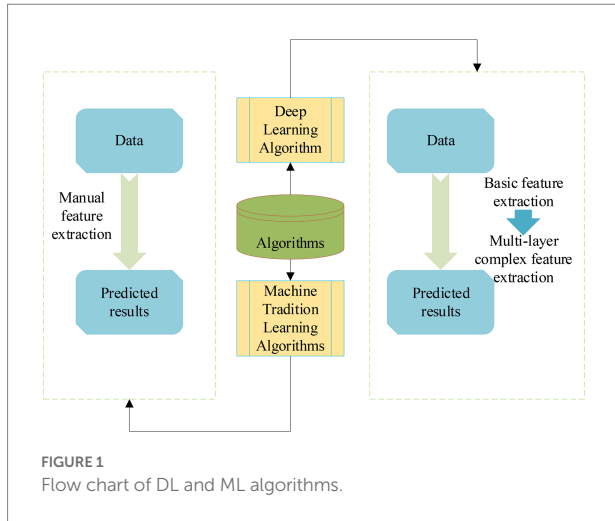


FIGURE 1
Flow chart of DL and ML algorithms.

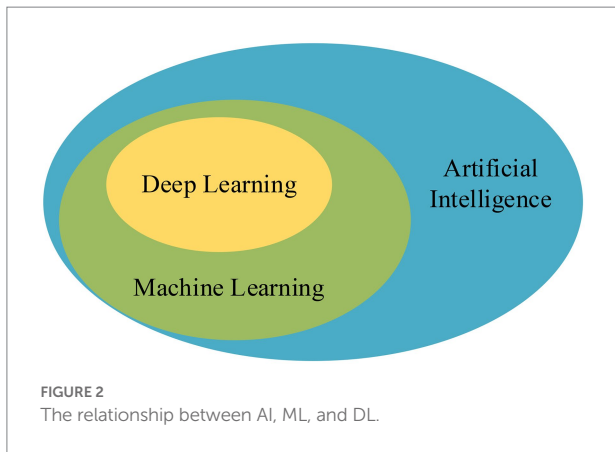


FIGURE 2
The relationship between AI, ML, and DL.

learning algorithm is an algorithm, in which any type of the variable must be input and produce an output variable. Its purpose is to approximate the mapping function to a level where it is possible to have new input data that consistently predict appropriate values for that particular data. In an unsupervised ML algorithm, only the input data are required and no output variables need to be provided. Its main purpose is to model the underlying structure or distribution of data to obtain more data. Semi-supervised algorithms are where most real-world problems lie, and these types of algorithms fall somewhere between supervised and unsupervised algorithms.

The flow of the DL algorithm and ML algorithm is displayed in Figure 1.

Deep learning algorithms have more advantages than traditional ML algorithms in implementing models and extracting feature data. Many problems in the development of artificial intelligence (AI) have been effectively solved with the development of DL. AI is to endow machines with human intelligence, so that computers have the ability to judge some situations like people; ML is to use algorithms to analyze data to achieve specific tasks, which is a method to achieve AI (Zhu et al., 2020). DL optimizes the early algorithms of ML and is a better technique, and their relationship is exhibited in Figure 2.

Individualized learning theory

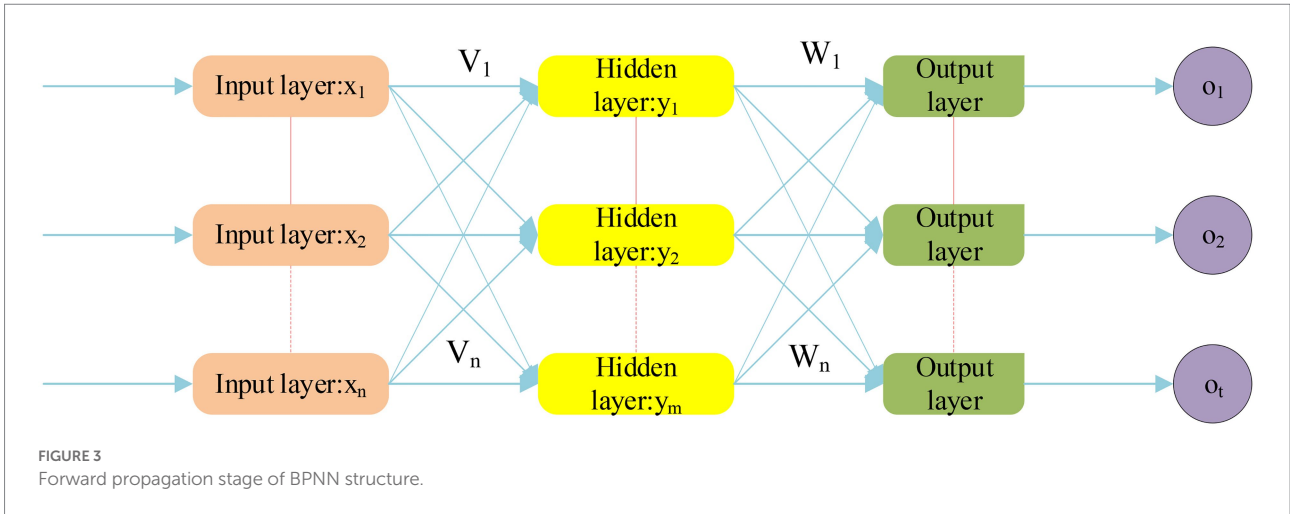
Personalized learning is to discover and solve the learning problems of students through the comprehensive evaluation of specific students, and tailor learning strategies and learning methods different from others for students. Every student is unique, with his unique talents, strengths, and weaknesses. Individualized methods should be used to solve children’s learning problems (Tsai et al., 2020). The idea of personalized learning has also been mentioned by more and more experts. With the rapid development of information technology, a variety of new technologies have also motivated solving the problem of personalized learning (Kurilovas, 2019). In addition to curriculum teaching design, student learning has other factors that affect academic performance, such as students’ gender, parental support, learning concepts, and learning orientation (Hwang and Fu, 2020). Some scholars have studied other factors that affect students’ academic achievement, including learning orientation, learning concepts, learning motivation, academic commitment, and parental support. According to different backgrounds and environments, it needs to study and analyze internal and external factors separately about students’ learning (Erwiza et al., 2019).

In summary, scholars’ research on personalized learning has gradually shifted from focusing on learning content and learning navigation to analyzing educational psychological factors such as learning styles. Data analysis of personalized learning requires the design of accurate personalized learning services for students. (McCarthy et al., 2020). Personalized learning is already one of the basic characteristics of current learning and a new direction of current education and learning.

Back propagation neural network

Back Propagation Neural Network (BPNN) mainly includes two stages, namely forward propagation and backpropagation. The forward propagation stage refers to a process in which the excitation signal is first introduced from the input layer, then processed by the hidden layer, and finally to the output layer (Ruan et al., 2020). If the output of the output layer is inconsistent with the actual output, it will enter another stage of backpropagation. The backpropagation stage is one of the cores of the BPNN model. The operating principle is to propagate the wrong information to the input layer through the hidden layer through its own characteristics. But in actual work to adjust it, the purpose is to output the correct value. Therefore, it mainly shows the forward propagation stage in the BPNN model, as expressed in Figure 3.

