

Research Article

A Fuzzy Multicriteria Assessment Mechanism towards Musical Courses Using Deep Learning

Na Li ¹ and Ting Gong²

¹College of Literature Media and Educational Science, Chaohu University, Chaohu 238000, Anhui, China

²College of Art, Chaohu University, Chaohu 238000, Anhui, China

Correspondence should be addressed to Na Li; 055020@chu.edu.cn

Received 22 April 2022; Revised 12 May 2022; Accepted 19 May 2022; Published 13 June 2022

Academic Editor: Zaoli Yang

Copyright © 2022 Na Li and Ting Gong. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, the musical courses have been quite prevalent spiritual activities in online or offline scenarios. However, the teaching quality is diverse and cannot be easily assessed by general nonprofessional audience. Limited by the amount of experts, it is supposed to investigate intelligent mechanisms that can automatically assess the teaching quality of musical courses. To deal with such issue, the combination of artificial intelligence and conventional music knowledge acts as a promising way. In this work, a fuzzy multicriteria assessment mechanism is used towards musical courses with the use of a typical deep learning model: convolutional neural network (CNN). Specifically, note that features inside the musical symbol sequences are expected to be extracted by residual CNN structure. Next, multilevel features inside the musical notes are further fused with neural computing structure, so that feature abstraction of initial musical objects can be further improved. On this basis, notes can be identified with use of bidirectional recurrent unit structure in order to speed up fitting efficiency of the whole assessment framework. Comprehensive experimental analysis is conducted by comparing the proposed method with several baseline methods, showing a good performance effect of the proposal.

1. Introduction

Music is an important way to spread and communicate culture, and sheet music, as one of its carriers, is the most direct way to learn, share, and spread music by recording notes and other related information in detail. However, many scores are not publicly available or published, and are kept in paper form, which is easily damaged or even lost when the environment changes and times change; therefore, the complete preservation of paper scores is of far-reaching significance. Usually, scanning or photographing is a better way to preserve paper scores, but it is easily limited by the quality of scanning and storage space [1]. Make CNN differentiable to realize the nonlinear mapping from low-dimensional simple features to high-dimensional complex features in the score image; the pooling layer reduces the amount of weight parameters on the premise of retaining the main features of the convolution layer, speeds up the

calculation speed prevent overfitting problems. Although with the rapid improvement of hardware performance such as scanners and memory, people can store more scores with better clarity, computers cannot directly use such digital scores. Only by extracting the symbolic content in the score images, people can use the scores more flexibly and conveniently in-depth to realize music arrangement, synthesis, and other operations. The development of theory and technology in the field of computer and image processing is constantly providing new ideas for extracting symbols from sheet music images.

Each course has accumulated a large amount of review data, which often truly reflects the learners' intuitive learning experience, as well as the real feedback and suggestions to the course and the teacher, and even the learning effect. Researchers can start from here and use the artificial intelligence-natural language processing technology, which is currently developing rapidly in the field of economy and

finance, to obtain more realistic results from the learners' perspective [2]. Some scholars have also proposed the research value of this breakthrough in their articles and have tried to mine the review data to conduct a "bottom-up" online course quality evaluation study (based on the learner's perspective), which has achieved certain results, but in general, the research method. This is also the main research direction of this study. By using deep learning technology to analyze the sentiment of review text, we can conduct a study on the quality evaluation of online courses based on learners' perspectives in the context of Big Data review and realize the development of a complete evaluation recommendation system to provide the direction of course selection for subsequent learners [3]. In terms of practical significance, learners' needs and preferences are important reference factors for platform quality improvement, and this study also hopes to provide reference and guidance for major platforms to improve and enhance course quality in the future [4]. In addition, the study will eventually form a website system based on the evaluation results, hoping to help learners quickly choose the right course for themselves without spending a lot of energy and time in this area [5].

The application of multimedia computers to piano teaching can, to a certain extent, solve the problem of the high cost of piano education and the scarcity of piano teachers and can also meet the market demand for more enthusiasts to learn this part of the piano. In the early stages of piano learning, the focus is on the grasp of the basics, playing proficiency, and the ability to play difficult repertoire in general. If you do not have any requirements for the tone of the piano or the emotion of the piece, you can use a MIDI keyboard or an electric piano to practice in the early semester. In other words, the electronic piano teaching system can replace the traditional way of piano learning for the primary stage. The hardware to build an electronic piano teaching system is very simple, only a multimedia computer and a MIDI keyboard or an electric piano is needed. The single-tone AMT task can be converted into audio pitch extraction and note onset detection, so that a piece of audio can be converted into a fundamental frequency sequence, and the fundamental frequency sequence can be segmented with the help of the onset position of the note, and converted to include pitch sequence of notes. The key to the system is the piano teaching software, and what the software must do to achieve better results. To solve the aforementioned problems, this paper proposes an intelligent student emotion evaluation method, which first collects students' expressions, summarizes the collected expressions by using expression recognition technology, and the teacher adjusts the teaching method after the students' learning emotion changes. By applying expression recognition technology in the classroom, teachers can provide objective, timely and comprehensive feedback on students' learning status, so that teachers do not have to spend too much time observing students' learning status and can focus on teaching and learning, which is conducive to improving teaching efficiency.

2. Related Jobs

Van and Chesnokova proposed an OMR algorithm based on a general framework, decomposing OMR technology into four subtasks: image preprocessing, note recognition, music information reconstruction, and final expression construction, and some breakthroughs in related research, all of which laid the foundation for subsequent OMR research [6]. They also summarized the software and hardware required for OMR systems, explored the potential application scenarios of OMR, and advanced the research of OMR. The widespread use of Big Data in recent years has also led researchers to focus on data-driven recognition approaches to accomplish various subtasks [7]. Owing to the excellent performance of deep learning in image processing, researchers have started to use neural network models for optical music recognition and gradually adopt end-to-end optical music recognition methods to simplify research methods based on generic frameworks, bringing new ideas and perspectives to OMR research [8]. Theoretical research on teaching quality evaluation is also becoming richer and more diverse [9]. Many scholars in the field of education have proposed various teaching evaluation models one after another. To ensure the quality of teaching, higher education teaching evaluation has been widely developed in different forms worldwide and plays an increasingly important and active role in the teaching process [10]. Higher education teaching quality evaluation is a nonlinear classification problem, and its results are influenced by the interaction of many factors. Therefore, when formulating the teaching quality evaluation system, the most basic factors that can directly reflect the teaching quality should be selected as the evaluation content. However, due to the different understanding and importance of teaching quality in each university, there are some differences in the content and methods of evaluation [11].

