Construction of an Art Education Teaching System Assisted by Artificial Intelligence

Xianyu Wang[1]*, Xiaoguang Sun[2]
School of Fashion, Henan University of Engineering, Zhengzhou, 451191, China

***Abstract*—With the continuous progress of art education and artificial intelligence technology, traditional music teaching models are facing transformation. This article aims to construct an art education and teaching system based on artificial intelligence, especially for teaching music sound recognition. Through in-depth research, we have designed a music sound recognition system that uses Mel frequency cepstral coefficient (MFCC) for feature parameter extraction, and combines BP neural network algorithm to construct a music sound learning model. The main purpose is to improve the efficiency and accuracy of music teaching through artificial intelligence technology. The main challenge we face in this process is how to effectively extract the features of music sounds and accurately identify different tones through algorithms. By using the MFCC algorithm, we have successfully solved this problem as it can effectively describe the time-frequency characteristics of music sound. Our proposed music sound learning model is based on a BP neural network, which trains the network to learn the mapping relationship between music sound and pitch. The experiment used piano sound as an example to verify the accuracy and reliability of the system. The simulation experiments conducted in MATLAB environment show that our system can accurately recognize and extract the main frequency of music, and has higher performance compared to traditional methods.**

Keywords—Feature extraction BP neural network; Tone recognition; smart art teaching; MEL frequency cepstral coefficient; MFCC algorithm; time frequency characteristics

I. INTRODUCTION

The Internet era's art education is progressively moving toward online and intelligent learning due to the economy's rapid growth. As a new discipline, artificial intelligence is mainly used to expand the methods and theories of human intelligence. The advent of artificial intelligence brings many opportunities for the development of science and technology but also makes art design face more challenges. The changes in market demand, the prominence of design immediacy problems, and the changes in design subjects make the existing art design face more challenges [1], [2].

Different listeners have their preferences for musical styles, and people usually experience and appreciate music from the tune, tone, and rhythm. In music dissemination, the music signal is transmitted to the human ear first; after receiving it, the human ear converts the music waveform into a bioelectric signal and then transmits it. Bioelectrical signals are transmitted to the brain through the nervous system. Finally, the brain analyzes the bioelectrical signals to further learn from the human experience of music. The brain plays a major role in the music appreciation process. In general, the role of the brain is to reduce high-dimensional musical signals to several musical dimensions that people care about, such as rhythm, melody, and harmonic intensity and intensity. These elements describe our experience and understanding of music. Therefore, learning a music data model based on music feature parameters can effectively reduce redundant data in music data and finally obtain more accurate data describing music. Tone recognition is a pattern recognition problem. There are rule-based audio classification methods, pattern matching methods, etc., but these methods still need to be improved. The rule-based audio classification method is simple to operate. Still, because of its simplicity, it is only suitable for identifying audio types with simple features, such as mute, which is difficult to satisfy for complex and multi-feature music classification applications. The pattern-matching method must establish a standard pattern for each audio type and then compare and match the input pattern with the standard pattern, which requires much computation and low classification accuracy. The hidden Markov model (HMM) method has a strong dynamic time series modeling ability, and the calculation amount is small, but the classification decision-making ability could be better. Artificial intelligence provides a good foundation for the research work of self-learning ability and automatic audio classification [3].

The modern art education system has gradually begun to be refined, including art education work evaluation, art education curriculum evaluation, and art education timeliness evaluation. In the quality evaluation of art theory courses, as early as the 1990s, some scholars in our country discussed the art education teaching system concept. Later, some domestic scholars explored and spread it, but it has yet to form a large-scale impact. The early research scholars analyzed how to improve the teaching quality of art courses from a macro level. They believed ideological teaching should integrate theory with practice, based on classroom teaching, enrich students' theoretical knowledge reserves, and strengthen display and application. Lin D. et al. [4] first proposed an intelligent system for teaching art theory courses in colleges and universities and tried to introduce artificial intelligence algorithms into art education. By exploring the application of artificial intelligence algorithms in art teaching, this paper proposes a music recognition method based on BP neural network for the problems existing in traditional music recognition methods. The cepstral coefficient method is used to extract the musical features, and the Mel cepstral coefficients are used to form the feature vector.