The input is a n-dimensional feature vector: $X = (x_1, x_2, \dots, x_n)^T$, and the output is a m-dimensional feature vector: $(y_1, y_2, \dots, y_m)^T$. V means the weight between the input layer and the hidden layer, and W means the weight between the hidden layer and the output layer. The training process of BPNN is mainly divided into two parts, signal forward propagation and error information back feedback. The Sigmoid function is used in the hidden layer, and the equation is as follows.



$$f(a) = \frac{1}{1 + e^{-x}} \tag{1}$$

The BPNN algorithm is divided into the following steps. First, initialize the network, take random numbers in $(-1, 1)$ to initialize (weight) and b (threshold). M means a maximum number of training times, μ means calculation accuracy, and e means the error function. The k -th input sample and expected output:

$$\begin{aligned} x(k) &= (x_1(k), x_2(k), \dots, x_n(k)) \\ d_o(k) &= (d_1(k), d_2(k), \dots, d_n(k)) \end{aligned} \tag{2}$$

The input and output of each neuron in the hidden layer and output layer:

$$\begin{aligned} hi_h(k) &= \sum_{i=1}^n w_{ih} x_i(k) - b_h & h &= (1, 2, \dots, p) \\ ho_h(k) &= f(hi_h(k)) & h &= (1, 2, \dots, p) \\ yi_o(k) &= \sum_{h=1}^p w_{ho} ho_h(k) - b_o & o &= (1, 2, \dots, p) \\ yo_o(k) &= f(yi_o(k)) & o &= (1, 2, \dots, p) \end{aligned} \tag{3}$$

Define the error function as:

$$\begin{aligned} e &= \frac{1}{2} \sum_{o=1}^q (d_o(k) - yo_o(k))^2 \\ \frac{\partial e}{\partial w_{ho}} &= \frac{\partial e}{\partial yi_o} \frac{\partial yi_o}{\partial w_{ho}} \\ \frac{\partial yi_o(k)}{\partial w_{ho}} &= \frac{\partial \left(\sum_h^p w_{ho} ho_h(k) - b_o \right)}{\partial w_{ho}} = ho_h(k) \\ \frac{\partial e}{\partial yi_o} &= \frac{\partial \left(\frac{1}{2} \sum_{o=1}^q (d_o(k) - yo_o(k))^2 \right)}{\partial yi_o} \\ &= \partial (d_o(k) - yo_o(k)) yo_o(k) \\ &= -(d_o(k) - yo_o(k)) f'(yi_o(k)) = \delta_o(k) \end{aligned} \tag{4}$$

The error function can be used to measure the training result. The smaller the error function value is better, so the weight value is modified to reduce the error function value. The equation is as follows:

$$\begin{aligned} \frac{\partial e}{\partial hi_h(k)} &= \frac{\partial \left(\frac{1}{2} \sum_{o=1}^q (d_o(k) - yo_o(k))^2 \right)}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial hi_h(k)} \\ &= \frac{\partial \left(\frac{1}{2} \sum_{o=q}^q (d_o(k) - f(yi_o(k)))^2 \right)}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial hi_h(k)} \\ &= \frac{\partial \left(\frac{1}{2} \sum_{n=1}^q (d_o(k) - f(\sum_{h=1}^p w_{ho} ho_h(k) - b_o))^2 \right)}{\partial ho_h(k)} \frac{\partial ho_h(k)}{\partial hi_h(k)} \\ &= -\sum_{o=1}^q (d_o(k) - yo_o(k)) f'(yi_o(k)) w_{ho} \frac{\partial ho_h(k)}{\partial hi_h(k)} \\ &= -\left(\sum_{o=1}^q (d_o(k) w_{ho}) f'(hi_h(k)) \right) \\ &= \delta_h(k) \end{aligned} \tag{5}$$

Then, update the connection weight w_{ho} :

$$\begin{aligned} \Delta w_{ho}(k) &= -\eta \frac{\partial e}{\partial w_{ho}} = \eta \delta_o(k) ho_h(k) \\ w_{ho}^{N+1} &= w_{ho}^N + \eta \delta_o(k) ho_h(k) \end{aligned} \tag{6}$$

η means learning rate and w_{ho} means connection weight. The threshold change is obtained as Equation (7). Finally, the global error is calculated as in Equation (8).

$$\begin{aligned} \Delta b_o(k) &= \eta * \delta_o(k) \\ \Delta b_h(k) &= \eta * \delta_h(k) \end{aligned} \tag{7}$$

$$E = \frac{1}{2m} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_o(k))^2 \tag{8}$$

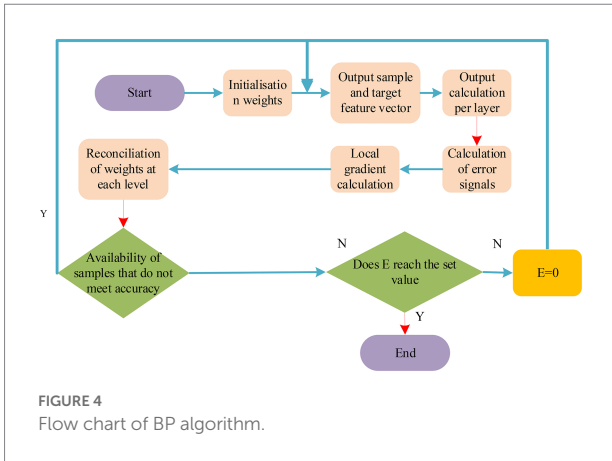


FIGURE 4 Flow chart of BP algorithm.

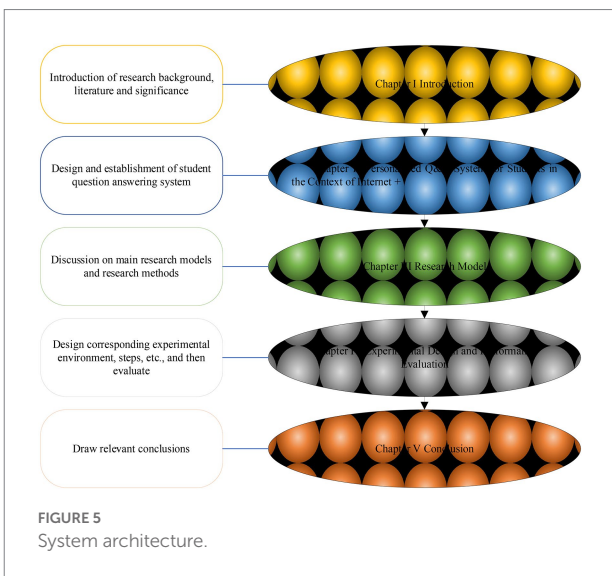


FIGURE 5 System architecture.

If the final prediction error of the neural network (NN) reaches the preset value or reaches the maximum number of training times, it will stop training. If it does not meet the requirements, the next sample is input for the next round of learning and training with the expected output value. Figure 4 shows the BP algorithm flow.

Research model

Convolutional neural network

Under the background of the Internet, the exploration of students' adaptive DL path of contemporary ceramic art and the analysis of teaching strategies are carried out. The specific system architecture is expressed in Figure 5.

The main advantage of CNN is that increasing the number of network layers can model complex non-linear relationships (Kattenborn et al., 2021). CNN includes the input layer, one or more convolutional layers, the output layer, the pooling layer,

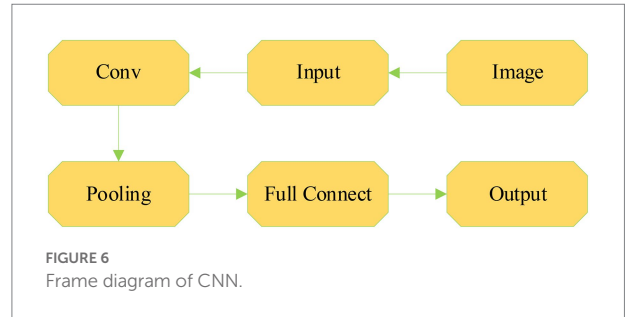


FIGURE 6 Frame diagram of CNN.

and the dropout. Figure 6 is a simplified structure diagram of CNN.