Based on the analysis of mudflow hazard evaluation factors, Dakhiel et al. established an evaluation model using the function approximation ability of the BP neural network [12]. This evaluation model can accurately evaluate the mudslide hazard and simulate the nonlinear functional relationship between some main evaluation indexes of the mudslide and the hazard degree. The experimental data show that the BP neural network improved by GA has a substantial improvement in accuracy and efficiency, and this model provides a new idea for evaluating the danger level of mudflow. AlJaser designed a three-layer BP neural network for the numeral recognition system and conducted experiments with the actual situation [13]. Let the network correct the appropriate weights and then predict the evaluation results of teaching quality according to the sample data. The results show that the system has a high correct recognition rate for not only printed numerals, but also has good results in recognizing handwritten numerals. Songbatumis predicted the gas outflow in the mining area by establishing a neural network prediction model, and the analysis results reflect that the model has good performance, and this model provides some basis for predicting and preventing gas disasters [14].

The establishment of hybrid music teaching quality evaluation model based on the BP neural network. First, after collecting the corresponding raw data according to the evaluation index system developed in Chapter 3, the collected raw data were normalized. Then the structure of the BP neural network was determined according to the evaluation system and the training of the samples was started, and the evaluation results were obtained followed by the error analysis of the model results. To address the shortcomings of the BP neural network model, GA-BP neural network is proposed to be used for hybrid teaching evaluation. The process of the GA-BP neural network algorithm is introduced and then the training and testing of the sample data are started to obtain the corresponding evaluation results. Finally, the BPS model and GA-BP model are compared and analyzed with the evaluation results of GA and BSA to obtain a hybrid evaluation model with better results.

3. Improved Neural Network Algorithm Design

There are two types of artificial neural networks in terms of how the models are connected. In a feed-forward network, information is passed only one way along the input neuron to the output neuron, and there are no loops or loops in the whole network, which is a simple and easy-to-implement structure. Unlike feedforward networks, each neuron in a feedback network feeds its output signal to the next neuron as an input signal, allowing each neuron to receive the input signal and output signal at the same time, which is a relatively more complex process.

In a network of artificial neurons, the transfer of information is inseparable from the role of the activation function [15]. The information is processed by the activation function to produce output information and transmitted to the next layer, a process that is essential for the transfer of information. To fully express the advantages of neural networks as parallel distributed processing systems, nonlinear activation functions are usually used, and several common activation functions are described as follows:

$$f(x) = \begin{cases} 0, & x \leq 1, \\ 1, & x > 1. \end{cases} \quad (1)$$

BP neural network, as a classical model in a multilayer feedforward neural network, is trained by error backpropagation, and its basic idea is to use gradient descent and error backpropagation to adjust the weights and thresholds in the model to obtain the minimum training error. BP neural network consists of an input layer, hidden layer, and output layer, and Figure 1 shows the topology of a three-layer BP neural network.

The training process of a BP neural network has two parts: one is the forward propagation of information, where the network takes input samples from the input layer, and

after processing them in the implicit layer, the results will be passed to the output layer. The training target error is 0.0001. Input the preprocessed sample data and generate the corresponding network model according to the BP neural network model structure designed in the previous section. The training process is realized by the function train, according to the input of the sample, the target output, and the preset training function parameters to train the network. If the error between the output at this point and the desired output does not meet the requirements, the second part—the backward propagation of the error—will begin. The second process is to send the error through the implicit layer to the input layer, and the error signal of each layer cell is used as the basis for modifying the weight of each cell. These two processes can be repeated until the output error of the network reaches the previously set range or reaches the predetermined number of times the network learns, and the whole learning process ends.

BP neural network is a nonlinear optimization method based on gradient descent, so for some complex problems, the training process may take a long time due to the slow convergence rate. From the training process, it is approximated downward along the slope of the error surface, while the error surface is generally complex and irregular in real problems, with many local minima, which will lead the network to fall into local minima.

The selection of the parameters of BP neural networks (such as the number of layers in the hidden layer, the number of neurons in the hidden layer, and the learning rate) has no clear theoretical basis so far and is generally determined by empirical formulas or continuous training experiments, thus may lead to long learning time and low efficiency [16]. The training learning and memory functions of the network are unstable. When the samples change, the already trained network model must retrain the network, affecting the samples that have been learned previously.

$$S(i, j) = \sum_m \sum_n I(m, n)K(i + m, j + n). \quad (2)$$

The downsampling layer, also called the pooling layer, is usually connected to the pooling layer after successive convolutional layer operations to gradually reduce the model size and thus the number of model parameters, while effectively preventing the occurrence of overfitting. The common pooling operations are max pooling and average pooling. Like the convolution kernel, the pooling operator scans the input feature map along the aspect of the feature map and acts on each depth of the feature map to maintain the dimensional consistency of the feature map. The difference in the pooling window size F and the pooling operator step size S affects the features, and the pooling window is too large to lose the feature map information, so $F=3, S=2$, or $F=2, S=2$ is chosen in most cases. The local information of the previous layer of the feature map is connected with all the output neurons, the parameters are trained by forwarding and backward