This article is dedicated to exploring the innovative application of artificial intelligence in the field of music sound recognition, providing intelligent and personalized solutions for art teaching. We use Mel frequency cepstral coefficient (MFCC) for feature parameter extraction and combine it with BP neural network algorithm to construct a music sound learning model to optimize the effectiveness of music teaching. Through the research in this article, we hope to bring revolutionary changes to the field of music education, achieve more efficient and precise teaching, and stimulate students' creativity and musical potential. This is not only an in-depth exploration of the application of artificial intelligence in the field of art education and teaching, but also a beneficial attempt to innovate music education and teaching methods.

II. Related Work

With the advancement of technology and the development of media art, artificial intelligence technology has penetrated into various art fields, especially in the application of complex situational teaching in folk music, which is becoming increasingly prominent. Li and Bin [5] discussed how artificial intelligence can assist in the teaching of complex folk music situations and promote the modernization of traditional folk music education from the perspective of media art. Media art, as a new field that integrates technology and art, has brought a new perspective to traditional folk music teaching. In this context, folk music teaching is no longer limited to traditional master apprentice inheritance or classroom teaching, but can realize the sharing of teaching resources and diversification of teaching methods through multimedia, Internet and other modern technical means. Complex situation teaching emphasizes teaching in actual or simulated real situations to improve students' practical abilities and ability to cope with complex situations. In the teaching of folk music, complex situational teaching is particularly important because folk music is often closely linked to specific cultural and historical backgrounds. The introduction of artificial intelligence technology has provided strong support for the teaching of complex situations in folk music. Computer assisted analysis, as a technical means, can quantitatively and qualitatively analyze artistic images. By utilizing algorithms such as image recognition, machine learning, and deep learning, CAA can deeply explore the style, techniques, themes, and other aspects of artistic works, providing insights that traditional methods find difficult to obtain. Shen [6] collected a large amount of image data from art works and analyzed them using CAA to discover the stylistic differences and evolutionary trends among different art genres and artists, thus constructing a more comprehensive and detailed theory of art history. CAA reveals the decision-making process, technological application, and style formation of artists in the creative process, which helps us understand the internal logic and laws of artistic creation and provides new theoretical support for artistic creation. Neural networks are computational models that simulate the structure of human brain neurons, with powerful feature learning and classification capabilities. In vocal speech recognition, neural networks can automatically extract sound features by learning a large amount of vocal data, achieving accurate recognition of vocal signals.

The vocal speech recognition technology based on neural networks can not only recognize basic elements such as melody and rhythm of songs, but also further analyze deeper information such as the singer's timbre, pitch accuracy, and emotions. Vocal voice evaluation is a quantitative evaluation of a singer's performance. Traditional vocal evaluation mainly relies on the subjective judgment of professional judges, while neural network-based vocal speech evaluation technology can provide more accurate and comprehensive evaluation through objective analysis of vocal signals. Neural networks can learn the evaluation criteria and preferences of professional judges, thereby achieving automatic evaluation of the singer's voice quality, skill application, emotional expression, and other aspects [7]. Mei et al. [8] explore the possibility of constructing art theory from the perspective of art images based on AI assisted analysis, and analyze its impact on art research and creation. AI assisted analysis refers to the automated and intelligent analysis of artistic images using artificial intelligence algorithms [9]. Through technologies such as deep learning and image recognition, AI can deeply explore the style, techniques, composition, and other aspects of artistic works, providing a more detailed and objective interpretation than traditional methods. This technology not only improves the accuracy and efficiency of analysis, but also provides a new perspective and method for art research. Traditional research on art history mainly relies on the knowledge and experience of experts, while AI assisted analysis can reveal the evolution laws of art styles, schools, and trends through mining and analyzing a large amount of art image data, providing a more comprehensive and objective theoretical basis for art history. AI can analyze the image data during the artist's creative process to reveal the artist's creative ideas, technical application, and style formation process. This helps us to have a deeper understanding of the internal logic and laws of artistic creation, providing new support for the theory of artistic creation [10].

