

Research Article

Piano Music Teaching under the Background of Artificial Intelligence

Nansong Huang ¹ and Xiangxiang Ding²

¹*Xi'an Conservatory of Music, Xi'an, 710061 Shaanxi, China*

²*Xi'an Institute of Physical Education, Art College of Xi'an Institute of Physical Education, Xi'an, 710068 Shaanxi, China*

Correspondence should be addressed to Nansong Huang; 04046@xacom.edu.cn

Received 9 October 2021; Revised 21 December 2021; Accepted 29 December 2021; Published 24 January 2022

Academic Editor: Narasimhan Venkateswaran

Copyright © 2022 Nansong Huang and Xiangxiang Ding. 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.

The work is aimed at solving the problems of easy trapping into local extremes and slow convergence speed of the traditional music teaching evaluation system on Backpropagation Neural Network (BPNN). The traditional note recognition methods are susceptible to high noise complexity. Firstly, the Levenberg Marquardt (LM) algorithm is used to optimize the BPNN; secondly, an improved endpoint detection algorithm is proposed by short-term energy difference, which can accurately identify the time value of each note in the piano playing audio. By the traditional frequency domain analysis method, a radical frequency extraction algorithm is proposed by the improved standard harmonic method, which can accurately identify the note's pitch. Finally, a piano performance evaluation model by BPNN is implemented, and the model is implemented by the Musical Instrument Digital Interface (MIDI) system. This evaluation model can be used to correct the errors of students' performances in the piano music teaching process and to perform overall evaluation, rhythm evaluation, and expressive evaluation. Teachers and students play *minuet* to collect experimental samples to train BPNN and test the performance of the evaluation model. The practical result shows that (1) after 3000 times of training, the neural network error is less than 0.01, and the network converges; (2) the evaluation results of the piano performance evaluation model designed are basically in line with the actual level of the performer and have specific feasibility; and (3) the optimized BPNN is used to correct errors during performances with an accuracy rate of 94.3%, which is 5.25% higher than the traditional method. The error correction accuracy rate for pitch is 92.9%, which is 5.21% higher than the traditional method. The optimized BPNN has significantly improved the error correction accuracy of the notes and pitches played by the player. The model can effectively help piano beginners correct errors and improve the accuracy and efficiency of the practice. The purpose of this study is to alleviate the scarcity of piano teachers, reduce the work intensity of piano teachers, realize automatic error correction and objective evaluation of playing, and provide necessary technical support for improving the efficiency of piano music teaching.

1. Introduction

As an essential tool for delivering music, the piano has been popularized worldwide and is favored by more and more people [1]. However, there are still some unsolved problems in piano music education. For example, due to the relative scarcity of piano teachers, most piano learners do not receive sufficient guidance in learning and practice. They are prone to misreading the score, playing the wrong keys, and making mistakes in fingering during the independent course [2]. In addition, the piano teacher mainly relies on his own experi-

ence to guide, evaluate, and correct errors in the students' performance. Piano teachers and students also have a different understanding of music and playing characteristics. The factors that affect the commission are the right or wrong notes and essential elements such as rhythm and expressiveness [3]. Therefore, traditional piano music teaching methods have shortcomings such as intense subjectivity, preliminary evaluation, and high uncertainty [4]. Artificial Neural Network (ANN) is an algorithmic mathematical model that imitates the behavioral characteristics of animal neural networks and performs distributed parallel information processing [5].

Among them, Backpropagation Neural Network (BPNN) is one of the more mature and widely used ANNs, and it has an excellent performance in data classification, prediction, and related evaluation [6].

Fei used BPNN to establish an evaluation model. The vocal evaluation system uses input to solve the problem that vocal players are greatly affected by subjective factors when scoring. MATLAB analyzes and processes the collected data and compares the results with personal evaluation methods. Simulation results show that this scheme significantly improves music classification and recognition accuracy and robustness. The system can reflect the player's true level [7]. Jia proposed to use BPNN to simulate the nonlinear mapping of various factors and establish a vocal music evaluation system to address the problem of multiple factors and subjective factors in the evaluation of vocal music teaching. Meanwhile, with the vocal music teaching evaluation system as input, BPNN establishes a teaching quality evaluation model. The model's validity has been verified through simulation experiments [8]. BPNN is introduced into the evaluation system, which can effectively avoid the influence of human subjective factors, make the evaluation results more objective and accurate, and help improve students' comprehensive ability.

Traditional BPNN has defects such as ease to fall into local extremum and slow convergence speed [9]. It is necessary to use appropriate methods to optimize it. The premise of piano performance evaluation is the recognition of the played notes. Traditional musical note recognition methods are mainly divided into the time and frequency domains; traditional time-domain analysis methods are susceptible to noise. The frequency-domain analysis method has disadvantages such as high algorithm complexity and extensive calculation [10]. Based on the above problems, the Levenberg Marquardt (LM) algorithm uses to optimize BPNN. Secondly, an improved endpoint detection algorithm is proposed by short-term energy difference, which can accurately identify the time value of each note in the piano playing audio. A radical frequency extraction algorithm by the improved standard harmonic method is put forward, which can accurately identify the note's pitch. Finally, a piano performance evaluation model by BPNN is implemented. The model is realized mainly by the Musical Instrument Digital Interface (MIDI) system. This system can assign a value to each note in the score, convert it into a MIDI standard signal, and store it as a MIDI file. The operation process of the performance evaluation model is to identify the time value and pitch of each note played by the performer using improved time domain and frequency domain algorithms. The corresponding MIDI standard file compares time values and pitch information read. Errors in playing are pointed out to realize the error correction function. This detection algorithm can perform overall evaluation, rhythm evaluation, and expressiveness evaluation of the performer through the comparison results. The study is aimed at realizing automatic error correction and objective assessment of playing and providing necessary technical support for improving the efficiency of piano music teaching.

2. Materials and Methods

2.1. Introduction to BPNN and Its Optimization

2.1.1. Introduction to BPNN. The BP neural network (BPNN) is a typical representative of ANN, and it is also the most widely used ANN [11]. BPNN is produced by simulating the structure of the human brain neuron network, which is a complex network composed of many nodes connected [12]. BPNN is a multilayer perceptron structure that mainly contains a three-layer structure of input, hidden, and output [13]. The three-layer BPNN structure is shown in Figure 1.

The input layer and output layer mainly store and transmit external information. All networks contain an input layer and an output layer. The main difference lies in the number of hidden layers in the middle. The hidden layer does not directly communicate with the outside world, but its change will directly impact the relationship between the input layer and the output layer [14]. The BP algorithm includes two signal forward and backward propagation processes in the learning process. The forward propagation is from the input layer to the output layer. If the actual and expected output signals are too far apart, it needs to enter the reverse transmission process [15]. Back-propagation is to pass the output error back layer by layer in the direction of the input layer through the hidden layer, distribute it to all units in each layer, and adjust the weight of each unit by the error signal obtained by each layer. Then, change the connection strength and threshold among the input layer, the output layer, and the hidden layer so that the error can be gradually reduced. Repeat the previous process and the learning will be terminated until the error tends to the allowable range or reaches the preset practice frequency [16].