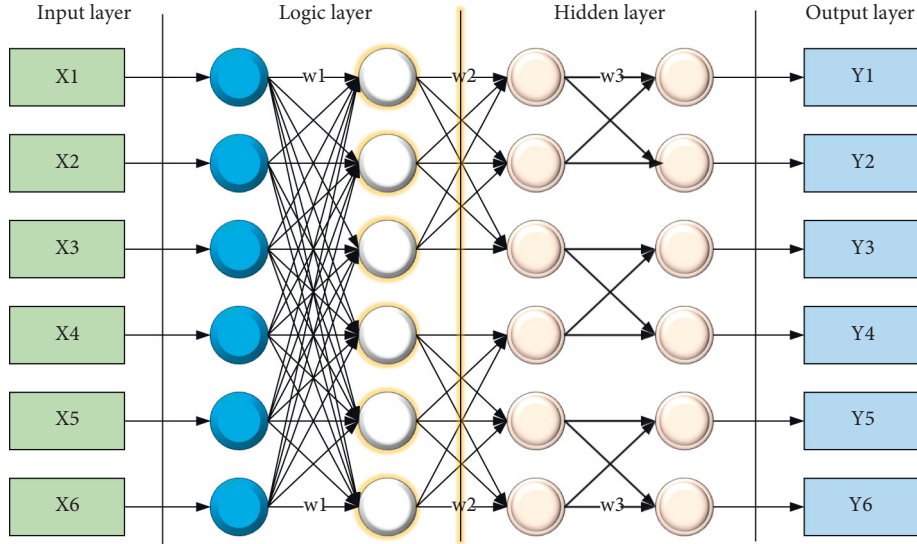


FIGURE 1: Topology of BP neural network.

propagation of weights and biases, and the activation function is used to achieve the classification.

$$f_t = \sigma(w_f [h_{t+1}, x_t] - b_f). \quad (3)$$

The “gate mechanism” effectively controls the degree of influence of the information in the previous moment on the subsequent moments, and the degree of influence decreases as the moment progresses so that the information in the earlier moments is restricted and the information in the later or current moments is retained to a higher degree, thus providing a reasonable distribution among contextual information.

$$h_t = o_t * \coth(c_t). \quad (4)$$

Neural networks mainly use backward propagation algorithms and gradient descent algorithms to achieve the adjustment of hyperparameters in the model during the training process of the model. The gradient descent method mainly targets the optimization of individual parameters, while the backward propagation algorithm uses the gradient descent algorithm to optimize all parameters in the model through the loss function thus reducing the value of the loss function, so the quality of the neural network model is directly related to the optimization of parameters. And the optimization process is divided into two stages: the first stage calculates the predicted value of the model by forwarding propagation and gets the distance between the predicted value and the correct value according to the defined loss function; the second stage calculates the gradient of the loss function for each parameter by backward propagation and updates it.

When a large amount of data is needed to optimize gradient descent algorithms, there are two main categories: momentum-based algorithms and adaptive-based algorithms. Common momentum-based algorithms include momentum and NAG, which are more likely to find optimal solutions in valley-type optimization surfaces, but if the trend is not obvious enough, it will certainly increase the

complexity of the core path of optimization parameters. In contrast, Adagrad, RMSProp, and Adadelta are adaptive algorithms, which are more likely to find a balance in various scenarios and are a compromise in optimization, simplifying the optimization of some complex scenarios but creating obstacles for the optimization of simpler scenarios, making the optimization speed decrease. Adam and his improved algorithms combine the two types of descent algorithms and are gradually used in various practical problems because of their better performance. Use numbers from 0 to 23 instead, you can more intuitively feel the change trend of each evaluation index.

$$\text{Loss}_{\text{reg}} = \text{Loss} + \frac{\lambda}{2n} \sum_w w^2. \quad (5)$$

Parameter regularization means that a penalty term is added to the loss function to constrain the coefficients so that the model cannot be fitted arbitrarily to the random noise in the training data. One term is added to the bias derivative for the weights w , while there is no change for the bias b . Thus, at gradient descent w is updated to

$$w = w + \eta \frac{\partial \text{Loss}_{\text{reg}}}{\partial w}. \quad (6)$$

The notes in the score image is discrete and evenly distributed, mainly consisting of straight lines or curves in multiple directions, solid and hollow near-circle shapes, and some of the notes have the same shape, with only differences in position, while some of the notes are small and easily confused with the noise in the score [17]. The convolutional layer in the CNN has local connectivity and weight sharing, which can facilitate the extraction of edge features and position information of the notes; the activation function layer can enhance the expressiveness of the CNN and make the CNN microscopic, to realize the nonlinear mapping from low-dimensional simple features to high-dimensional complex features in sheet music images; the pooling layer

can reduce the weights while retaining the main features of the convolutional layer. The pooling layer reduces the number of weight parameters while preserving the main features of the convolutional layer to speed up the computation and prevent the overfitting problem. Usually, the width or depth of CNN layers is increased to improve the model accuracy, but the gradient disappearance/explosion problem easily occurs in the process of parameter update, resulting in model nonconvergence, as shown in Figure 2.

When the conditional probability distribution is determined, the network can classify the given input sequence by selecting the most probable label and obtain the final target sequence by maximizing the probability of the label sequence, that is, for a given sheet music image, the probability of the output notes are different, and the selection of each position note in the note sequence constitutes a selectable output path, and the difference in the conditional probability of the output notes leads to the difference in the probability of the selected path. Selecting the note sequence with the highest probability will also increase the probability of outputting the correct sequence. When the environment changes and the times change, it is very easy to be damaged or even lost. Therefore, the complete preservation of paper scores has far-reaching significance. By traversing multiple paths to select the path with the highest probability, we can achieve accurate recognition and classification of notes.

That is, the sound signal of music is converted into a musical score representation. In the sight-singing scenario, since the human voice is monophonic, that is, it can only produce a single-frequency sound per unit of time, the AMT task of monophonic can be converted into pitch extraction and note onset detection of audio, thus converting a piece of audio into a fundamental frequency sequence and slicing the fundamental frequency sequence into a sequence of notes containing pitch with the help of note onset location. Since the existing algorithms for single-pitch extraction are relatively mature, the solution to this problem requires the design of a note onset detection algorithm that satisfies the sight-singing scenario.

The note sequences extracted from the audio need to be aligned with the note sequences of the standard sheet music and then compared one by one. This process needs to consider the deviations in pitch and duration between the sighted singer's pronunciation and the standard score in the sighted singing scenario, as well as the possibility of polyphony and missing notes during the sighted singing process. Therefore, to solve this problem, we need to design a robust note sequence alignment algorithm that meets the sight-singing scenario.