The basic two-dimensional convolution is as in Equation (9), where x means the sample matrix, size of filter w is represented by $U \times V$, and output y means convolution of signal sequence x and filter w .

$$y_{i,j} = \sum_{u=1}^U \sum_{v=1}^V w_{u,v} x_{i-u+1, j-v+1} \quad (9)$$

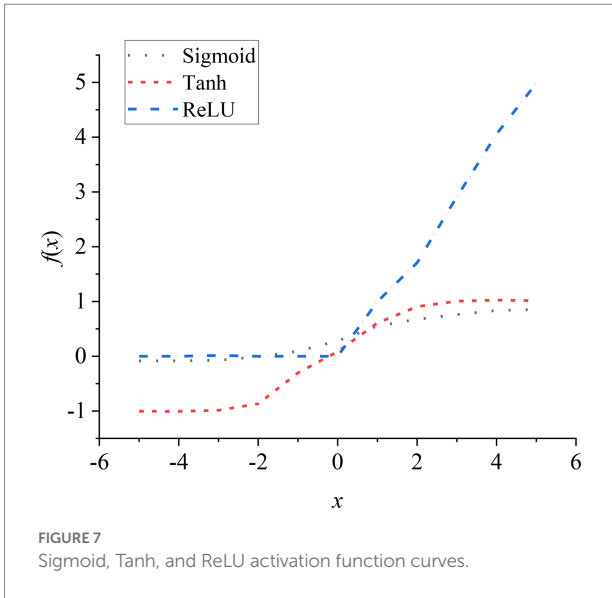
Assuming that the output sample y starting from (3,3), $y_{3,5} = \sum_{u=1}^3 \sum_{v=1}^3 w_{u,v} x_{3-u+1, 3-v+1} = -1$ can be gained, then the output matrix can be obtained. The continuous and discrete forms are as follows.

$$\begin{aligned} y(n) &= (f \otimes w)(n) = \int_{-\infty}^{\infty} f(x)w(n-x)dx \\ y(n) &= (f \otimes w)(n) = \sum_{x=-\infty}^{\infty} f(x)w(n-x) \end{aligned} \quad (10)$$

\otimes means convolution operation, f means input function, w represents weighting function, y means feature map, x represents a map, and w means filter. Input the result of feature mapping into the next layer as follows.

$$\begin{aligned} z^1 &= w^1 \otimes x + b^1 \\ y^1 &= f(z^1) \end{aligned} \quad (11)$$

Characteristics of convolutional layer: Local connections are shared with convolution kernels. The purpose of the local connection is to reduce the number of parameters, increase the calculation speed, and effectively reduce the probability of overfitting. Convolution kernel sharing means that when extracting feature maps, the same convolution kernel is shared between locations, which is also to reduce the number of parameters and further increase the calculation speed. The pooling layer mainly compresses data, reduces parameters, and improves calculation speed. Fully connected layer: It is the hidden layer of the traditional NN. Each neuron in this layer needs to be connected to the previous neuron. In CNN, the convolutional layer, pooling layer, and fully connected layer need to add



activation functions. The activation function is to activate neurons to reduce the probability of overfitting. Commonly used activation functions mainly include Sigmoid-type functions and ReLU functions. Among them, the commonly used forms of Sigmoid-type functions are Logistic function and Tanh function, as shown in the following equation:

$$a = f(z^1) \tag{12}$$

$$\begin{aligned} \sigma(x) &= \frac{1}{1 + e^{-x}} \\ \text{Tanh}(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned} \tag{13}$$

$$\text{ReLU}(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{14}$$

Figure 7 is a graph of Sigmoid, Tanh, and ReLU functions.

Tanh function: The output value of the function is centered at 0. Although it has a faster convergence speed, there is a gradient dispersion. ReLU function is currently a commonly used activation function, which solves the gradient dispersion. And it has a fast convergence speed and it can reduce the possibility of overfitting. The LeakyReLU activation function is an improved version of the ReLU function. The calculation equation is as follows, a generally takes 0.01 as a constant value.

$$f(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \tag{15}$$

The weight update in the NN is generally the BP algorithm, and CNN also uses the BP algorithm to update the weight and propagate forward. The output of the i -layer is as the following equation:

$$x^i = f(u^i) \tag{16}$$

$$u^i = W^i x^{i-1} + b^i \tag{17}$$

W means i -layer weight and b represents i -layer bias. Suppose the overall loss function of the NN:

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \tag{18}$$

N means the total number of samples and c means the number of sample categories. t_k^n represents the n th sample label of the k -th dimension, and y_k^n represents the output of the n -th sample of the k -th dimension. Each layer of CNN uses the gradient descent method to update the weight, and the calculation formula for the weight update and bias update:

$$W_n^i = W_o^i - \eta \frac{\partial E}{\partial W_o^i} \tag{19}$$

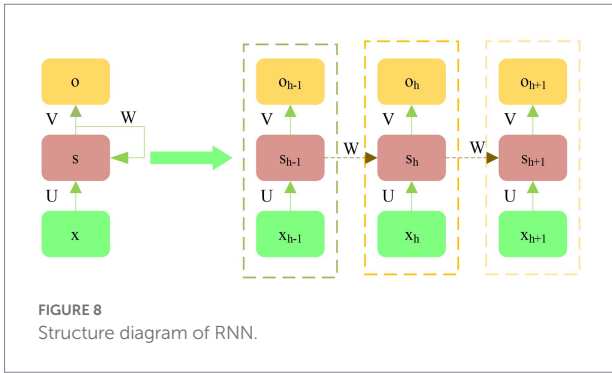
$$b_n^i = b_o^i - \eta \frac{\partial E}{\partial b_o^i} \tag{20}$$

W_o^i and b_o^i represent Weight and bias before update; W_n^i and b_n^i represent Weight and bias after the update. η means the learning rate in the gradient descent method. The convolutional layer in CNN propagates forward, and the output feature map of each i -layer convolution:

$$x_j^i = f \left(\sum_{l \in M_j} x_l^{i-1} * k_{lj}^i + b_j \right) \tag{21}$$

M_j means input feature map; k_{lj}^i represents the convolution kernel, and f means activation function. With CNN backpropagation, pooling layer error backpropagation as Equation (22), convolutional layer error direction propagation Equation (23), convolutional weight update, bias update as (Equations 24 and 25).

$$\delta^{i-1} = \text{upsample}(\delta^i) \odot \sigma'(u^{i-1}) \tag{22}$$



$$\delta^{i-1} = \delta^i \left(\partial \delta^i / \partial \delta^{i-1} \right) = \delta^i * \text{rot180}(W^i) \odot \sigma'(u^{i-1}) \quad (23)$$

$$\partial E / \partial W^i = \left(\partial E / \partial u^i \right) \left(\partial u^i / \partial W^i \right) = x^{i-1} * \delta^i \quad (24)$$

$$\partial E / \partial b^i = \sum_{u, v} \left(\delta^i \right) u, v \quad (25)$$

upsample(δ^i) means sampling, \odot represents Hadamard product, δ represents error value, i means i layers, and E means loss function (Bocci and Carlini, 2022).