In short, the BP algorithm can transform signal input and output problems into nonlinear optimization problems [17]. BPNN has simple learning rules and a solid nonlinear fitting ability. The trained neural network can also give appropriate output for the input near samples not in the sample set. The BP algorithm includes two signal forward and backward propagation processes in the learning process.

(1) Forward Transmission. The forward propagation is carried out from the input to the hidden layer and then to the output layer. The input α_i of the i th node in the hidden layer can be expressed as

$$\alpha_i = \sum_{j=1}^m w_{ij}x_j + \theta_i, \quad j = 1, 2, 3, \dots, m, \quad (1)$$

where x_j is the input of the j th node of the input layer, w_{ij} is the weight between the i th node in the hidden layer and the j th node in the input layer, and θ_i is the threshold of the i th node in the hidden layer.

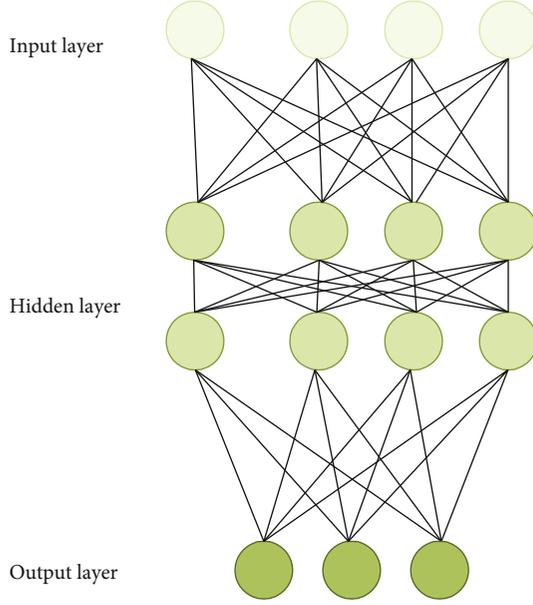


FIGURE 1: Three-layer BPNN structure.

The output β_i of the i th node in the hidden layer can be expressed as

$$\beta_i = \phi(\alpha_i) = \phi\left(\sum_{j=1}^m w_{ij}x_j + \theta_i\right), \quad (2)$$

where $\phi(\alpha_i)$ represents the activation function of the hidden layer.

The input of the k th node in the output layer δ_k can be expressed as

$$\delta_k = \sum_{i=1}^q w_{ki}\beta_i + a_k = \sum_{i=1}^q w_{ki}\phi\left(\sum_{j=1}^m w_{ij}x_j + \theta_i\right) + a_k, \quad i = 1, 2, 3, \dots, q, \quad (3)$$

where w_{ki} represents the weight between the k th node in the output layer and the i th node in the input layer and a_k represents the threshold of the k th node in the output layer.

The output y_k of the k th node of the output layer can be expressed as

$$y_k = \psi(\delta_k) = \psi\left(\sum_{i=1}^q w_{ki}\beta_i + a_k\right) = \psi\left(\sum_{i=1}^q w_{ki}\phi\left(\sum_{j=1}^m w_{ij}x_j + \theta_i\right) + a_k\right), \quad (4)$$

where $\psi(\delta_k)$ represents the excitation function of the output layer. If the error between the actual and expected output signals is too large, it needs to enter the backpropagation process.

(2) *Backpropagation.* Backpropagation is to pass the output error back layer by layer in the direction of the input layer through the hidden layer and distribute it to all units in each layer and adjust the weight of each team based on the error signal obtained by each layer. Meanwhile, the connection strength and threshold between the input layer,

output layer, and hidden layer are adjusted so that the error can be gradually reduced. This process is repeated until the error is within the allowable range or reaches the preset practice frequency. The learning will be terminated [18]. The quadratic error criterion function E_p of each sample p can be expressed as

$$E_p = \frac{1}{2} \sum_{k=1}^l (T_k - y_k)^2, \quad (5)$$

where T_k represents the expected output and y_k represents the actual output.

The model's total error criterion function for P training samples can be expressed as

$$E = \frac{1}{2} \sum_{P=1}^P \sum_{k=1}^l (T_k^P - y_k^P)^2. \quad (6)$$

According to the error gradient descent method, the weights and thresholds of the output layer and the hidden layer are sequentially modified. The weight correction of the output layer is Δw_{ki} , and the threshold correction is Δa_k ; the weight correction of the hidden layer is Δw_{ij} , and the threshold correction is $\Delta \theta_i$, as shown in

$$\begin{aligned} \Delta w_{ki} &= -\eta \frac{\partial E}{\partial w_{ki}}, \\ \Delta a_k &= -\eta \frac{\partial E}{\partial a_k}, \\ \Delta w_{ij} &= -\eta \frac{\partial E}{\partial w_{ij}}, \\ \Delta \theta_i &= -\eta \frac{\partial E}{\partial \theta_i}, \end{aligned} \quad (7)$$

where η represents the learning rate, as shown in

$$\begin{aligned} \Delta w_{ki} &= -\eta \sum_{P=1}^P \sum_{k=1}^l (T_k^P - y_k^P) \cdot \psi(\delta_k) \cdot \beta_i, \\ \Delta a_k &= \eta \sum_{P=1}^P \sum_{k=1}^l (T_k^P - y_k^P) \cdot \psi(\delta_k), \\ \Delta w_{ij} &= \eta \sum_{P=1}^P \sum_{k=1}^l (T_k^P - y_k^P) \cdot \psi(\delta_k) \cdot w_{ki} \cdot \phi(\alpha_i) \cdot x_j, \\ \Delta \theta_i &= \eta \sum_{P=1}^P \sum_{k=1}^l (T_k^P - y_k^P) \cdot \psi(\delta_k) \cdot w_{ki} \cdot \phi(\alpha_i). \end{aligned} \quad (8)$$

To improve accuracy, reduce network errors, and avoid the phenomenon of "overfitting," the network architecture usually only contains an input layer, an output layer, and a

hidden layer [19], and the number of nodes is generally definite in the input layer and output layer. Therefore, the number of nodes needs to be determined in the hidden layer. The available number of hidden layer nodes is used to determine, as shown in

$$y = \sqrt{a + b} + c, \quad (9)$$

where y represents the number of hidden layer nodes, a represents the number of input layer nodes, b represents the number of output layer nodes, and c represents a constant between 1 and 10.