4. Analysis of Music Teaching Evaluation Model

A typical BP neural network consists of an input layer, an implicit layer, and an output layer. The number of neurons in each layer and the number of layers in the implicit layer need to be adjusted according to the actual situation [18]. A reasonable network structure can reduce the number of network training and improve the accuracy of network learning. In the hybrid teaching evaluation of this paper, the

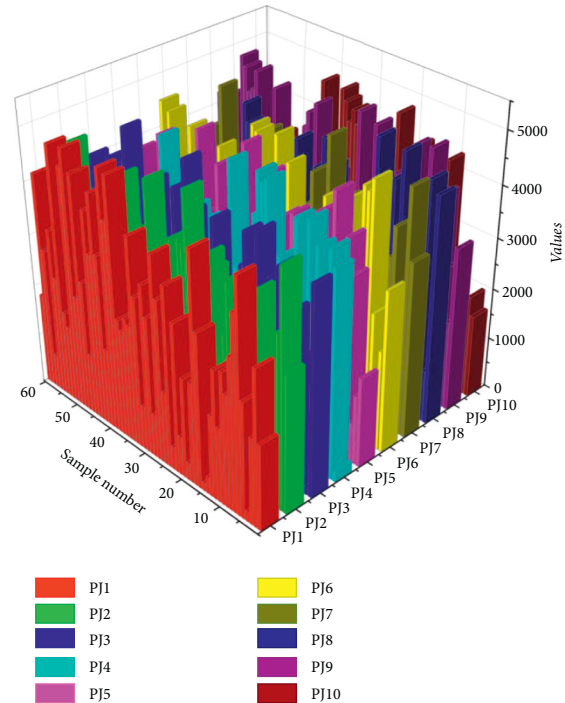


FIGURE 2: The evaluation value of the sample data and its corresponding frequency.

index values in the evaluation system are used as the input values of the BP neural network, and the evaluation results are used as the output values. With enough samples to train, the network can correct the appropriate weights and then predict the evaluation results of teaching quality based on the sample data.

When training modeling with BP neural networks, a range of initial weights and thresholds in the network needs to be set beforehand, which ensure the training does not start by falling on those flat areas that fall into local minima. When setting the weights, a relatively small random number is usually taken, which can effectively shorten the learning time of the network. In this paper, the initial range of connection weights and threshold values of the BP neural network is set to $[-0.5, 0.5]$. A hybrid evaluation model based on the BP neural network is shown in Figure 3.

In the establishment of the hybrid teaching evaluation model, the sample occupies a crucial position, and the sample selection will directly affect the training results of the neural network and the scientific degree of the model establishment. The data generated from the implementation of blended learning this semester were preprocessed to obtain 85 sets of samples. From these data, 70 samples were selected as the training set and input to the designed BP network for training, and the remaining 15 samples were used as the test set to test the trained BP neural network.

After setting the initial values of the weights and thresholds of the BP neural network, MATLAB can automatically mobilize the unit function to initialize it. Some scholars have also put forward the research value prospect of this new breakthrough in their articles, and tried to mine the review data to conduct a “bottom-up” online course quality

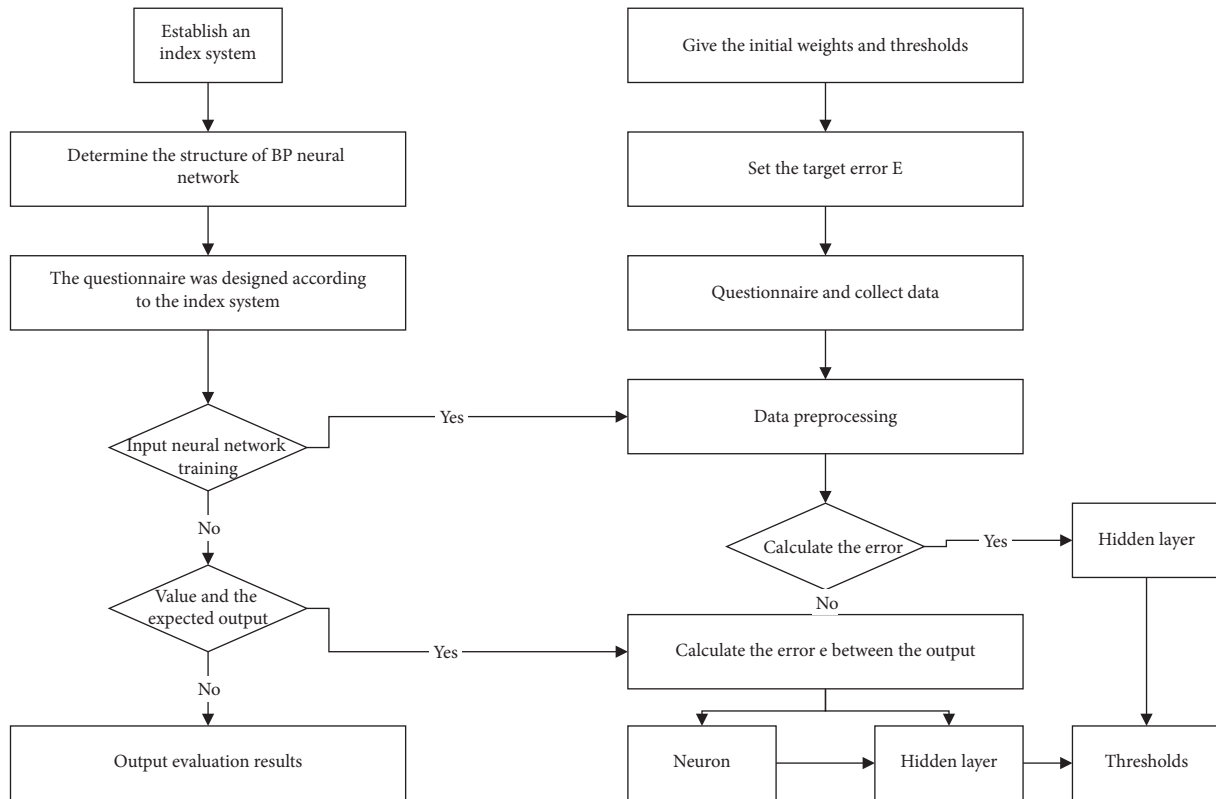


FIGURE 3: Hybrid teaching evaluation model based on BP neural network.

evaluation research, which has achieved certain results. It is relatively simple, and the amount of research is relatively small, which is also the main research direction of this time. The input sample data are preprocessed, and the corresponding network model is generated according to the BP neural network model structure designed in the previous section, and the training process is realized by the function train, which trains the network according to the sample input, target output, and the parameters of the training function that have been set in advance. The training process implements the function train, which trains the network according to the sample input, target output, and the preset parameters of the training function.