Although artificial intelligence has made significant progress in the field of music recognition, it still faces many challenges and limitations. Traditional voice recognition methods often rely on manually extracted features, which are time-consuming and difficult to ensure the comprehensiveness and effectiveness of the features. In addition, musical sound has a high degree of complexity and diversity, and different instruments and performance styles can bring difficulties to sound recognition. Therefore, it is particularly important to develop a system that can adaptively, efficiently, and accurately recognize music sounds.

This article proposes a music sound learning model based on MFCC and BP neural network. Firstly, the MFCC algorithm is used to extract feature parameters of music sound, which can effectively describe the time-frequency characteristics of music sound. Then, these feature parameters are used as inputs for the BP neural network, and the mapping relationship between music sound and pitch is learned through training the network. Finally, the trained network is used to recognize new music sounds, thereby achieving the recognition and extraction of the main frequency of music.

III. Artistic Music Tone Recognition and Feature Extraction

A. Principles of Art Music Tone Recognition

There are two main recording formats for artistic musical tone information: WAVE format and Musical Instrument Digital Interface (MIDI) format. MIDI is a standard for transmitting music signals between electronic musical instruments. It includes hardware interface standards and asynchronous serial transmission protocols for electronic music signals between different hardware. The transfer rate is 31.25k (±1%) baud. Since the music file in MIDI format records the whole process of the score and performance, many basic features of music can be directly extracted. Therefore, the identification of music features mostly adopts the music files in MIDI format, and this paper mainly aims at identifying the music files in MIDI format.

From the physical level analysis, music consists of the fundamental frequency, frequency multiplication, and signal amplitude. These three factors correspond to three important concepts in musicology. The physical quantity determined by the fundamental frequency of the musical tone signal is called pitch. The fundamental frequency and the multiplied frequency signal jointly constitute the requirements for the timbre. Finally, the amplitude of the signal is the sound intensity in musicology. From the level of emotional characteristics, pitch, length, timbre, fundamental tone, tone name, and rhythm are six vectors that reflect the characteristics of music.

1) Pitch: The higher the pitch of the signal, the higher the dominant frequency, and vice versa. However, the corresponding relationship between pitch and dominant frequency is not a simple linear relationship but an approximate logarithmic relationship.

Changes in pitch can be expressed in tunes. For singers, the pitch mainly reflects the changes in the singer's voice. For music sung by a soprano singer, the frequency of the music signal is higher. For bass singers, the signal frequency is very low during music analysis. In the process of picking up musical tones, it is common to detect the length of time between peaks and the amplitude of the peaks. The smaller the amplitude, the smaller the pitch. The amplitude cannot represent the pitch, and the frequency must be considered comprehensively. In this way, the displayed tones are the high and low frequencies.

Define f as a function that evaluates the pitch attribute of a MIDI file. The musical emotion of a MIDI song file is mainly reflected in the master track. Let the number of audio tracks be m. In the process of identifying the emotion of the music, the relevant extraction of the pitch characteristics of each audio track, such as the average value of the pitch characteristics of each audio track, is carried out, and determine the pitch feature vector x_f of the piece according to the corresponding properties of the main audio track. Therefore, suppose that the type of music emotion to be detected is n, $n \geq 1$, and the eigenvalue of the emotion type under the action of the current pitch value is $pi(xf)$, $i \in [1, 2, ..., n]$, then each emotion the part of the type related to the pitch attribute index of the song, as shown in Formula (1).

$$f(x_f) = Max(p_i) \tag{1}$$

2) Length: To measure the pitch characteristics of a musical piece, a function g is defined as an evaluation standard. Generally speaking, the longer the pitch, the more sad and melancholy elements it expresses in the emotional components. In the pitch identification of MIDI files, the pitch attribute is mainly determined according to the length of its duration. Define the threshold value of the pitch length switch as x_{switch} and the interval of note switch as x_g, then the value of the pitch length evaluation function g is as follows:

$$g(x_g) = \begin{cases} long, x_g < x_{switch} \\ short, x_g > x_{switch} \end{cases} \tag{2}$$