2.1.2. LM Optimization Algorithm. Traditional BPNN has defects such as easy to fall into local extreme value and slow convergence speed. The LM algorithm is chosen to optimize the BPNN. The LM algorithm obtains the extreme importance of the function through iteration, which is the product of the combination of the Gauss-Newton iteration method and the gradient descent method. It combines the former's local convergence and the latter's global characteristics [20]. Generally, the gradient descent method drops faster at the beginning. As the target approaches the optimal value, the gradient tends to zero, making the decline of the objective function slower. The Gauss-Newton iteration method can produce an ideal search direction near the optimal value. Therefore, the LM algorithm is used to optimize BPNN. It can solve the problem that it is easy to fall into extreme local importance, effectively reducing the computational complexity, reducing the number of network iterations, and speeding up the convergence speed.

x_k and x_{k+1} are two points in the function $f(x)$, and the relationship is shown in

$$x_{k+1} = x_k + \Delta x. \quad (10)$$

Newton's method is shown in

$$\Delta x = -[\nabla^2 E(x)]^{-1} \nabla E(x), \quad (11)$$

where $E(x)$ is the error-index function, $\nabla^2 E(x)$ is the Hessian matrix of $E(x)$, and $\nabla E(x)$ is the gradient. $E(x)$ can be expressed as

$$E(x) = \frac{1}{2} e^2(x), \quad (12)$$

where $e(x)$ represents the error, as shown in Equation (13):

$$\nabla E(x) = J^T(x)e(x), \quad (13)$$

$$\nabla^2 E(x) = J^T(x)e(x) + S(x), \quad (14)$$

$$S(x) = \sum e_i(x) \nabla^2 e_i(x), \quad (15)$$

where $S(x)$ represents the Jacobian (J) matrix, as shown in

$$J(x) = \begin{bmatrix} \frac{\partial e_{1(x)}}{\partial x_1} & \frac{\partial e_{1(x)}}{\partial x_2} & \dots & \frac{\partial e_{1(x)}}{\partial x_n} \\ \vdots & \ddots & \dots & \vdots \\ \frac{\partial e_{n(x)}}{\partial x_1} & \frac{\partial e_{n(x)}}{\partial x_2} & \dots & \frac{\partial e_{n(x)}}{\partial x_n} \end{bmatrix}. \quad (16)$$

The calculation method of the Gauss-Newton iteration method is shown in

$$\Delta x = [J^T(x)J(x)]^{-1} J(x)e(x). \quad (17)$$

The improved Gauss-Newton iteration method is the LM algorithm, as shown in

$$\Delta x = [J^T(x)J(x) + \mu I]^{-1} J(x)e(x), \quad (18)$$

where μ represents the proportionality coefficient, a constant greater than 0, and I is the identity matrix.

2.2. Analysis of Improved Endpoint Detection Algorithm by the Short-Term Energy Difference. Endpoint detection is aimed at finding the start and endpoints of the musical tone signal in a segment of the audio signal to obtain the characteristic of the length of each note in the audio [21]. Audio is inevitably mixed with noise during the recording process. The core of the endpoint detection algorithm is to accurately identify the beginning and end of the music segment from the background noise [22]. At present, the most commonly used endpoint detection algorithm is the double-threshold method by signal time-domain characteristic parameters. It mainly finds endpoints by setting three thresholds and adopting a secondary decision. However, the dual-threshold detection algorithm has shortcomings such as excessive reliance on threshold setting, difficulty in threshold setting, and poor noise resistance [23]. This study proposes an improved endpoint detection algorithm by short-term energy difference by the above problems.

This algorithm mainly uses the short-term energy difference to find the energy mutation information to determine the starting point of the note. Then, according to the starting point position, the end position corresponding to each starting point is determined by designing two layers of judgment. Regarding each note pressed during piano playing as a short-term energy pulse, the audio signal needs to be framed and windowed and its short-term energy calculated, as shown in

$$E_i = \sum_{N=0}^{L-1} |x(n)|^2. \quad (19)$$

In Equation (19), $x(n)$ is the amplitude of the n th point in the i th frame signal and L is window length; the value is related to the sampling frequency.

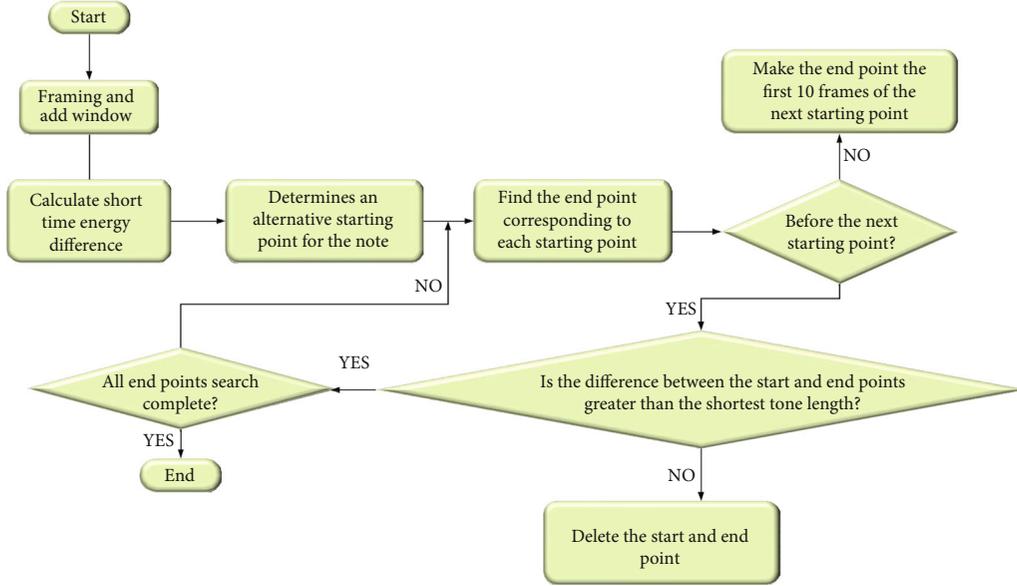


FIGURE 2: Architecture diagram of improved endpoint detection algorithm by the short-term energy difference.

Then calculate the short-term energy difference ΔE_i between two adjacent frames, as shown in

$$\Delta E_i = E_i - E_{i-1}. \quad (20)$$

This algorithm is not to calculate the energy difference between two sampling points but to calculate the energy difference between two frames to filter out the small energy fluctuations in the audio signal. Meanwhile, the difference operation can better reflect the sudden change of energy, and it is easier to judge the starting position of the note. Finding the endpoint corresponding to the starting point of each message is mainly by setting two thresholds, short-term energy and short-term zero-crossing rate. When the two parameters of the signal are both lower than the threshold, the point is determined as the rough endpoint corresponding to the start point of the current note. Then, this study makes a two-level judgment on the rough judgment endpoint to improve the algorithm's accuracy. The first level of assessment is mainly for the rough end position obtained in the previous step. If the endpoint corresponding to the current start point is located after the start point of the following note, the endpoint search is wrong. The n frames before the start of the following message are regarded as the endpoint corresponding to the start point, $n = 10$. The second-level judgment is for the difference between each pair of start and endpoints. The difference between the start and endpoints is calculated for each pair. If the difference is less than the shortest duration of the note, then the pair of start and endpoints are judged as noise. It is deleted from the collection. The architecture of the improved endpoint detection algorithm by the short-term energy difference is shown in Figure 2.

2.3. Analysis of Radical Frequency Extraction Algorithm by the Improved Standard Harmonic Method. Radical fre-

quency refers to the pure tone with the lowest frequency in each musical style, and its intensity is the largest. The frequency of the fundamental tone is the radical frequency, which directly determines the pitch of the entire tone [24]. Extreme frequency extraction methods mainly include time-domain-based algorithms, frequency-domain-based algorithms, and statistics-based algorithms. The frequency-domain-based radical frequency extraction algorithms are mainly used here. The extraction algorithm by the frequency domain is divided into two types: the harmonic peak method and the confidence coefficient. Among them, the harmonic peak method is a typical algorithm by Fast Fourier Transform (FFT), which mainly reflects the relationship between signal frequency and amplitude and is widely used to calculate signal spectrograms. The Fourier Transform $F(x)$ of the nonperiodic continuous signal $f(t)$ is shown in

$$F(x) = F[f(t)] = \int_{-\infty}^{+\infty} f(t)e^{-j\omega t} dt. \quad (21)$$

In Equation (21), ω refers to frequency. The harmonic peak method considers that the peak with the highest amplitude in the spectrogram corresponds to the fundamental wave of the audio signal and uses its frequency value as the radical frequency value; that is, if the extreme frequency value is x_0 , then there is

$$x_0 = F^{-1}(\max(F(x))). \quad (22)$$

The harmonic peak method has the advantages of low time complexity and low space complexity. In practical applications, even in the sonic spectrum of a single piano tone, the amplitude of the radical frequency is not necessarily the highest, so the accuracy of this method is low.