The Sim function in MATLAB implements the simulation process, and after training, it can perform simulation calculations on the test set based on the network that has been trained. When a player plays music with an MIDI instrument, the MIDI instrument converts the player's actions on the instrument into MIDI signals, which are transmitted through a cable to the sequencer [19]. The sequencer is a piece of software that stores, edits, and forwards MIDI signals. The MIDI signal stored by the sequencer is the MIDI file, which can be edited and finally output to the sound source.

The MIDI signals coming from an MIDI instrument contain all the possible actions of the player on the instrument, which are acquired and analyzed to extract the characteristics of the sound. In an MIDI system consisting of a personal computer, the MIDI input device is connected to

the computer's sound card, and a sequencer is software installed on the computer, which acquires the MIDI signals from the computer's sound card by calling the operating system's API. The MIDI signal that needs to be extracted for this project can be obtained in the same way, and after obtaining the signal, MIDI signal can be analyzed to obtain the characteristics of the sound, as shown in Figure 4.

As shown in Figure 4, which represents most of the cases processed by the alignment module, where the black squares represent the sight-sung notes processed by the system and the green squares represent the standard notes in the score, the real notes sung by the sight-singers may deviate from the standard notes of the score in terms of absolute pitch and duration. During the early semester, there is no requirement to grasp the timbre of the piano and the emotion of the music, and you can use the MIDI keyboard or electric piano for practice. If the absolute pitch and duration of the sung notes are directly used as features and the DTW and HMM are used to match the note sequences, it is easy to cause a mismatch. Based on the analysis of the sight-singing samples, it is found that the sight-singers, facing the notes in the pentatonic score, can judge the relative pitch height and duration between two adjacent notes through the basic music theory knowledge, and ensure that the relative deviation of duration and pitch between adjacent notes is small, so it can be assumed that the pitch difference between two adjacent notes in the sight-singing audio always keeps the same sign [20].

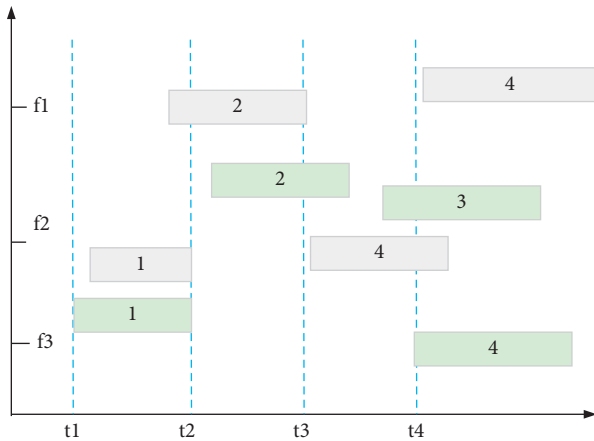


FIGURE 4: Schematic diagram of real singing notes and syncopated notes.

Therefore, the relative pitch relationship between adjacent notes is used to encode to get the melody-like sequence and use this sequence to align with the melody-like sequence extracted according to the same rules of the score, to achieve the effect of matching using only the relative pitches of two adjacent notes. After extracting the fundamental frequencies of all frames to get the fundamental frequency sequence, some points in the sequence exist abnormally. There are mainly two types of cases: one is the extraction of frequencies in the voiceless segment due to background noise and other factors, and the other is the error of candidate values due to the amplitude maxima in the spectrum not belonging to the fundamental frequency and its harmonics. The presence of anomalous frequencies can be reduced by simple post-processing in the sequence.

5. Analysis of Results

5.1. Optimized Neural Network Performance Analysis. The RCL is equivalent to a CNN with elastic paths, and the path length can be extended by adjusting the depth T value. Next, the optimal depth and the optimal number of layers of the RCL are tested. Since the recognition rate of the two-layer RCL is too low, the test will be shown starting from the number of layers $L=3$. Figure 5 shows the accuracy, precision, recall, and F1 scores of the RCL for different depths and a different number of layers. The results in Figure 5 are presented in the form of line graphs, where 24 combinations of layers L from 3 to 6 and depths T from 3 to 8 are used instead of numbers from 0 to 23, respectively, to get a more intuitive feeling of the trend of each evaluation index.

It can be seen from Figure 5 that all the evaluation indexes reach their maximum values at $L=5$ and $T=6$, except for the highest accuracy rate at $L=4$ and $T=6$. It can be seen from Figure 5 that the overall trend of the line graph increases and then decreases, and at the horizontal coordinate of 15, all the indexes except the accuracy rate reach the highest point, and the horizontal coordinate of 15 corresponds to the depth $L=5$ and $T=6$. Therefore, it can be concluded that when the depth is less than $L=5$ and $T=6$,

the performance increases with the depth, and when the depth is greater than $L=5$ and $T=6$, the performance decreases with the depth and the training time also decreases. When the depth is greater than $L=5$ and $T=6$, the performance decreases gradually with the increase of depth and the training time also increases gradually. Therefore, the overall performance of the model is optimal when $L=5$ and $T=6$. The next experiments are conducted based on $L=5$ and $T=6$.

The overall recognition rate is 77.36% when $\text{Batch_size}=64$, and 80.44% when $\text{Batch_size}=32$. In addition, the recognition rate of individual sentiment in the four confusion matrices is above 60%, and the recognition rate of individual sentiment can even reach 90%. The recognition rate of individual sentiment can even reach more than 90%.

The activation function is a nonlinear function that exists between the upper and lower neuron nodes and has an important role in neural networks: Without the activation function, the output of the network is just a linear combination of the inputs and the approximation capability of the network will be very limited. Therefore, activation functions are needed in neural networks to improve the nonlinearity of the network, and different functions will have different effects on the performance of the model. The structure is simple and easy to implement. For example, the Back-propagation neural network is a typical feedforward network. Unlike feedforward networks, each neuron in a feedback network feeds back its output signal as an input to the next neuron.

The variance of the output values is smaller than the variance of the actual values, indicating that the fluctuations between the output values and the mean are somewhat smaller, and the reason for this situation is due to the small amount of data with higher and lower scores at the end of the period in this training set, which adds difficulty to the machine learning and relatively poor learning. Also, the result tells another aspect that deep learning using neural networks must require a sufficiently large number of training samples, as shown in Figure 6.