3) Tone: In the MIDI standard, 128 timbres are defined. 128 timbres are divided into k categories, and the function h for extracting the timbre features of the current music is as follows:

$$h_{x_i} = \begin{cases} 1, x_i \in C_1 \\ 2, x_i \in C_2 \\ \dots \\ M, x_i \in C_M \end{cases} \tag{3}$$

4) Fundamental and overtones: The fundamental tone is the main frequency, which is the trigger frequency when the piano strings vibrate over the full length; the overtone, also known as the harmonic frequency, is the vibration frequency at which the strings are triggered in the non-full field. From the physical characteristics of the above analysis, it can be known that the fundamental frequency is the lowest frequency of vibration triggered by the strings. The values of overtones are integer multiples of the fundamental. In the normal operation of the piano, the fundamental tone determines the pitch of the tone, and the fundamental tone and the overtone together determine the timbre of the tone of the signal.

5) Sound name: The so-called sound name is the name attribute of a certain musical sound. Often referred to in music, octaves correspond to octaves in signal processing. That is, the pitch of two single notes differs by one octave, and the main frequency value differs by one time. Divide one octave of the signal into 12 semi-treble pitches: C, #C, D, #D, E, F, #F, G, #G, A, #A, B. There is the tone name of the tone signal or the fundamental tone level. The name of the piano is based on the twelve-tone equal temperament system.

6) Rhythm: Different music has different emotional matting speeds. The basic understanding is that some music is fast and others are slow. For example, Disco music gives us a sense of speed. In contrast, blues music is slower and emotionally sad, and this music can display the number of syllables output per second or minute by digitizing the rhythm. After noise reduction, the computer detects the number of audio signal peaks per minute. The number of peaks per minute can accurately express the rhythm of speech. Then, by extracting the signal envelope and marking the peaks of the envelope contour, the envelope peak value per second of the whole music is calculated and recorded as the peak frequency, which is used to measure the rhythm intensity of the music. Peak frequency means that the music has a strong rhythm. Generally speaking, this music is mainly disco.

B. *Extraction of Musical Tone Features*

As shown in Fig. 1, the identification of musical functions can also be divided into three levels: basic musical functions, complex characteristics, and general musical characteristics.

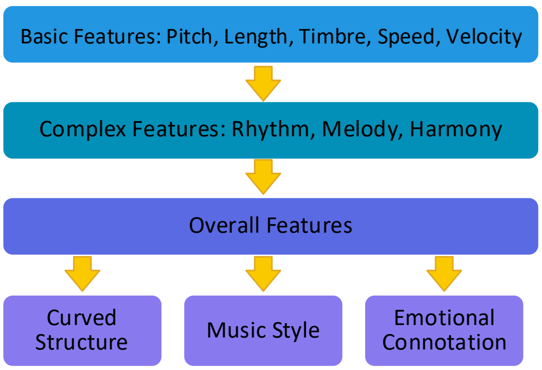

Fig. 1. The characteristic structure of music.

After preprocessing the musical tone signal, the next thing to do is to digitize the musical tone signal and extract its characteristic parameters, that is, replace the entire musical tone signal with a set of characteristic parameters the computer can process. This link is crucial and directly related to the performance of the recognition model in the next link. Feature extraction, or front-end processing, is crucial in recognizing musical tones. Measuring the distance between feature parameters to determine the intrinsic features of musical sounds is the fundamental step in its extraction process. In musical tone recognition, the characteristic parameters generally used are parameters that can reflect the short-term spectral envelope. The mainstream algorithms include Linear Prediction Cepstrum Coefficient (LPCC) and Mel Frequency Cepstral Coefficients (MFCC) f.

When configuring the implemented tone recognition algorithm, a series of parameters need to be set for each algorithm. These parameters will affect the performance and accuracy of the algorithm. The following are several aspects that typically need to be considered when configuring parameters for linear prediction cepstral coefficients (LPCC) and Mel frequency cepstral coefficients (MFCC):

1) Filter order: This refers to the order of the filter or model used for linear predictive analysis. It determines the complexity of the dynamic characteristics of the sound signal that the model can capture. A higher filter order can capture finer structures, but it may also lead to overfitting and increased computational complexity.