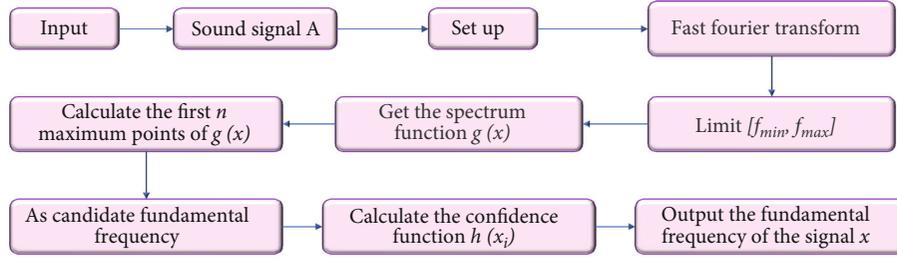


FIGURE 3: The calculation flow chart of the improved standard harmonic method.

The confidence coefficient is mainly proposed for the problem that the peak amplitude of the harmonic is higher than the peak value of the fundamental wave. In the confidence coefficient method, the entire wave or the n th ($n \leq 5$) harmonic component has the most significant peak amplitude. Therefore, a factor of 1 to 5 can be obtained for the maximum peak frequency as a candidate radial frequency. The sum of the amplitudes of the n th harmonics of each radical candidate frequency has tremendous confidence that the sum is the largest. That is, it is the most likely extreme frequency. The confidence coefficient is shown in

$$L(N) = \frac{f_p}{N}, \quad 1 \leq N \leq 5, \quad (23)$$

$$R(N) = \sum_{i=1}^N P(i). \quad (24)$$

In Equations (23) and (24), $L(N)$ is radical candidate frequency, f_p is maximum peak frequency, $R(N)$ is the confidence level, $P(i)$ is the amplitude of a particular harmonic, and n is the number of harmonics.

The confidence coefficient solves the problem that the maximum peak amplitude component is harmonics to a certain extent. When dealing with low-frequency and high-frequency sound waves, there is still a problem of the high probability of misjudgment, that is, low accuracy. The study proposes an improved standard harmonic way to extract the radial frequency using the above harmonic peak method and confidence coefficient issues. Using discrete first and second derivatives, find the first n maximum points x_1, x_2, x_{n-1}, x_n with higher spectrogram peaks as candidate fundamental frequencies. Construct the confidence function $h(x_i)$ to reflect the possibility that x_i is the radical frequency:

$$h(x_i) = \sum_{k=1}^n g(x_k) t(x_k, x_i) \quad (i = 1, 2, 3, \dots, n). \quad (25)$$

In Equation (25), $g(x_k)$ is the energy value corresponding to the candidate radical frequency x_k and $t(x_k, x_i)$ is the closeness of x_k to an integer multiple of x_i . It is defined as

$$t(x_k, x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(\theta_k)^2/2\sigma^2} \quad (k = 1, 2, 3, \dots, n). \quad (26)$$

In Equation (26), σ is the scale parameter and θ_k is

defined as follows:

$$\theta_k = \begin{cases} \min\left(\frac{x_k}{x_i} - \left\lfloor \frac{x_k}{x_i} \right\rfloor, 1 + \left\lfloor \frac{x_k}{x_i} \right\rfloor - \frac{x_k}{x_i}\right), & x_k \geq x_i, \\ 0.5, & x_k < x_i. \end{cases} \quad (27)$$

Given the candidate radical frequency interval $[f_{\min}, f_{\max}]$, the number of candidate fundamental frequencies n , and the scale parameter, firstly, the frequency spectrum function of the input audio is calculated to determine the radical candidate frequency. Secondly, calculate the probability that each candidate's revolutionary frequency is extreme. Finally, the value of x_i that maximizes $h(x_i)$ is used as the radical frequency value of the signal. The calculation process of the improved standard harmonic method is shown in Figure 3.

- (1) After performing a Fast Fourier Transform on A , take the modulus and limit its frequency range to $[f_{\min}, f_{\max}]$. The frequency spectrum function equation is as shown in

$$g(x) = |FFT(A)|(x \in [f_{\min}, f_{\max}]) \quad (28)$$

- (2) Calculate the first n maximum points x_1, x_2, x_{n-1}, x_n of $g(x)$ and use them as the radical candidate frequency
- (3) Calculate the confidence function $h(x_i)$ of each radical candidate frequency
- (4) $x = \max(h(x_i))$

2.4. Construction of Piano Performance Evaluation Model by BPNN

2.4.1. Music Performance by MIDI Standards. MIDI is a standard for electronic music. When a player uses a MIDI instrument to play, the MIDI instrument will convert the player's operations into MIDI signals and then pass these MIDI signals to the sequencer. A sequencer is a device that organizes, edits, and outputs the timbre, rhythm, notes, etc., required by a piece of music to the sound source for sound production. The stored MIDI signal is a MIDI file [25]. After obtaining the MIDI signal, the characteristics of

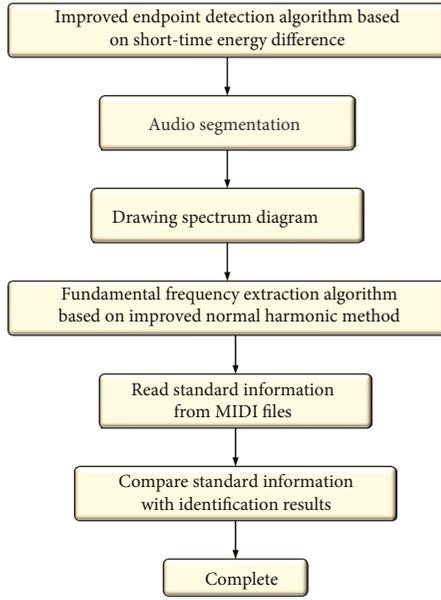


FIGURE 4: Flow chart of error correction for piano playing music by MIDI files.

the sound can be analyzed, such as time value and pitch. In a MIDI system composed of a personal computer, the MIDI input device of various MIDI musical instruments is directly connected to the computer sound card. A sequencer is computer software that acquires MIDI signals obtained mainly through the computer operating system. The Application Programming Interface (API) is obtained from the computer sound card. The MIDI messages received by the computer system are composed of multiple bytes [26].

2.4.2. Error Correction of Piano Playing Music by MIDI Files. After obtaining a piece of piano playing audio, first, get each note's start and endpoints through the improved endpoint detection algorithm by the short-term energy difference mentioned above. At this time, the time value information of the note is also determined. The report divides the original signal to obtain all the letters in the audio to obtain the movement. Second, through the radical frequency extraction algorithm by the improved standard harmonic method, identify the pitch information of each note. Finally, the note time value and pitch information obtained by the two steps are compared with the corresponding standard information in MIDI music to find out the wrong note played by the player. The error correction process of piano playing sound by MIDI file is shown in Figure 4.