As a variant of RNN, the computation of the gate state is independent of the previous moment information output, so it can increase the parallel computation and accelerate the convergence speed of the model by solving the dependence problem of long-range information. In this paper, SRU combined with CTC for note recognition, and the experimental results show that SRU converges nearly three times faster than the LSTM network with guaranteed model accuracy. Considering that there is no publicly available data set containing audio pitch annotation, onset annotation, and score information, this paper builds a test set for sight-singing evaluation by randomly selecting 30 audio tracks from the self-built data set HUST-Solfege.

The corresponding relationship between the true sight-singing pitch and the notes of the score is marked in the test set for each sample to obtain the smallest training error. BP neural network consists of input layer, hidden layer, and output layer. The audio in the test set was graded according to the CRS evaluation method introduced in the subsection,

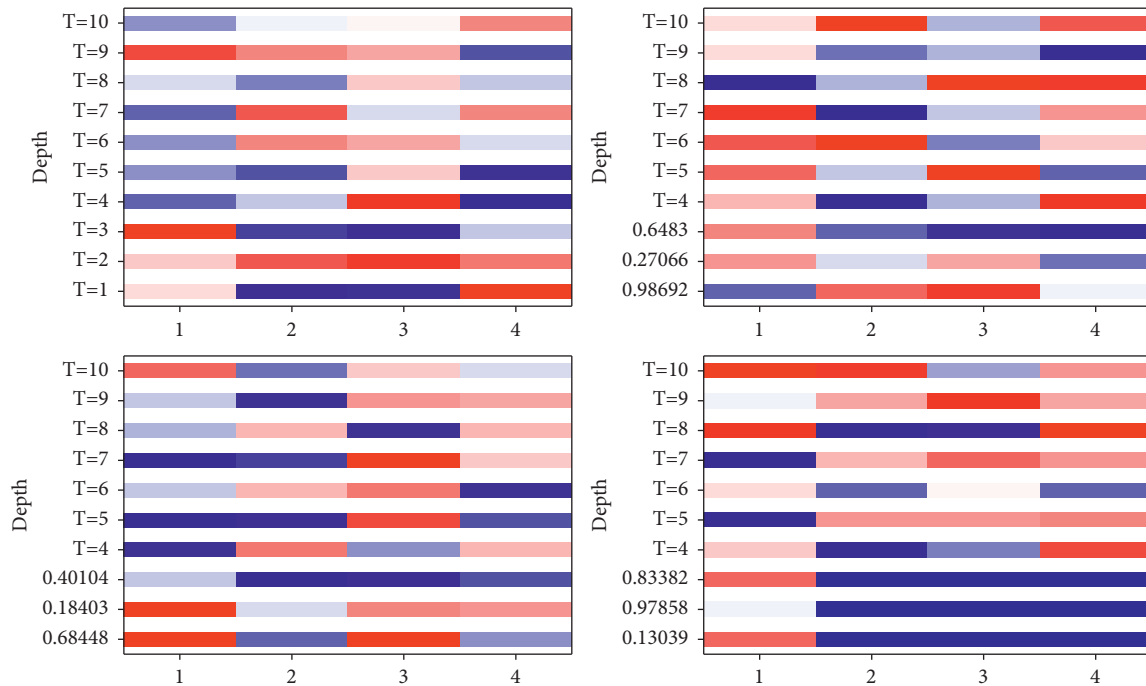


FIGURE 5: Same depth comparison results.

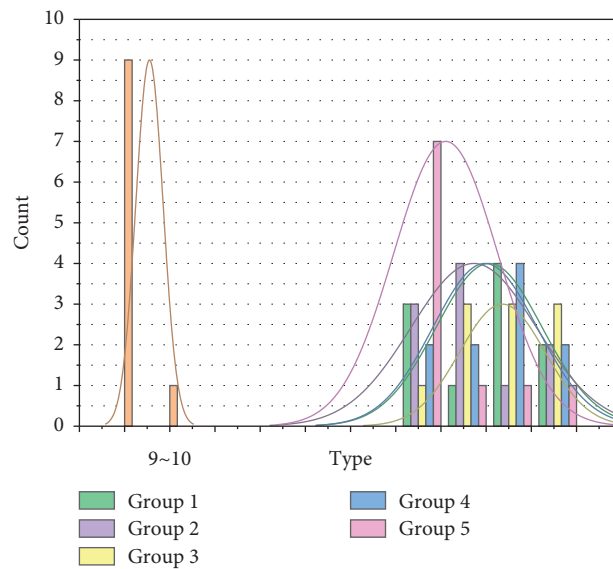


FIGURE 6: Distribution of absolute errors between actual and predicted values of two sets of predicted data.

and the gradation of the test set and the correct rate of each song were published in the HUST-Solfege data set.

Design of model topology based on the aforementioned design idea, building a good neural network that can correctly predict the target sample through training-like learning of 150 students is the first problem to be solved. The neuron (processing unit) is the core of the whole network, and it is important to make a reasonable design for it. The design of the neuron must be based on its four core elements, namely, the weight (a positive value means the neuron activated, a negative value means the neuron is inhibited), the summation unit (to find the weighted sum of the input

signals, which is essentially a linear combination), the excitation function (to realize the nonlinear mapping), and the output range of the neuron and limits the output range of the neuron), and the threshold is also called bias or intercept (used to measure the degree of positive or negative excitation generated by the neuron).