Long short-term memory

Compared with traditional NNs, Recurrent Neural Network (RNN) mainly adds time to the loop structure, so that when processing data, the processed information can retain some “memory,” and it can be free from the limit of sequence length when processing data. Figure 8 shows the structure of RNN.

x : the input at the current time h ; s : hide the node status in real-time; o : output (RNN processing). The specific equation is:

$$s_t = f(V_{x_h} + W_{s_{h-1}} + b) \quad (26)$$

$$o_t = \text{softmax}(V_{s_h} + c) \quad (27)$$

f represents activation function sigmoid; V and W are weight matrixes between layers; and b and c : bias value. The shortcomings of RNN are also obvious, which includes unstable model parameter update, the gradient exploding or disappearing, and the memory being short. The LSTM model is an improvement of the cyclic NN model. In the structure, the “gate” structure is added, so that the problems caused by the long distance can be solved, even if the length of the data sequence is different. LSTM model neuron

is composed of the unit state, the output gate, the input gate, and the forget gate. The working model of the forget gate: The sigmoid function assigns the weighted calculated value of the input pt. at the current time t and the output $kt-1$ at the time $t-1$, and uses the above to control the influence of the previous output sequence information on the input stream. The equation is as (28):

$$g_t = \sigma(W_f \cdot [k_{t-1}, p_t]) + b_g \quad (28)$$

The sigmoid function is used to weight the input pt. and the output $kt-1$ at $t-1$ to obtain the value s , as shown in the following Equation (29). The new state candidate value of the unit is generated by the non-linear Tanh function, as shown in the following Equation (30). The new state A_t of the unit only needs to add the two, and then passes through the forget gate and the input gate, as shown in Equation (31):

$$s_t = \sigma(W_i \cdot [k_{t-1}, p_t]) + b_s \quad (29)$$

$$\widetilde{A}_t = \text{tann}(W_c \cdot [k_{t-1}, p_t]) + b_A \quad (30)$$

$$A_t = g_t \cdot A_{t-1} + s_t \cdot \widetilde{A}_t \quad (31)$$

The value qt output by the gate output needs to use the sigmoid function to weight the input pt. and the output k_{t-1} at $t-1$, as shown in the following Equation (32). Next, the output of the LSTM unit is controlled by the non-linear Tanh function calculation, and finally, the output value kt is shown in the following Equation (33). The advantage of LSTM is to solve the data problem due to the long distance.

$$q_t = \sigma(W_q \cdot [k_{t-1}, p_t]) + b_q \quad (32)$$

$$k_t = q_t \cdot \text{tann}(A_t) \quad (33)$$

Through the previous introduction of different network models, the LSTM model is combined with the CNN model to obtain a hybrid network model based on the CNN model. The structure of the hybrid neural network (HNN) model is shown in Figure 9 below.

Gated recurrent unit

Gated Recurrent Unit (GRU) is a variant of LSTM. The input gate and forget gate in LSTM are merged into update gates and

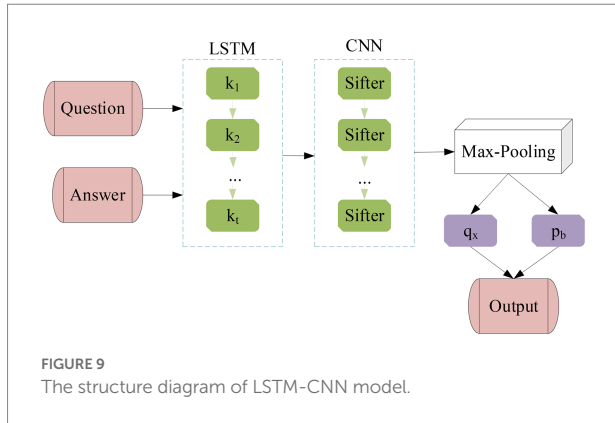


FIGURE 9 The structure diagram of LSTM-CNN model.

some other changes to form GRU (Shahid et al., 2020). GRU has few parameters and fast convergence. Input the sequence $x = \{x_1, x_2, x_3, \dots, x_t\}$ to the GRU, the internal state is as follows:

$$z_t = \sigma(W_z [h_{t-1}, x_t]) \tag{34}$$

$$r_t = \sigma(W_r [h_{t-1}, x_t]) \tag{35}$$

$$h_t = \tanh(W_h [r_t * h_{t-1}, x_t]) \tag{36}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * h_t \tag{37}$$

$$y_t = \sigma(W_o \cdot h_t) \tag{38}$$

x_t represents input vector; h_{t-1} represents hidden layer output at the previous moment; z_t means update gate; r_t means reset gate; $*$ represents multiply by point; σ represents sigmoid activation function; W represents weight matrix; and dropout is added during training to prevent overfitting.

The Attention mechanism simulates the attention model of the human brain, which is essentially a resource allocation model. Basic working principle: reasonable allocation of attention resources, more allocations to key parts, and less allocation to the rest, which can reduce refers to eliminating the adverse effects of non-key parts. Commonly used methods are used to score functions for soft attention and experiment with three weight value calculations. The first is to input all attention models to score and sum, as in Equation (39):

$$\alpha_t = \frac{\exp(\text{score}(h_t))}{\sum_{i=1}^T \exp(\text{score}(h_i))} \tag{39}$$

α_t : The weight value of the t-th input; h_t : the t-th input; score(): the score of the input. The second type is to calculate the

input first, and then the calculated input model, such as Equation (40):

$$y_t = \tan h(W h_t) \tag{40}$$

$$\alpha_t = \frac{\exp(\text{score}(y_t))}{\sum_{i=1}^T \exp(\text{score}(y_i))} \tag{41}$$

y_t is gained by calculating h_t ; W : the output obtained by inputting the single-layer NN; and α_t is the input weight value. The first two methods are after obtaining α_t , multiply α_t and h_t , and finally accumulate, as in Equation (42):

$$S_t = \sum_{i=0}^n \alpha_t h_t \tag{42}$$

S_t : the output of the input model. The third type is obtained based on the second type, and its own input for calculation, such as Equation (43):

$$S_t = \sum_{i=0}^n (\alpha_t h_t + h_t) \tag{43}$$

According to the GRU and Attention mechanism, a NN model based on BiGRU-Attention (Bidirectional Gated Recurrent Unit Network-Attention) is constructed, as shown in Figure 10.