2.4.3. Confirmation of Input Parameters

(1) Determine the Parameters of the Pitch Feature. Johann Sebastian Bach's *minuet* is an example of implementing a piano performance evaluation model. There are five notes in the first bar of this piano music. The high notes are D, G, A, B, and C. MIDI messages are 74, 77, 70, 71, and 72. The input parameter is $5/5 = 1$ when playing five tones, $4/5 = 0.8$ when playing four styles, and so on. The total score

has 48 bars, so the input parameters of the high pitch feature require 48 neurons in the input layer.

(2) Determine the Parameters of Rhythm Characteristics. Rhythm is a feature that describes the length of the sound, and the time represented by the score is relative. It needs to be converted to absolute time to get the difference between the length of the sound [27]. Take the first measure of *minuet* as an example. The first note is a quarter note with a duration of 1, and the last four notes are all eighth notes with a period of 0.5. Suppose each note pronunciation's total time is 0 s and 0.75 s, and the duration is 0.75 s and 0.25 s, respectively. If the sounding time of the note is 0.1 s and 0.8 s and the course is 0.7 s and 0.3 s, respectively, the player's sense of rhythm can be calculated by

$$f(x) = \sum \text{abs}(S_i - B_i)Q_i + \text{abs}(E_i - D_i)Q_i. \quad (29)$$

In Equation (29), S_i is the sounding time or duration of quarter notes when playing, B_i is the absolute time or the hypothetical quarter note pronunciation, E_i is the pronunciation time or duration of the eighth note when playing, D_i is the absolute time or duration of the hypothetical eighth note pronunciation, and Q_i is weight.

According to Equation (29), the player's sense of rhythm is shown in

$$\begin{aligned} \text{abs}(0.1 - 0) + \text{abs}(0.8 - 0.75) &= 0.15, \\ \text{abs}(0.7 - 0.75) + \text{abs}(0.3 - 0.25) &= 0.1. \end{aligned} \quad (30)$$

Since the duration has little influence on the performance effect, multiplying by the weight of 0.1, the rhythm input parameter is $0.15 + 0.1 * 0.1 = 0.16$. The input parameters of the rhythm also require 48 neurons in the input layer.

(3) Determine the Parameters of the Beat Feature. The beat is a feature that describes the strength of the sound. The score has a minimal representation of the power of the sound, and the term of the force is a relative value [28]. Artificially judge the average value of input samples with excellent beat control as the MIDI standard signal. The beats of the second bar of the piano score of *minuet* are strong, weak, and weak. Let the traditional values of the strength of each note be 100, 70, and 80, respectively, and the values obtained by the performer are 95, 75, and 70, respectively. The beat input parameter in this section is $5 + 5 + 10 = 20$. The input parameters of the beat require 48 neurons in the input layer.

(4) Determine the Parameters of the Chord Characteristics. Chords are composed of simultaneous pronunciation at a time point [29]. The time point of each chord can be obtained in advance from the music score and standard MIDI files. The chord judgment is performed on each chord time point during the playing process. The difference in pitch is used to judge whether the fundamental tone is right or wrong. The number of chord judgment errors is 1, and if

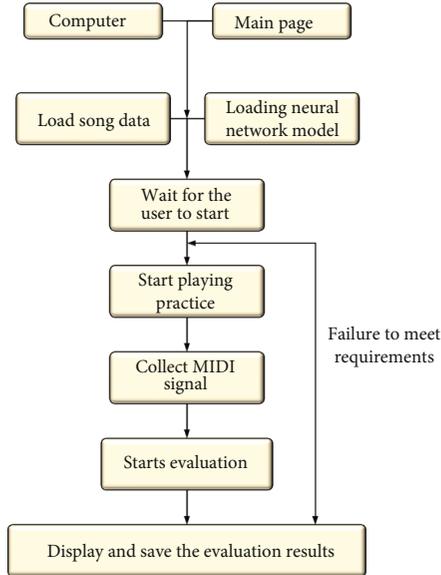


FIGURE 5: Architecture diagram of piano performance evaluation model based on BPNN.

the other one is wrong, it is considered 0.5 when playing a wrong chord. Finally, divide the total number of chord judgment errors by the total number of chords as the input parameter of the chord pitch, which corresponds to an input layer neuron. The method of judging the strength and timing of the chord is the same as the pitch, and the input parameters correspond to two input layer neurons.

(5) *Determine the Parameters of Melody Characteristics.* The melody is the artificial grouping of music [30]; the minuet can be divided into two pieces, and the first melody can be divided into four phrases. The second melody is the reproduction part. The characteristic song parameters can be obtained by summing the pitch, time value, and strength attributes of each segment's notes. The typical melody parameters of *minuet* should correspond to the neurons of the six input layers.

2.4.4. *Determination of the BPNN Structure.* The input parameters determined above correspond to a total of 153 neurons in the neural network's input layer. The total evaluation of the player's performance requires an output neuron to conform to it. Two indicators of rhythm and expressiveness are also commonly used for performance evaluation. Two neurons are used to correspond to them, and the output layer of the neural network is determined to have three neurons. Therefore, as long as the number of hidden layer neurons is determined, the number of hidden layer nodes must be greater than the sum of the input layer and the output layer nodes, according to experience. Then, the number of remote layer nodes should be increased until the performance requirements reach the ideal state. For the neural network model of the *minuet*, the number of hidden layer nodes finally determined after repeated experiments are 168.

The structure of the piano performance evaluation model by BPNN designed is shown in Figure 5.

2.5. Experimental Analysis Method

2.5.1. *Training of BPNN Model.* The Mean Square Error (MSE) is used in the neural network training process, and its definition is shown in

$$MSE = \frac{1}{mp} \sum_{p=1}^p \sum_{j=1}^m (\widehat{y}_{pj} - y_{pj}). \quad (31)$$

In Equation (31), m is the number of output nodes, p is the number of training samples, \widehat{y}_{pj} is the expected output value of the network, and y_{pj} is the actual output value of the network.

After determining the input parameters that affect the performance effect and the parameters and structure of BPNN, the neural network needs to be trained to achieve the desired accuracy requirements. First, the standard information of each characteristic is obtained by the piano teacher's playing and MIDI files. Then, features are extracted and input through different levels of piano performance. The training method of the neural network obtains training samples mainly through two piano teachers and three students playing the *minuet* piano several times. The input data of the neural network are received, and then, the overall performance, rhythm, and expressiveness of the commission are evaluated manually. The data input ranges between 0 and 1, and 10 samples are collected in this training. Set the learning rate to 0.8, the momentum factor to 0.5, and the error to 0.01 to start neural network training.

2.5.2. Validation of Piano Performance System Model by BPNN

(1) *Model Performance Test.* Use the *minuet* piano music to test the performance of the model. Unlike the neural network training sample collection, the model's performance is tested by another piano teacher. Student A (level 6 piano) and student B (level 5 piano) play ten times as input samples. The BPNN piano performance system model is used to evaluate the three people's performance overall rhythm and expressiveness.