5.2. Music Teaching Evaluation Results. In this paper, the structures of the two models designed separately, the error analysis of the evaluation results obtained from the training was conducted, and the hybrid teaching evaluation models

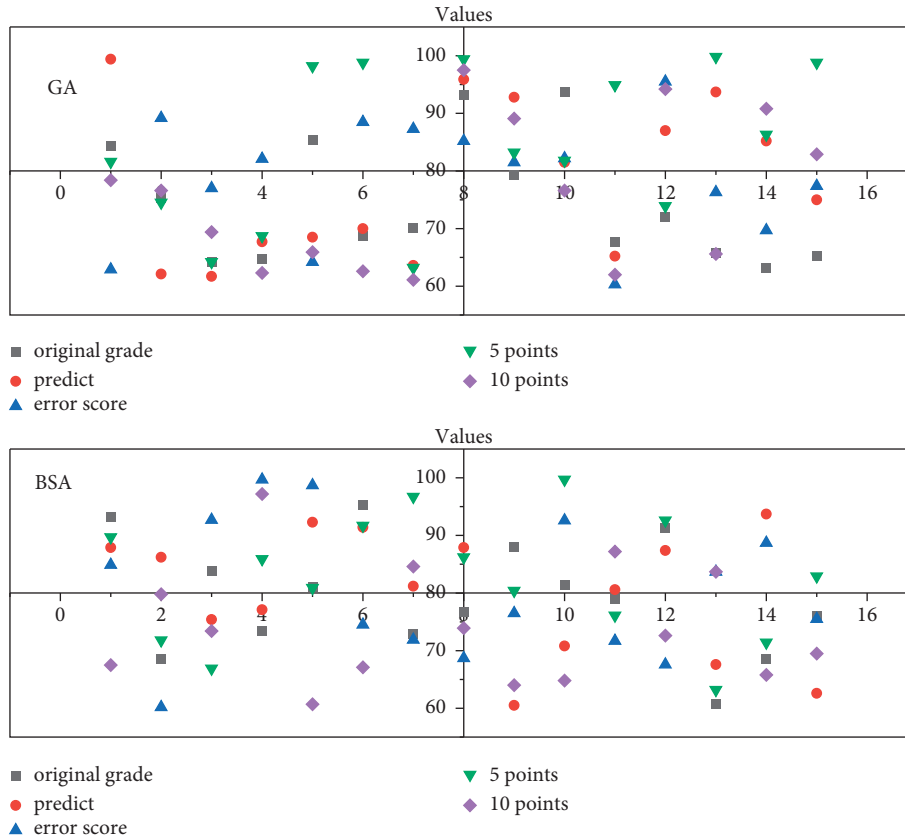


FIGURE 7: Evaluation results.

based on BP neural network and GA-BP were established. To better reflect the effects of the two neural network evaluation models, BP and GA-BP, the original GA and BSA algorithms were also used in this section to predict the same 15 sets of test set data in the sample data for comparative analysis, and the evaluation results of GA and BSA are shown in Figure 7.

Although the traditional evolutionary algorithms (GA, BSA) can produce prediction results based on students' performance in each category, the evaluation results of GA and BSA are highly variable, and some evaluation results are even outside the normal scoring range (100 points), while the BP neural network algorithm can predict students' performance within the scoring range regardless of the improvement. Figure 7 compares the evaluation results of the GA-BP neural network evaluation model with BP, GA, and BSA in the last 15 sets of test sample data. This result is more obvious when the evaluation results as well as the errors are made into a comparative box plot.

The evaluation of the original BP neural network algorithm fluctuated more, and the evaluation results of GA and BSA algorithms, though a little smaller than those of the BP neural network, all showed extreme data, while the error of the evaluation results of the GA-BP model was controlled in a good range, and the maximum error did not exceed 8 points. These two processes can be repeated until the output error of the network reaches the previously set range, or reaches a predetermined number of network learning times, and the entire learning process ends. Therefore, it can be

obtained that the hybrid teaching quality evaluation model based on GA-BP neural network is better in evaluation accuracy and can make a more effective evaluation of hybrid quality.

Comparing the recognition results of the four methods, the accuracy of the three methods, except SmartScore, is higher for the recognition of the pitch in notes and the recognition of the six consecutive notes. Capella-scan is better than PhotoScore in the recognition of legato lines and sustains lines, although it has errors in the recognition of legato lines and sustains lines. The proposed algorithm has less error in the identification of legato and sustains lines, the relative positions of pitch marks and notes, and note beams, but it has problems in the identification of some of the clef positions.

As shown in Figure 8, the algorithm in this paper has a low symbol error rate for the data set containing noisy and deformed musical score images. The enhanced recognition ability of this algorithm for the presence of noisy score examples indicates that the robustness of the model is improved by this algorithm. Compared with commercial software, this algorithm has a lower symbol error rate and is superior in the recognition of rests, note beams, legato lines, and sustain lines.

Firstly, the sheet music image data set is expanded to increase the complexity of the data set; secondly, a residual CNN is used in the note feature extraction to solve the potential gradient disappearance problem of the model;

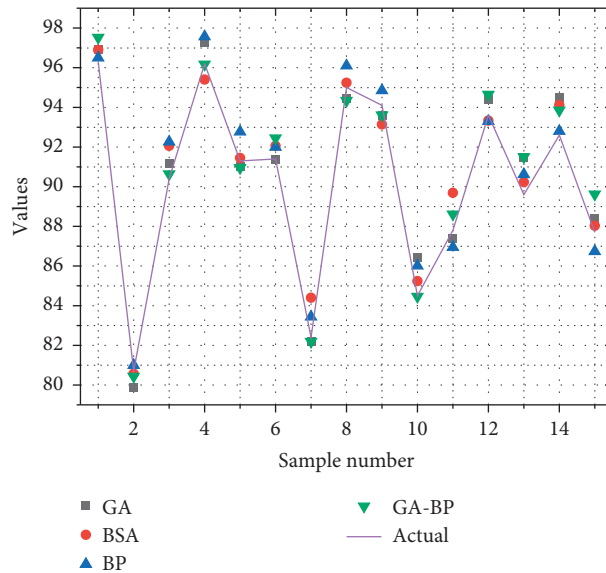


FIGURE 8: Comparison of evaluation results.

subsequently, a multi-scale feature fusion technique is used in the feature extraction to fuse the feature information extracted from different convolutional layers in the same feature to improve the model generalization ability and enhance the model feature characterization ability; finally, a variant SRU network of RNN is used to accelerate the model convergence speed. Finally, a variant of RNN, the SRU network, is used to accelerate the convergence speed of the model.

The ordinary deep convolutional network can extract the feature information of difficult notes, but at the same time, some detailed information is lost. To address the characteristics of note recognition requiring detailed information, the deep convolutional layer feature information is fused with the shallower layer information to obtain a feature map containing rich information, which enhances the learning ability of the model for detailed features and improves the subsequent recognition ability of the model. Extensive experimental results show that the feature extraction network combining multi-scale feature fusion and residual CNN can effectively reduce the symbol error rate of the model and solve the model non-convergence problem.