Experimental design and performance evaluation

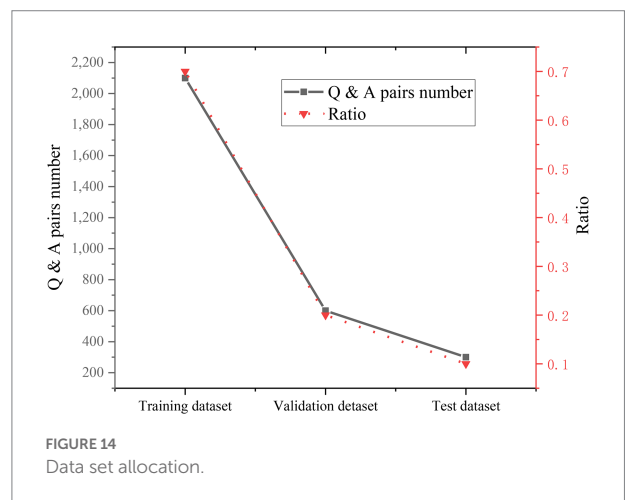
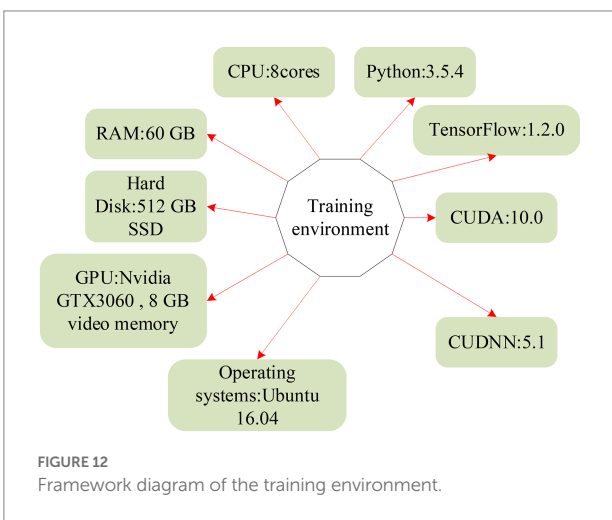
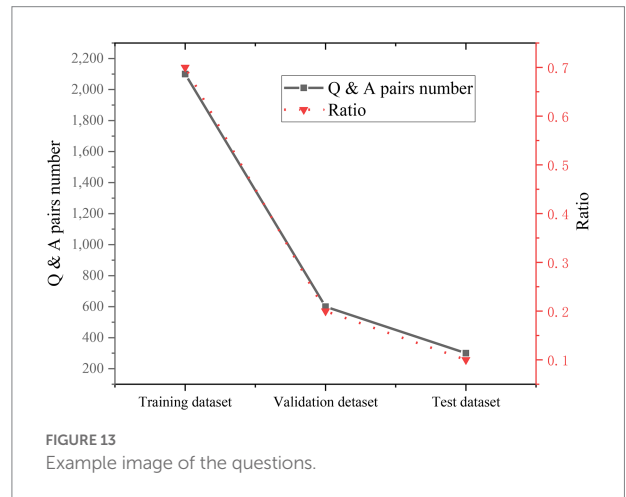
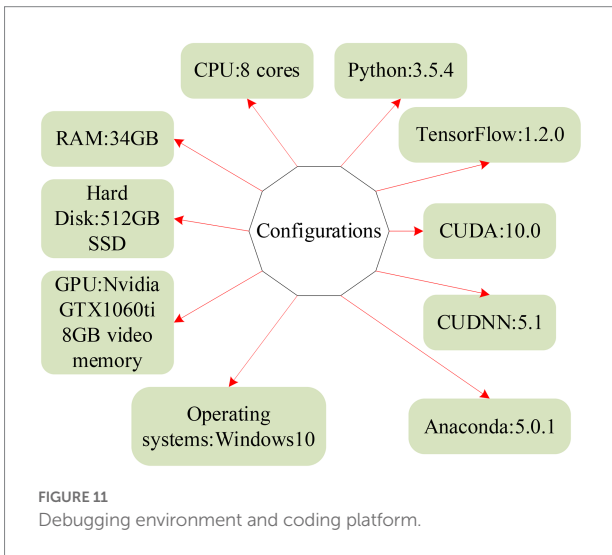
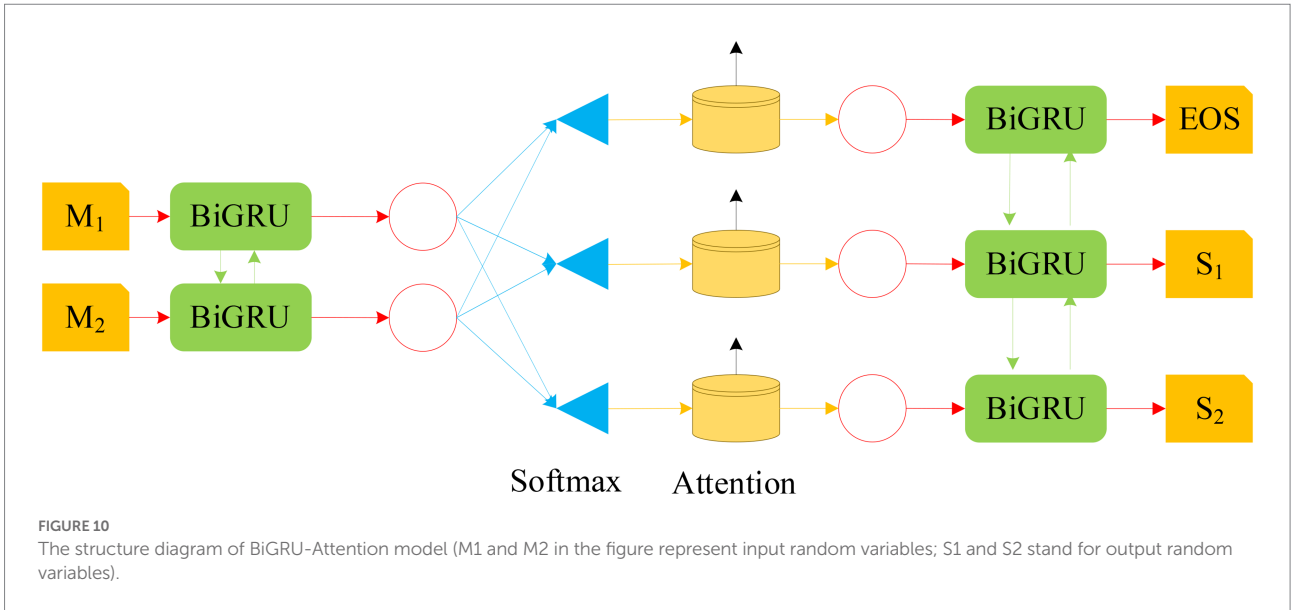
Experimental environment and data processing

There are two experimental environments, as shown in Figure 11 for the platform for writing and debugging code.

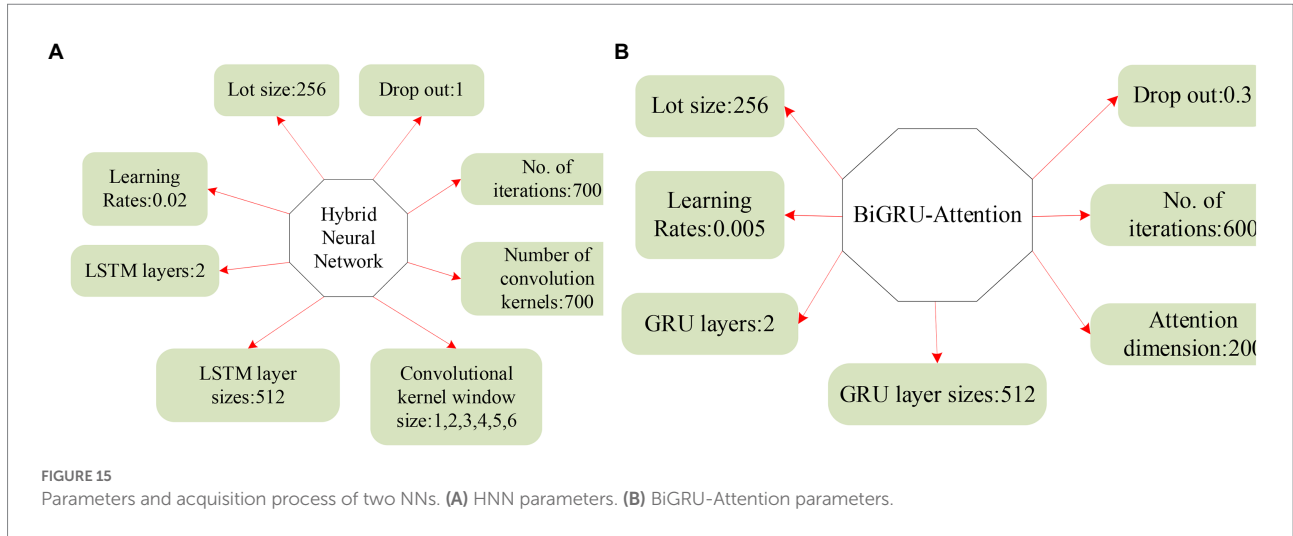
The platform used for training data is shown in Figure 12, and the operating system is Ubuntu.