(2) *Error Correction Accuracy Rate Test.* After recording the model performance test sample, let the piano teacher play *minuet* 10 times as the model error correction of the accuracy test sample. Among them, all 5 groups played correctly; 2 groups deliberately misplayed several pitches, and the timing values were all correct; 2 groups deliberately misplayed several timing values, and the angles were all correct; 1 group deliberately misplayed several pitches. The total score of *minuet* piano music has 162 notes, played 30 times in total, and 4860 note samples. In the test process, the traditional BPNN is used as a control, and the results are compared and analyzed of the optimization algorithm output. This further verifies the error correction performance of the optimized BPNN in piano performance.

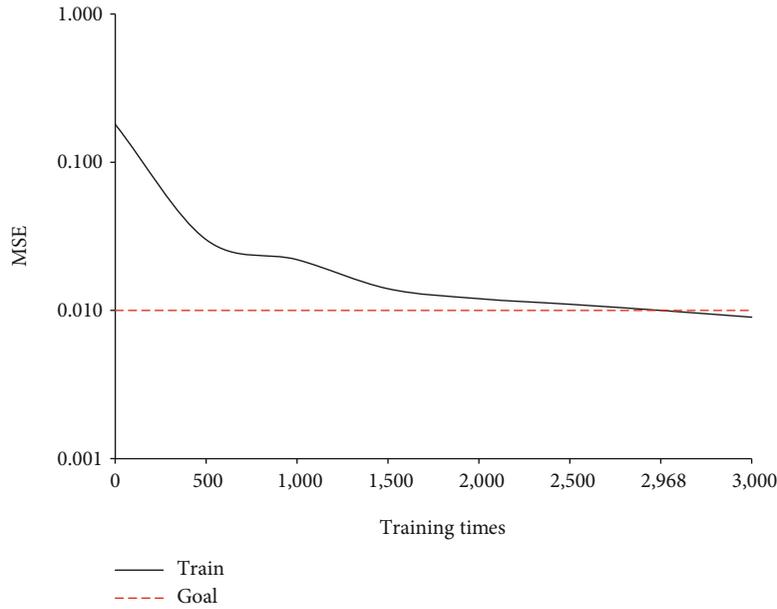


FIGURE 6: Analysis of neural network training results.

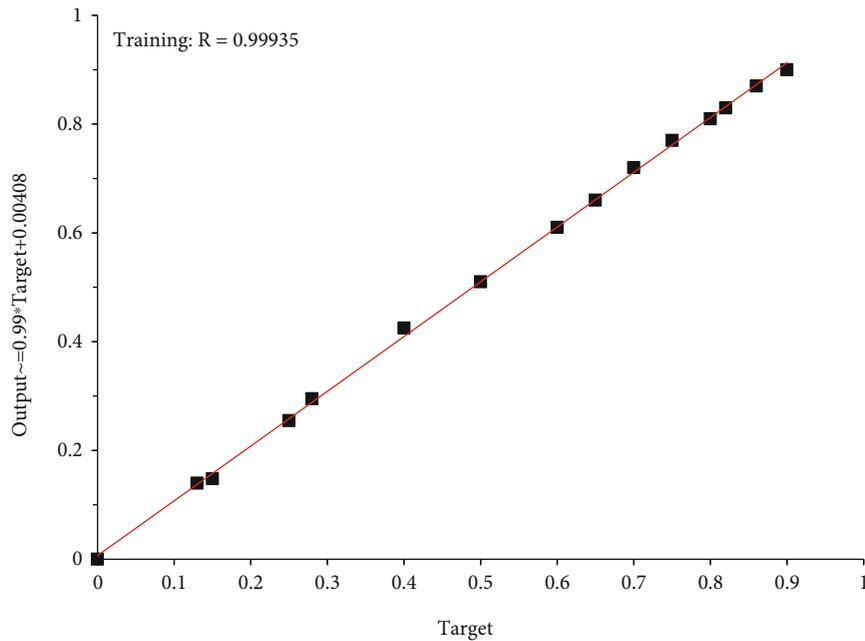


FIGURE 7: Neural network training error curve.

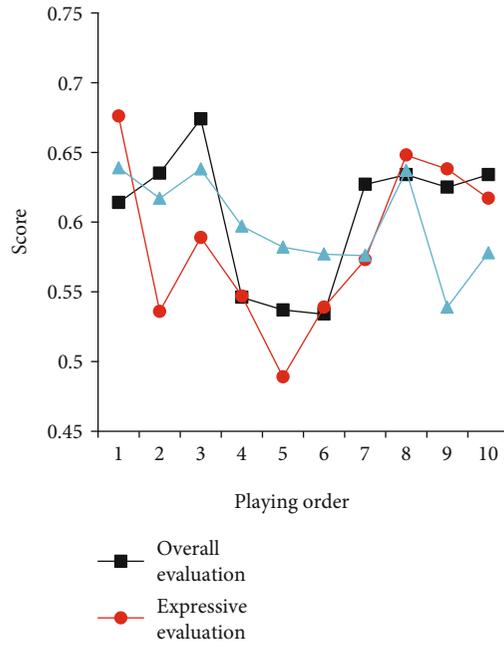
3. Results

3.1. *BPNN Model Training Results.* The training results of BPNN designed are shown in Figures 6 and 7.

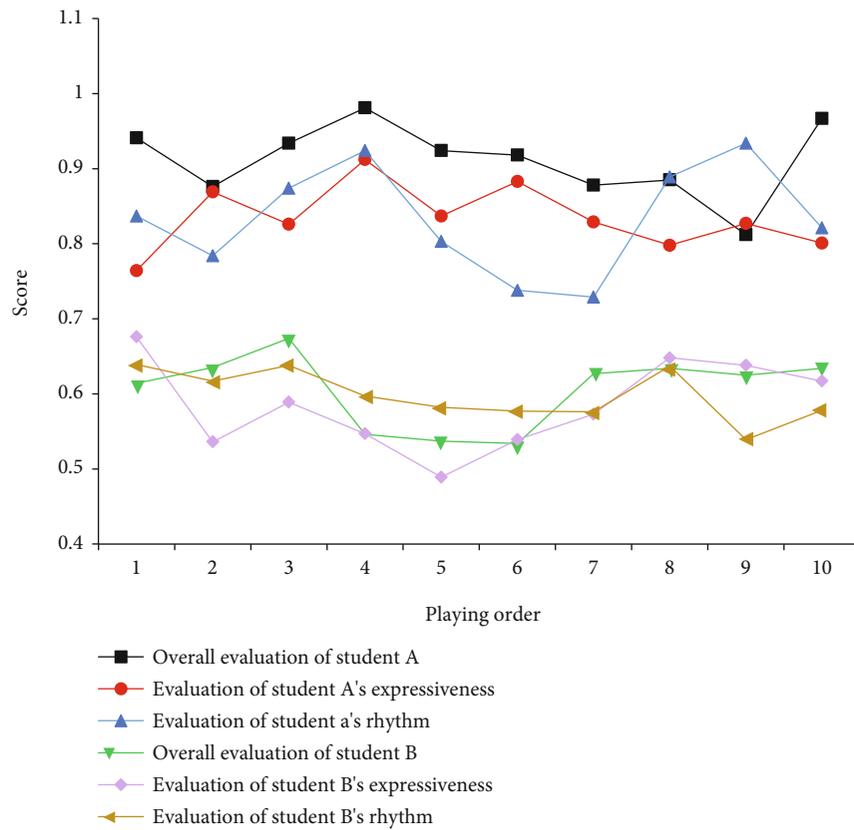
Figures 6 and 7 show that during the training process of BPNN designed, the error value shows a downward trend as the number of training increases. After 3000 times, the error is less than the preset 0.01, and the network convergence accuracy is achieved. The correlation coefficient between the network output and the target is as high as 0.99935, and the degree of fit is high. Therefore, the performance of BPNN designed can meet the actual requirements.