6. Conclusion

Then the factors influencing the quality evaluation of blended teaching were analyzed from the perspective of two main subjects, teachers, and students. Therefore, in most cases, choose $F = 3, S = 2$ or $F = 2, S = 2$. The local information of the feature map of the previous layer is all connected to the output neurons, and the parameters are trained through the forward propagation and back propagation of the weights and biases, and the classification is realized by the activation function. Finally, the data analysis results and offline questionnaires generated in the process of implementing blended teaching are used as reference bases to develop a blended teaching quality evaluation index table

and establish a blended teaching evaluation system with preclass, in-class, and postclass teaching evaluation as the primary indexes and 20 items in the index table as the secondary indexes. In the process of recognition, Capella-scan tends to identify rests as numbers and score numbers as rests, and such errors are concentrated, especially in the data set containing noisy and deformed score images. By designing the structure of the BP neural network and analyzing the error of the training results, a hybrid teaching evaluation model based on BP neural network was established. To obtain a more accurate evaluation model, the GA-BP algorithm is proposed to be used in the hybrid teaching quality evaluation by analyzing the shortcomings of the BP neural network, which is optimized by the genetic algorithm and has better performance and less error in the training results, thus establishing a hybrid teaching evaluation model based on the GA-BP neural network. Finally, the evaluation results of these two models are compared with those of single GA and BSA algorithms, and the error analysis is performed, which proves that the hybrid teaching evaluation model based on GA-BP neural network has better results.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Humanity Project of Anhui Province (project nor: SK2021A0622): Seeking the Path for the Inheritance and Innovation of Luzhou Song Guided by Song with Peking Opera.

References

- [1] K. Yu, L. Tan, S. Mumtaz et al., "Securing critical Infrastructures: Deep-learning-based Threat Detection in IIoT," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 76–82, 2021.
- [2] L. Tan, K. Yu, L. Lin et al., "Speech emotion recognition enhanced Traffic efficiency solution for Autonomous Vehicles in a 5G-Enabled space-Air-Ground Integrated intelligent Transportation system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 2830–2842, 2021.
- [3] F. Ding, G. Zhu, M. Alazab, X. Li, and K. Yu, "Deep-learning-empowered Digital Forensics for edge consumer electronics in 5G HetNets," *IEEE Consumer Electronics Magazine*, vol. 11, no. 2, pp. 42–50, 2022.
- [4] Z. Guo, Y. Shen, S. Wan, W. Shang, and K. Yu, "Hybrid intelligence-Driven Medical image recognition for Remote Patient Diagnosis in Internet of Medical Things," *IEEE Journal of Biomedical and Health Informatics*, p. 1, 2021.
- [5] C. Feng, B. Liu, K. Yu, S. K. Goudos, and S. Wan, "Blockchain-empowered Decentralized horizontal Federated learning for 5G-Enabled UAVs," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 5, pp. 3582–3592, 2022.
- [6] P. W. Van and A. Chesnokova, "Reading and rereading: insights into literary evaluation," *Advanced education*, vol. 5, no. 9, pp. 39–46, 2018.
- [7] S. Zuhriyah and B. W. Pratolo, "Exploring students' Views in the Use of Quizizz as an assessment Tool in English as a Foreign language (EFL) class," *Universal Journal of Educational Research*, vol. 8, no. 11, pp. 5312–5317, 2020.
- [8] S. Sulistyani and R. Riwayatinationsih, "Modeling online classroom interaction to support student language learning," *IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature*, vol. 8, no. 2, pp. 446–457, 2020.
- [9] D. P. Shea, "Compelled to speak: Addressing student reticence in a university EFL classroom," *The Asian Journal of Applied Linguistics*, vol. 4, no. 2, pp. 173–184, 2017.
- [10] S. L. Campbell and M. Ronfeldt, "Observational evaluation of teachers: Measuring more than we bargained for?" *American Educational Research Journal*, vol. 55, no. 6, pp. 1233–1267, 2018.
- [11] T. Ounis, "Exploring secondary teachers' perceptions of classroom assessment in a Tunisian context," *International Journal of Language and Linguistics*, vol. 4, no. 2, pp. 116–124, 2017.
- [12] M. A. Dakhiel, "Factors affecting the quality of English language teaching in Preparatory Year, university of Jeddah," *English Language Teaching*, vol. 10, no. 7, p. 43, 2017.
- [13] A. M. Aljaser, "Effectiveness of using flipped classroom Strategy in Academic Achievement and self-Efficacy among education students of Princess Nourah bint Abdulrahman university," *English Language Teaching*, vol. 10, no. 4, p. 67, 2017.
- [14] A. M. Songbatumis, "Challenges in teaching English faced by English teachers at MTsN Taliwang, Indonesia," *Journal of foreign language teaching and learning*, vol. 2, no. 2, pp. 54–67, 2017.
- [15] D. Zou, "Gamified flipped EFL classroom for primary education: student and teacher perceptions," *Journal of Computers in Education*, vol. 7, no. 2, pp. 213–228, 2020.
- [16] M. K. Ibrahim and Y. A. Ibrahim, "Communicative English language teaching in Egypt: classroom practice and challenges," *Issues in Educational Research*, vol. 27, no. 2, pp. 285–313, 2017.
- [17] M. C. Limlingan, C. M. McWayne, E. A. Sanders, and M. L. Lopez, "Classroom language contexts as Predictors of Latinx Preschool Dual language learners' School Readiness," *American Educational Research Journal*, vol. 57, no. 1, pp. 339–370, 2020.
- [18] M. Ahmad, A. Shakir, and A. R. Siddique, "Teacher-student interaction and management practices in Pakistani English language classrooms," *Journal of Language and Cultural Education*, vol. 7, no. 3, pp. 115–134, 2019.
- [19] N. Yürük, "Edutainment: using Kahoot! As A review activity in Foreign language classrooms," *Journal of Educational Technology and Online Learning*, vol. 2, no. 2, pp. 89–101, 2019.
- [20] D. Meng, Y. Xiao, Z. Guo et al., "A data-driven intelligent planning model for UAVs routing networks in mobile Internet of Things," *Computer Communications*, vol. 179, pp. 231–241, 2021.