At present, the Q&A database in the field of ceramic art on the Internet is very fragmented. Several ceramic art question bank websites are selected to crawl and clean the captured data. Finally, about 300 multiple-choice questions are selected as training data. First, visit the target page to obtain the page data, and then process and extract the required data, mainly extracting multiple choice questions. And it should be noted that the answer to the question is not given directly, so it is necessary to jump to another page to obtain the question with the answer option, as shown in Figure 13, and finally, the grabbed data are saved to a txt file.

After the captured multiple-choice questions are processed to a certain extent, the data set is divided into 70% training data set,



20% verification data set, and 10% test data set. The specific data are shown in Figure 14.



Model establishment

The parameters of the CNN-based hybrid model and the BiGRU-Attention model are demonstrated in Figure 15.

Objective function and evaluation index

The objective function of the model is as follows. a^+ represents positive example answer vector; a^- represents negative example answer vector; and m represents threshold parameter.

$$L = \max(0, m - \cos(q, a^+) + \cos(q, a^-)) \quad (44)$$

The evaluation metrics used are Accuracy (Acc) and Mean Reciprocal Rank (MRR). Acc is the ratio of the correct number of samples to the total number of samples. MRR is the effect index of the general search algorithm, as displayed in Equations (45) and (46).

$$\text{Accuracy} = \frac{\text{Number of samples with correct answers}}{\text{Total number of samples tested}} \quad (45)$$

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i} \quad (46)$$

Q : the set of sample queries; $|Q|$: the number of queries; rank_i : ranking of the first correct answer of the i -th query.

Due to some particularities of the adaptive DL algorithm of ceramic art, the traditional classification accuracy cannot be used for evaluation, so two indicators, F1 score and Area Under Curve (AUC), are introduced. The confusion matrix is shown in Table 1.

As shown in Table 1, when the true-positive (TP) BP model predicts positive, the true label is also positive. When

TABLE 1 Confusion matrix.

Forecast category	Real category	
	BP	EP
	BN	EN

the true-negative (TN) EN model predicts negative, the true label is also negative. The false-positive (FP) EP model predicts positive, and the true label is negative. The false-negative (FN) BN model predicts negative, and true labels are positive.

The F1 score is an important indicator to measure the effect of the binary classification model in statistics, which can be regarded as the harmonic average of the precision rate and recall rate of the model. Its maximum value is 1 and its minimum value is 0. According to the confusion matrix in Table 1, the corresponding expressions can be obtained, as expressed in Equations (47)–(49).

$$P = \frac{BP}{BP + EP} \quad (47)$$

$$H = \frac{BP}{BP + BN} \quad (48)$$

$$F_1 = \frac{2 \cdot P \cdot H}{P + H} \quad (49)$$

P stands for precision and H refers to recall.

In addition, AUC is a vital indicator to measure the quality of the two-class model, and it is the area under the Receiver Operating Characteristics (ROC) curve. It indicates the probability that positive examples are ranked ahead of negative examples. Its

TABLE 2 Metrics for AUC.

Range	Significance
AUC = 1	Perfect classification
1 > AUC > 0.5	Good predictive power
AUC = 0.5	No predictive power
0.5 > AUC ≥ 0	It is better to guess at random

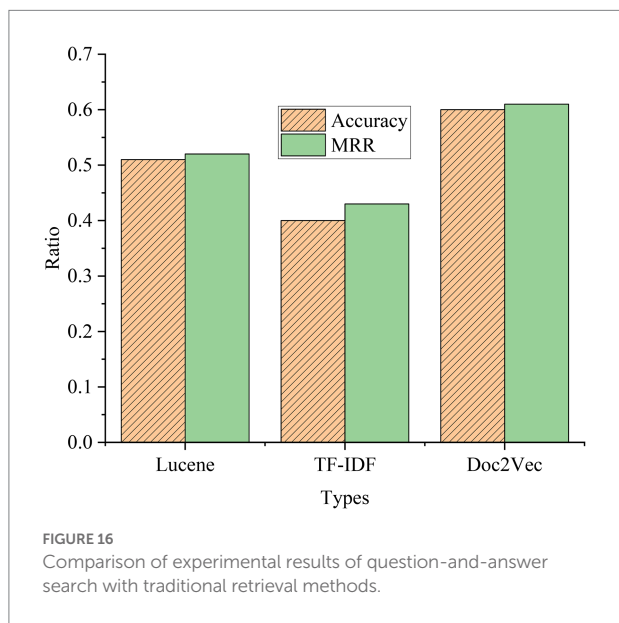


FIGURE 16 Comparison of experimental results of question-and-answer search with traditional retrieval methods.

maximum value is 1 and its minimum value is 0. Its measurement standard is shown in Table 2.

The closer the AUC value is to 1, the better and more perfect the classification effect is, and the closer the AUC value is to 0, the worse the classification effect. The specific calculation method is indicated in Equation (50).

$$AUC = \frac{\sum_{i \in \text{positiveClass}} \text{rank}_i - \frac{C(1+C)}{2}}{C \cdot D} \quad (50)$$

C stands for the number of positive samples, and D refers to the number of negative samples.

Comparative analysis of experimental results of traditional retrieval methods

The comparison results of the question-and-answer search experiment using the traditional retrieval system method are shown in Figure 16.

Lucene is an open-source full-text search engine toolkit that provides a complete query engine, indexing engine, and partial text analysis engine (English and German in two

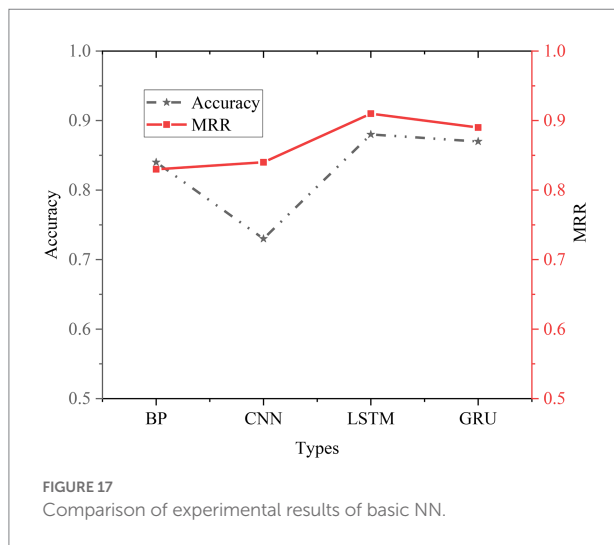


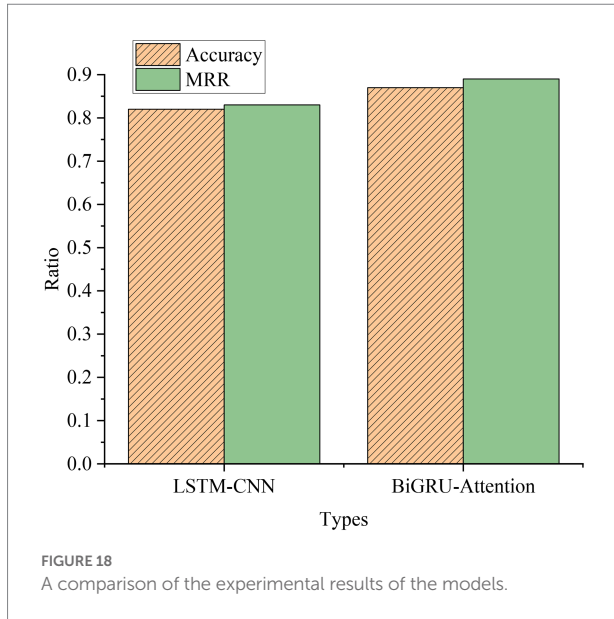
FIGURE 17 Comparison of experimental results of basic NN.