3.2. *Model Performance Test Results.* The evaluation results of a piano teacher, student A (level 6 piano), and student B (level 5 piano) using the model designed are shown in Figure 8.

Figure 8 shows that the overall evaluation average of piano teachers is 0.9116, the average expressiveness evaluation is 0.8346, and the rhythmic evaluation average is 0.8333; the evaluation averages of student A are 0.7169, 0.6516, and 0.6694, respectively; the standards of student B are 0.606, 0.5852, and 0.5980, respectively. Figure 9 reveals that the evaluation values given by the model from high to low are piano teacher, student A, and student B, which are

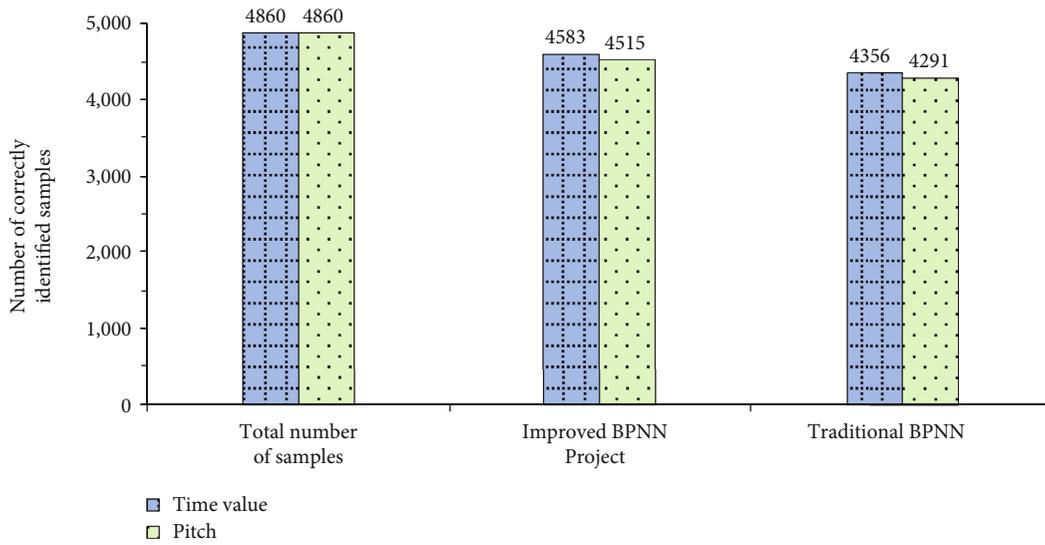


(a)

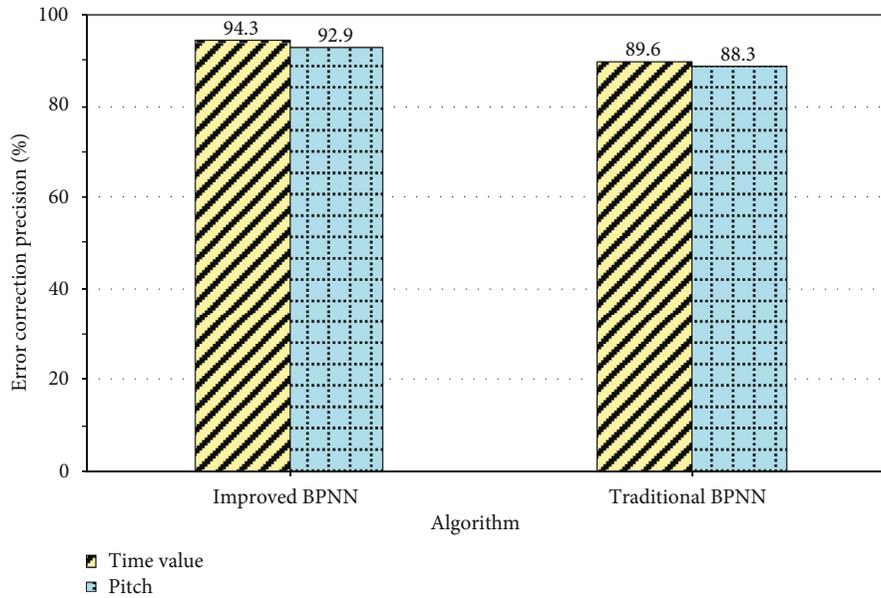


(b)

FIGURE 8: Performance evaluation result graph: (a) piano teacher's performance evaluation result graph; (b) student A&B's performance evaluation result graph.



(a)



(b)

FIGURE 9: Test results of the error rate of the evaluation model.

generally in line with the actual level of the performer. It is proved that the model’s output can meet the requirements and be used in piano music teaching.

3.3. The Accuracy Test Result of Error Correction. The error correction rate result of the piano performance evaluation model by BPNN is shown in Figure 9.

Figure 9 collected a total of 4860 samples. The number of correct correction samples of the traditional BPNN is 4356. The accuracy rate is 89.6%, the number of accurate correction samples of pitch is 4291, and the accuracy rate is 88.3%. The number of correctly corrected pieces of the BPNN optimized is 4583, and the accuracy rate is 94.3%, which is 5.25% higher than the traditional method. The number of correct pitch correction samples is 4515, and

the accuracy rate is 92.9%, which is 5.21% higher than the conventional method. The optimized BPNN has significantly improved the error correction accuracy of the notes and pitches played by the player. The model can effectively help piano beginners correct errors and improve the accuracy and efficiency of the practice.

4. Conclusions

In traditional piano music teaching, teachers play an essential role. However, due to various reasons, the shortage of piano teachers has caused piano learners to practice by themselves most of the time. Without the guidance of professional teachers, it is difficult for beginners to find the wrong playing in their practice, and the learning efficiency