2) Window function and window length: The window function and window length used to analyze sound signals determine the time resolution and frequency resolution of the analysis. The commonly used window functions include the Hamming window and the Hanning window. The window length is usually selected based on signal characteristics and the required time-frequency resolution.

3) Pre emphasis: In order to eliminate possible DC bias during the sound production process, pre emphasis is usually performed before signal analysis. Pre emphasis is usually achieved by applying a high pass filter.

4) Iteration count: When solving the LPC coefficient, multiple iterations may be required to converge to a stable solution. The number of iterations should be sufficient to ensure convergence, but excessive iterations may increase computation time.

The linear prediction cepstral parameters (LPCC) map the linear prediction parameters (LPC) in the cepstral domain. Cepstral technology is a well-known homomorphic signal processing method, and its calculation steps generally need to go through the following three steps: fast Fourier transform, logarithmic operation, and phase correction. The LPCC feature extraction algorithm assumes that the musical sound signal is autoregressive and uses the LPC parameters to obtain the cepstral coefficients. The LPCC parameters are insufficient for the anti-noise performance of the system. In the normal range of use, noise signals will inevitably be, which may lead to large errors in the final recognition results. Therefore, MFCC parameters with an anti-noise performance ratio are generally used as a feature extraction method for musical sound signals in practical applications.

5) MFCC: Based on Fourier and cepstral analysis, the spectrum in the time domain can be transformed through the nonlinear spectrum and finally converted to the cepstral spectrum for research. One of its major features is its good identification effect and anti-noise ability, but its calculation amount is complicated compared with LPCC. MFCC reflects the pitch auditory characteristics of the human ear, and it is not computationally intensive and is widely used in speech processing. The research results show that MFCC can be used as an audio classification feature and can improve the accuracy of audio classification. Among them, the relationship between the Mel scale and the specific time domain frequency is shown in Fig. 2.

The calculation process of the MFCC feature is as follows:

The MFCC parameters rely on the Mel frequency filter bank of equal bandwidth set by it, perform D transformation on each frame of signal to calculate the amplitude spectrum, and then transform the amplitude spectrum into the Mel domain with the Mel scale, and filter by the Mel filter bank of equal bandwidth.

$$e[j] = log\left(\sum_{k=0}^{N-1} w_j[k] \times |s[k]|\right), j = 1,2,\ldots.,p \tag{4}$$

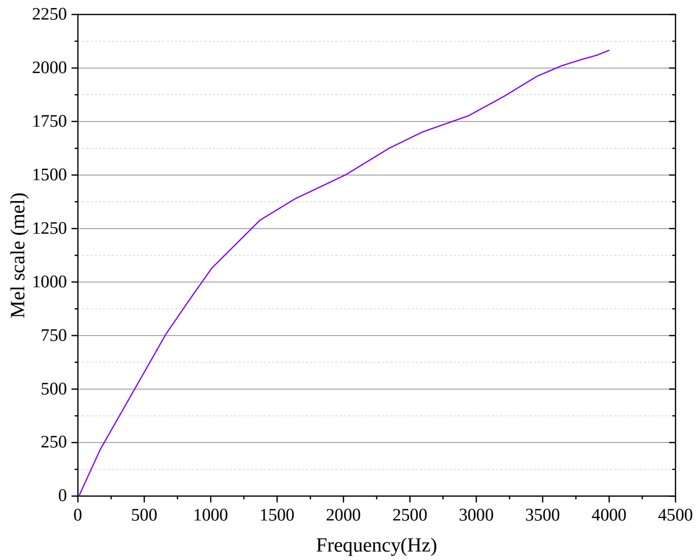

Fig. 2. Relationship between mel scale and frequency in the time domain.