Western languages). TF-IDF (Term Frequency–Inverse Document Frequency) is a statistical method to assess the importance of a word to a document set or a document in a corpus. Doc2Vec is a vectorized representation of the created document. As shown in Figure 15, among the three retrieval methods, the Lucene engine is based on keyword search, with an accuracy rate of 0.51 and an MRR of 0.52; the accuracy rate of the TF-IDF method is only 0.4, and the MRR is only 0.43; the accuracy rate of the Doc2Vec method is 0.6, and the MRR is 0.61. The principle of the Lucene engine is similar to that of the TF-IDF method, and the search effect is general. Doc2Vec mainly converts words into word vectors, and the search effect is the best compared to the other two methods.

Comparative analysis of experimental effects of basic neural network

The experimental results of the basic NN are shown in Figure 17.

To better compare the performance of different models, this experiment also conducts an experimental comparative analysis of the basic NN model. It can be seen from the experimental results that the Acc of BPNN is 0.84, and the MRR is 0.83. Compared with the traditional method, the performance of BPNN is improved. Compared with the BPNN model, the performance of the CNN model has declined, and both LSTM and GRU are better than the CNN model in processing text tasks. Compared with the LSTM model, the GRU model has a simpler structure, fewer parameters, and faster training speed. The LSTM model has the best effect, but the structure is more complex than the GRU model, with more parameters and slower training speed. Therefore, when choosing a NN model, it should be selected according to the actual situation.

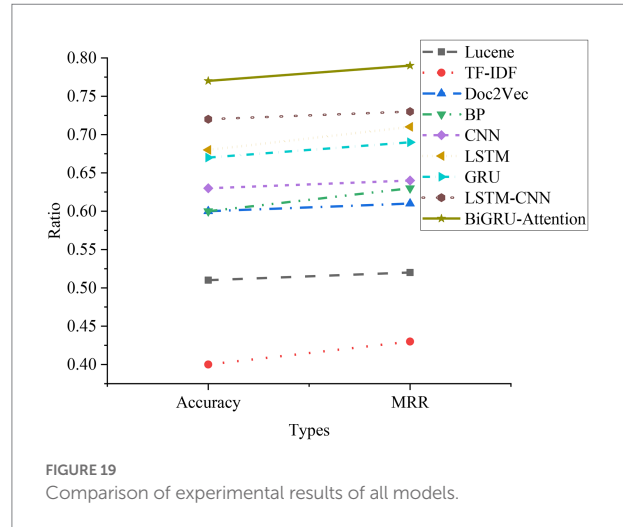


Experimental comparison and analysis of LSTM-CNN model and BiGRU-Attention model

Figure 18 signifies the comparative analysis of the experimental results of the LSTM-CNN model and the BiGRU-Attention model.

As shown in Figure 18, compared with the traditional model and the NN model, the used LSTM-CNN and BiGRU-Attention models are significantly better. The Acc of the LSTM-CNN model is 0.82, and the MRR is 0.83. Compared with the traditional model and the basic NN model, the Acc and MRR are much improved, and better results can be achieved. The Acc of the BiGRU-Attention model is 0.87, and the MRR is 0.89. Compared with the LSTM+CNN model, the Acc and MRR are also improved. The attention mechanism can retain more effective features after the feature weights of each step are calculated, which greatly improves the model effect. The appropriate combination of different models and the advantages of different models can effectively improve the overall effect. The comprehensive effect of all models is compared in Figure 19.

It can be clearly seen from Figure 19 that the effect of the traditional method is obviously not as good as that of the basic NN model. The main reason is that the structure of Deep Neural Networks (DNN) is superior to traditional methods, and NN can better describe and extract semantic features when processing sentence information. RNN can also retain contextual information during processing, which can represent sentences more accurately. The LSTM-CNN model and the BiGRU-Attention model are better than the basic NN model. After the CNN model is combined with the LSTM model, the sentence features can be better extracted. Even in the face of complex sentences, the HNN model still works well. BiGRU can obtain more time series information. Combined



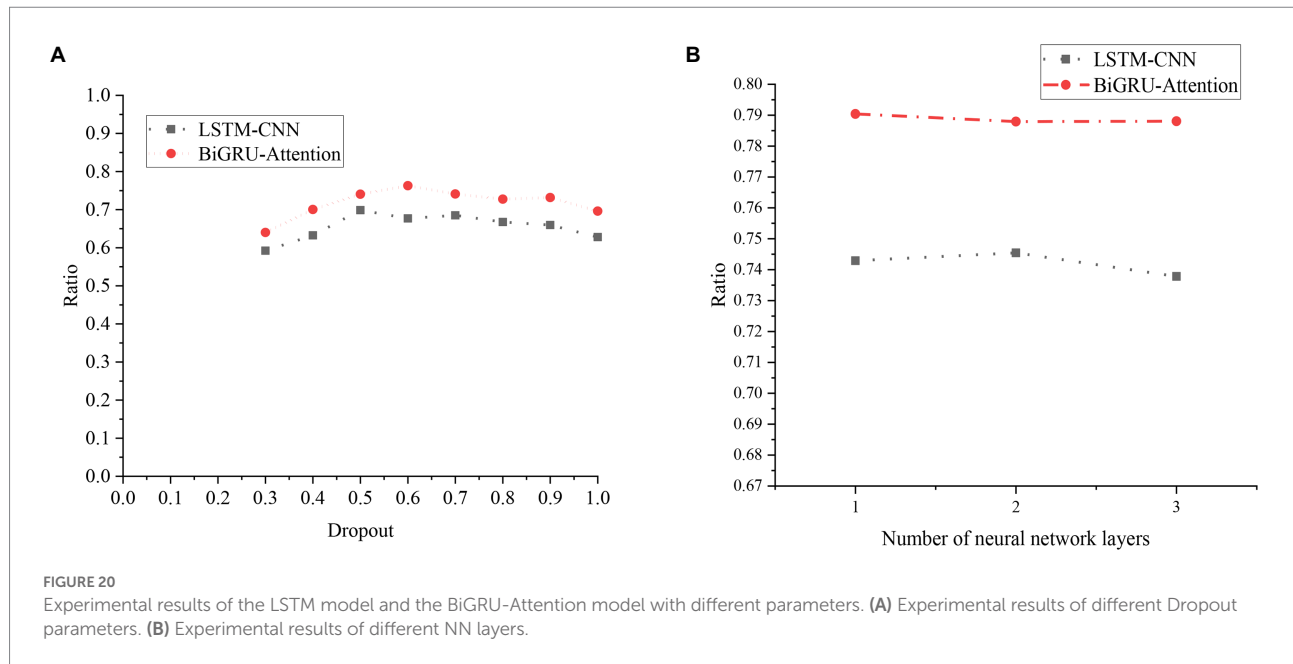
with contextual semantic features and attention mechanisms, it can retain more effective features during calculation, so that the model works well when dealing with long sequences.

For the two models, the comparison chart of different Dropout parameters and different numbers of NN layer models is shown in Figure 20.