is low. Therefore, it is essential to establish a piano performance evaluation system with an error correction function. Firstly, the Levenberg Marquardt (LM) algorithm is used to optimize the BPNN of the traditional music teaching evaluation system. Secondly, an improved endpoint detection algorithm is proposed based on short-term energy difference, which can accurately identify the time value of each note in the piano playing audio. By the traditional frequency domain analysis method, a radical frequency extraction algorithm with an improved standard harmonic method is proposed, which can accurately identify the note's pitch. Finally, a piano performance evaluation model by BPNN is implemented. The experimental results show that after 3000 times, the neural network error is less than 0.01, and the network converges. The evaluation result of the piano performance evaluation model designed is basically in line with the actual level of the performer and has specific feasibility. The error correction accuracy of the optimized BPNN is 94.3% during the performance, which is 5.25% higher than the traditional method. The pitch error correction rate is 92.9%, 5.21% higher than the conventional method. The optimized BPNN has significantly improved the error correction accuracy of the time value and pitch of the notes played by the player. The error correction of the designed model piano is helpful for beginners. It can improve the accuracy and efficiency of exercises. The shortcoming is that all the experimental results are only for the song *minuet*, which does not indicate that the evaluation model designed is also applicable to other songs. Therefore, it needs to expand the experimental sample for further verification. The purpose of this study is to alleviate the scarcity of piano teachers, reduce the work intensity of piano teachers, realize automatic error correction and objective evaluation of playing, and provide necessary technical support for improving the efficiency of piano music teaching.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. R. Griffith, A. L. Ramos, K. E. Hill, and C. F. Miguel, "Using equivalence-based instruction to teach piano skills to college students," *Journal of Applied Behavior Analysis*, vol. 51, no. 2, pp. 207–219, 2018.
- [2] L. Li, "Study on the innovation of piano teaching in normal colleges and universities," *Creative Education*, vol. 9, no. 5, pp. 697–701, 2018.
- [3] J. Wang, "The application of intelligent speech recognition technology in the tone correction of college piano teaching," *Journal of Physics: Conference Series*, vol. 1852, no. 2, article 022086, 2021.
- [4] L. Yan, "Design of piano teaching system based on internet of things technology," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 5, pp. 5905–5913, 2019.
- [5] Q. Zhang, H. Yu, M. Barbiero, B. Wang, and M. Gu, "Artificial neural networks enabled by nanophotonics," *Light: Science & Applications*, vol. 8, no. 1, pp. 1–14, 2019.
- [6] J. Jiang, G. Hu, X. Li, X. Xu, P. Zheng, and J. Stringer, "Analysis and prediction of printable bridge length in fused deposition modelling based on back propagation neural network," *Virtual and Physical Prototyping*, vol. 14, no. 3, pp. 253–266, 2019.
- [7] H. Fei, "Application research of neural network technology in vocal music evaluation," in *2021 6th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 176–179, Kunming, China, 2021.
- [8] W. Y. Jia, "Research on the application of BPNN in vocal music teaching quality evaluation," in *2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*, pp. 306–309, Zhangjiajie, China, 2020.
- [9] A. Dimas, "Emotion classification by EEG signal generated by brain using discrete wavelet transform and artificial neural network backpropagation with classical music stimulus," *International Journal of Engineering Technology and Natural Sciences*, vol. 1, no. 2, pp. 1–4, 2020.
- [10] X. Li and H. Robin, "Construction and analysis of hidden Markov model for piano notes recognition algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 3, pp. 3293–3302, 2019.
- [11] L. N. Li, X. F. Liu, F. Yang, W. M. Xu, J. Y. Wang, and R. Shu, "A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis," *Spectrochimica Acta Part B Atomic Spectroscopy*, vol. 180, no. 1, 2021.
- [12] B. Wang, X. Gu, L. Ma, and S. Yan, "Temperature error correction based on BP neural network in meteorological wireless sensor network," *International Journal of Sensor Networks*, vol. 23, no. 4, pp. 265–278, 2017.
- [13] L. Xu, J. Wang, H. Wang, T. Aaron Gulliver, and K. N. le, "BP neural network-based ABEP performance prediction for mobile Internet of Things communication systems," *Neural Computing and Applications*, vol. 32, no. 20, pp. 16025–16041, 2020.
- [14] C. Yuan, X. Sun, and Q. J. Wu, "Difference co-occurrence matrix using BP neural network for fingerprint liveness detection," *Soft Computing*, vol. 23, no. 13, pp. 5157–5169, 2019.
- [15] J. Zou, K. T. Bui, Y. Xiao, and C. V. Doan, "Dam deformation analysis based on BPNN merging models," *Geo-spatial information science*, vol. 21, no. 2, pp. 149–157, 2018.
- [16] Z. Wang, B. Wang, C. Liu, and W. S. Wang, "Improved BP neural network algorithm to wind power forecast," *The Journal of Engineering*, vol. 2017, no. 13, pp. 940–943, 2017.
- [17] Z. Zhao, Q. Xu, and M. Jia, "Improved shuffled frog leaping algorithm-based BP neural network and its application in bearing early fault diagnosis," *Neural Computing and Applications*, vol. 27, no. 2, pp. 375–385, 2016.
- [18] Y. Lv, W. Liu, Z. Wang, and Z. Zhang, "WSN localization technology based on hybrid GA-PSO-BP algorithm for indoor three-dimensional space," *Wireless Personal Communications*, vol. 114, no. 1, pp. 167–184, 2020.
- [19] Y. Zhang, J. Tang, R. P. Liao et al., "Application of an enhanced BP neural network model with water cycle algorithm on

- landslide prediction,” *Stochastic Environmental Research and Risk Assessment*, vol. 35, no. 6, pp. 1273–1291, 2021.
- [20] L. Wang, S. Hu, K. Liu et al., “A hybrid genetic algorithm and Levenberg–Marquardt (GA–LM) method for cell suspension measurement with electrical impedance spectroscopy,” *Review of Scientific Instruments*, vol. 91, no. 12, article 124104, 2020.
- [21] J. Lu and X. Han, “Novel speech endpoint detection algorithm for voice detectors in interaction of intelligent terminals,” *Sensors & Transducers*, vol. 242, no. 3, pp. 1–5, 2020.
- [22] M. Ying, L. Kaiyong, H. Jiayu, and G. Zangjia, “Analysis of Tibetan folk music style based on audio signal processing,” *Journal of Electrical and Electronic Engineering*, vol. 7, no. 6, p. 151, 2019.
- [23] T. C. Yeh and J. S. R. Jang, “AutoRhythm: a music game with automatic hit-timing generation and percussion identification,” *IEEE Transactions on Games*, vol. 12, no. 3, pp. 291–301, 2020.
- [24] L. Zimman, “Gender as stylistic bricolage: transmasculine voices and the relationship between fundamental frequency and/s,” *Language in Society*, vol. 46, no. 3, pp. 339–370, 2017.
- [25] K. János, “Midicine: using MIDI (musical Instrument Digital Interface) in medicine for treating patients who experienced a stroke-review article,” *Mentálhigiéné és Pszichoszomatika*, vol. 18, no. 4, pp. 365–379, 2017.
- [26] D. M. Howard, “The vocal tract organ: a new musical instrument using 3-D printed vocal tracts,” *Journal of Voice*, vol. 32, no. 6, pp. 660–667, 2018.
- [27] A. F. Yap, Y. H. Kwan, C. S. Tan, S. Ibrahim, and S. B. Ang, “Rhythm-centred music making in community living elderly: a randomized pilot study,” *BMC Complementary and Alternative Medicine*, vol. 17, no. 1, pp. 1–8, 2017.
- [28] J. London, R. Polak, and N. Jacoby, “Rhythm histograms and musical meter: a corpus study of Malian percussion music,” *Psychonomic Bulletin & Review*, vol. 24, no. 2, pp. 474–480, 2017.
- [29] H. P. Utomo, “MIDI command formula design for chord based music applications,” *Jurnal TIARSIE*, vol. 15, no. 1, pp. 1–4, 2018.
- [30] H. Zhu, Q. Liu, N. J. Yuan, K. Zhang, G. Zhou, and E. Chen, “Pop music generation,” *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 14, no. 5, pp. 1–31, 2020.