Where: $e[j]$ represents the logarithmic energy output of the jth filter; $w_j[k]$ represents the weight corresponding to the kth point of the jth triangular filter; $|s[k]|$ represents the transformation to Mel scale Discrete Fourier Transform (DFT) spectrum amplitude on; P is the number of filters, generally 24. Taking the discrete cosine transform of the logarithmic energy of the filter, the following cepstral domain MFCC coefficients can be obtained:

$$x_i = \sqrt{\frac{2}{p}} \sum_{j=1}^{p} \left(e[j] \times cos\left(\frac{i\pi}{p}(j-0.5)\right)\right), i = 1,2,\dots,L \quad (5)$$

Among them, L is the dimension of the MFCC. Generally, $L \leq P$, this paper takes 12 dimensions.

The MFCC dynamic characteristic parameters of the music signal can mainly be described by the different spectrums of these static characteristics, combining the first-order and second-order differences as the dynamic feature. These dynamic and static information complement each other, which can greatly improve the recognition performance of the system [11].

C. Research on Tone Recognition Model

After introducing the first two important links of the music sound recognition system, the related algorithms of preprocessing and feature parameter extraction, the system needs to match the model, which is the most important link in the music sound recognition process, matching the optimization degree of model training. The recognition rate of matching using this model is closely connected to it. Consequently, the training of the model is crucial for accurately recognizing musical tones, and users must supply numerous original databases. Therefore, it is necessary to establish a musical tone database before starting the musical tone recognition research. A collection of 88 piano syllables recorded in various settings is available in the database. The individual file should be distinct from others and accurately represent the original musical sound data in a well-balanced way. The essence of model training is to extract the inherent connections and laws of the data using a large amount of data under certain training standards. The unknown musical tone signal is used as input to act on the model, and it is matched and compared to see which template library data is more closely related to obtaining the recognition result. The training process related to the recognition rate of the entire system must meet the training of large data volume and the high efficiency of learning speed.

Artificial neural network (ANN) is a frontier research field for simulating human brain structure and thinking process. ANN can continuously adjust their parameters in training process weights and topology structure to adapt to the environment and the demand of the system performance optimization; it has the characteristics of fast speed and high recognition rate in pattern recognition and has been the research direction and hotspot of music recognition system at home and abroad.

In general, the neural network-based music sound recognition system has great potential for development. Still, the research on the application method for large-scale problems has just started, especially for the actual music sound recognition system. Neural networks generally have the disadvantage of too long training and recognition time.

IV. Realization of Intelligent Art Education System

A. Analysis of BP Neural Network Algorithm

The core algorithm model of the intelligent art teaching system adopts the back-propagation network. The learning algorithm of error backpropagation to train and adjust the weights of the nonlinear differentiable functions of the multi-layer network. It belongs to the multi-layer forward feedback neural network, and the activation function of its neurons belongs to the sigmoid function, which can complete any nonlinear mapping from the input layer to the output layer.

Theoretically, any nonlinear signal can be approximated if the system has one or more S-functions. Therefore, the recognition accuracy can be enhanced through increasing the number of network layers. Still, with the increase of the network layers, the network learning time and calculation amount will also increase significantly, leading to a decrease in the relative performance of the system. Therefore, some balance must be struck between recognition progress and system speed. Generally speaking, the three-layer structure of the BP network, that is, the single-hidden layer network, can achieve a balance between the two by increasing or decreasing the number of neurons in the hidden layer. In the specific design, different numbers of neurons in the hidden layer can be used for experimental comparison to obtain the optimal number of neurons in the hidden layer. Fig. 3 shows the basic structure of the BP neural network, which includes three-layer architecture of input layer, hidden layer, and output layer.

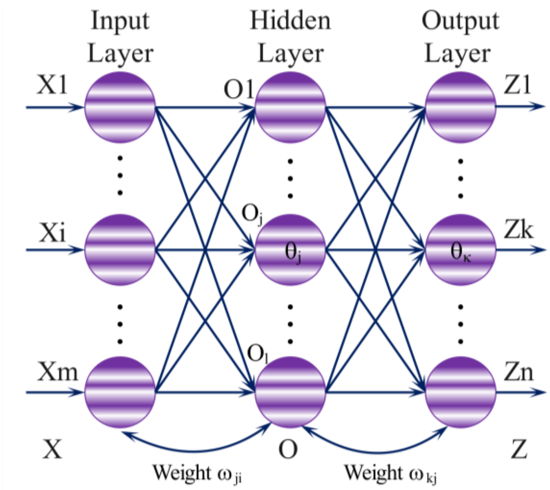

Fig. 3. BP neural network structure.