Figure 20A indicates that setting different Dropouts has a certain impact on the model effect. When the Dropout is 0.5, the LSTM-CNN model has the best effect, and when the Dropout is greater than 0.5, the effect of the LSTM-CNN model begins to decline. The BiGRU-Attention model has the best effect when the Dropout is 0.6, and the effect decreases when the Dropout is greater than 0.6. Figure 20B reveals that the number of layers of the NN is selected as 1, 2, and 3 for the experiment. The effects of the two models have not changed much, and the increase in the number of layers will not significantly affect the experimental results. The increase in the number of NN layers will also slow down the training speed, so finally, the number of NN layers is 1.

Conclusion

Based on the automatic QA system established by DL, new teaching strategies are analyzed. The Internet is combined with the automatic QA system to help students solve problems encountered in the process of learning. Firstly, the advantages and disadvantages of BPNN, CNN, LSTM, and GRU are analyzed one by one. Secondly, models are constructed using different NNs. The LSTM is combined with the CNN to implement the LSTM-CNN model, which is compared and analyzed with the BiGRU-Attention model. Finally, experiments manifest that the accuracy and MRR of traditional retrieval methods can only reach about 0.5 on average. In the basic NN, BPNN, CNN, LSTM, and GRU, the average Acc of the BPNN model and the CNN model is about 0.81, and the average MRR is about 0.84. The average Acc of the LSTM model and the GRU model is about 0.82, and the average MRR is 0.83, which is



obviously better. After the CNN model is combined with the LSTM model, the sentence features can be better extracted. Even in the face of complex sentences, the HNN model still works well. BiGRU can obtain more time series information. Combined with contextual semantic features and attention mechanism, it can retain more effective features during calculation, so that the model works well when dealing with long sequences. The DL-based automatic QA system not only helps teachers explore new teaching strategies, but also assists students customize personalized learning plans to improve students' learning effects. However, the data of the Chinese QA system is still very small. For the DL model, the more data in the sample database, the more accurate the experimental results. Moreover, after solving the problem of individualization of students, the improvement of learning effect and the exercises between course content and teaching strategies still need further practice tests. In the future, research in this area will continue to improve the experiment to achieve better results.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by Xinyang Normal University Ethics Committee. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained

from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

Funding

This work was supported by the 2022 National Art Fund Youth Artistic Creation Talent Project: "Praise to Dabie Mountain" pottery creations series [2022-A-06-(129)-598] and the Henan Province Philosophy and Social Science Planning Project: "Research on Pottery Unearthed from the Chu Tomb of Warring-States Age Nobles at Yangcheng Site in Xinyang, Henan" (2021BYS041).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed

or endorsed by the publisher. The author confirms being the sole contributor of this work and has approved it for publication.

References

- Alamri, H., Lowell, V., Watson, W., and Watson, S. L. (2020). Using personalized learning as an instructional approach to motivate learners in online higher education: learner self-determination and intrinsic motivation. *J. Res. Technol. Educ.* 52, 322–352. doi: 10.1080/15391523.2020.1728449
- Al-Razgan, M., and Alotaibi, H. (2019). Personalized mobile learning system to enhance language learning outcomes. *Indian J. Sci. Technol.* 12, 1–9. doi: 10.17485/ijst/2019/v12i1/139871
- Arnesen, K. T., Graham, C. R., Short, C. R., et al. (2019). Experiences with personalized learning in a blended teaching course for preservice teachers. *J. Online Learn. Res.* 5, 275–310.
- Bocci, C., and Carlini, E. (2022). Hadamard products of hypersurfaces. *J. Pure Appl. Algebra* 226:107078. doi: 10.1016/j.jpaa.2022.107078
- Chang, S. J., Yang, E., Lee, K. E., and Ryu, H. (2021). Internet health information education for older adults: a pilot study. *Geriatr. Nurs.* 42, 533–539. doi: 10.1016/j.gerinurse.2020.10.002
- Erwiza, E., Kartiko, S., and Gimin, G. (2019). Factors affecting the concentration of learning and critical thinking on student learning achievement in economic subject. *J. Educ. Sci.* 3, 205–215. doi: 10.31258/jes.3.2.p.205-215
- Hussain, S., Munir, J. P. K., Munir, M. T., and Zuyeva, A. (2020). A quasi-qualitative analysis of flipped classroom implementation in an engineering course: from theory to practice. *Int. J. Educ. Technol. High. Educ.* 17, 1–19. doi: 10.1186/s41239-020-00222-1
- Hwang, G. J., and Fu, Q. K. (2020). Advancement and research trends of smart learning environments in the mobile era. *Int. J. Mob. Learn. Organ.* 14, 114–129. doi: 10.1504/IJMLO.2020.103911
- Kattenborn, T., Leitloff, J., Schiefer, F., and Hinz, S. (2021). Review on convolutional neural networks (CNN) in vegetation remote sensing. *ISPRS J. Photogramm. Remote Sens.* 173, 24–49. doi: 10.1016/j.isprsjrs.2020.12.010
- Kurilovas, E. (2019). Advanced machine learning approaches to personalise learning: learning analytics and decision making. *Behav. Inform. Technol.* 38, 410–421. doi: 10.1080/0144929X.2018.1539517
- Luo, S. (2021). The impact of internet-based patient self-education of surgical mesh on patient attitudes and healthcare decisions prior to hernia surgery. *Surg. Endosc.* 34, 5132–5141.
- McCarthy, K. S., Watanabe, M., Dai, J., and McNamara, D. S. (2020). Personalized learning in iSTART: past modifications and future design. *J. Res. Technol. Educ.* 52, 301–321. doi: 10.1080/15391523.2020.1716201
- Pang, G., Shen, C., Cao, L., and Hengel, A. V. D. (2021). Deep learning for anomaly detection: a review. *ACM Comput. Surv.* 54, 1–38. doi: 10.1145/3439950
- Ruan, X., Zhu, Y., Li, J., and Cheng, Y. (2020). Predicting the citation counts of individual papers via a BP neural network. *J. Informet.* 14:101039. doi: 10.1016/j.joi.2020.101039
- Safdar, G., Javed, M. N., and Amin, S. (2020). Use of internet for education learning among female university students of Punjab, Pakistan. *Univer. J. Educ. Res.* 8, 3371–3380. doi: 10.13189/ujer.2020.080809
- Shahid, F., Zameer, A., and Muneeb, M. (2020). Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos, Solitons Fractals* 140:110212. doi: 10.1016/j.chaos.2020.110212
- Tsai, Y. S., Perrotta, C., and Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assess. Eval. High. Educ.* 45, 554–567. doi: 10.1080/02602938.2019.1676396
- Uroikova, S. B. (2020). Advantages and disadvantages of online education. *ISJ Theor. Appl. Sci.* 89, 34–37. doi: 10.15863/TAS.2020.09.89.9
- Wang, G., Ye, J. C., and De Man, B. (2020). Deep learning for tomographic image reconstruction. *Nat. Mach. Intell.* 2, 737–748. doi: 10.1038/s42256-020-00273-z
- Xu, J. (2021). Usage of internet by university students of Hispanic countries: analysis aimed at digital literacy processes in higher education. *Eur. J. Contemp. Educ.* 10, 53–65.
- Zdorovets, M. V., Arbuz, A., and Kozlovskiy, A. L. (2020). Synthesis of LiBaZrOx ceramics with a core-shell structure. *Ceram. Int.* 46, 6217–6221. doi: 10.1016/j.ceramint.2019.11.090
- Zhu, W., Xie, L., Han, J., and Guo, X. (2020). The application of deep learning in cancer prognosis prediction. *Cancers* 12, 603. doi: 10.3390/cancers12030603