The first layer is the input layer. For the intelligent art teaching system, the quantities input to the BP neural network can be summarized as the characteristic parameters of the music signal. Set the input sample pair as (X, Y), where the input layer variable X is m, corresponding to n expected values Y.

The second layer is the hidden layer. The number of neurons in the hidden layer O is l. According to the function transfer formula of the hidden layer, the output expression of the jth hidden layer neuron O_j is:

$$O_j = f(\sum_{i=1}^{m} \omega_{ji} x_i - \theta_j), \quad j = 1,2,\cdots,l \tag{6}$$

Among them, ω_{ji} is the weight between the jth hidden layer neuron and the ith input layer neuron, θ_j is the threshold of the jth hidden layer neuron, and f is the hidden layer transfer function [11].

The third layer is the output layer, which outputs the evaluation index of the note signal. The output expression of the kth neuron z_k of the output layer is:

$$z_k = g(\sum_{j=1}^{l} \omega_{kj} O_j - \theta_k), \quad k = 1,2,\cdots,n \tag{7}$$

Among them, ω_{kj} is the weight between the kth output layer neuron and the jth hidden layer neuron, θ_k is the threshold of the kth output layer neuron, and g is the output layer transfer function [12], [13].

It is clear that the error E between the output value z_k and the expected value y_k is:

$$E = \frac{1}{2}\sum_{k=1}^{n}(y_k - z_k)^2$$
$$= \frac{1}{2}\sum_{k=1}^{n}\left\{y_k - g\left[\sum_{j=1}^{l}\omega_{kj} f(\sum_{i=1}^{m}\omega_{ji}x_i - \theta_j) - \theta_k\right]\right\}^2 \tag{8}$$

In the BP neural network, iteration aims to correct the weights and thresholds so that the error function E can decrease at the fastest speed. Taking the negative gradient direction makes the error function decrease the fastest, which is also very consistent with the teaching model of the influencing factors of interactive art learning [14], [15].

It can be seen that to ensure that the predicted value is as close to the expected value as possible, the correction formulas of the weights and thresholds in the iterative process must satisfy the following formulas

$$\begin{cases} \omega_{ji}(t+1) = \omega_{ji}(t) + \Delta\omega_{ji} = \omega_{ji}(t) - \lambda_1 \frac{\partial E}{\partial \omega_{ji}} \\ \omega_{kj}(t+1) = \omega_{kj}(t) + \Delta\omega_{kj} = \omega_{kj}(t) - \lambda_2 \frac{\partial E}{\partial \omega_{kj}} \\ \theta_j(t+1) = \theta_j(t) + \Delta\theta_j = \theta_j(t) - \lambda_1 \frac{\partial E}{\partial \theta_j} \\ \theta_k(t+1) = \theta_k(t) + \Delta\theta_k = \theta_k(t) - \lambda_2 \frac{\partial E}{\partial \theta_k} \end{cases} \tag{9}$$

where, t is the number of iterations, and λ_1 and λ_2 are the hidden and output layers' learning rates, respectively [16], [17].

B. BP Neural Network Learning Steps

Learning the BP neural network involves the following special steps s follows:

The first step is to initialize the connection weights and thresholds of the entire network, assign the connection weights ω_{ji}, ω_{kj} and the thresholds θ_j, θ_k to any number within (-1, 1) respectively, set the error function E, and determine the calculation accuracy and the maximum learning times M of the BP neural network [18]–[20].

The second step randomly selects the ith input sample and the corresponding expected output: X_i= [x_1, x_2, ... , x_i], Y_i=[y_1, y_2, ... ,y_i].

The third step is to calculate the input value and output value of each neuron in the hidden layer.

The fourth step is to calculate the partial derivative of the error function for each neuron in the output layer according to the error between the expected output value of the BP neural network and the theoretical value, as shown in Formula (10).

$$\frac{\partial E}{\partial z_k} \quad (k = 1,2,\cdots,n) \tag{10}$$

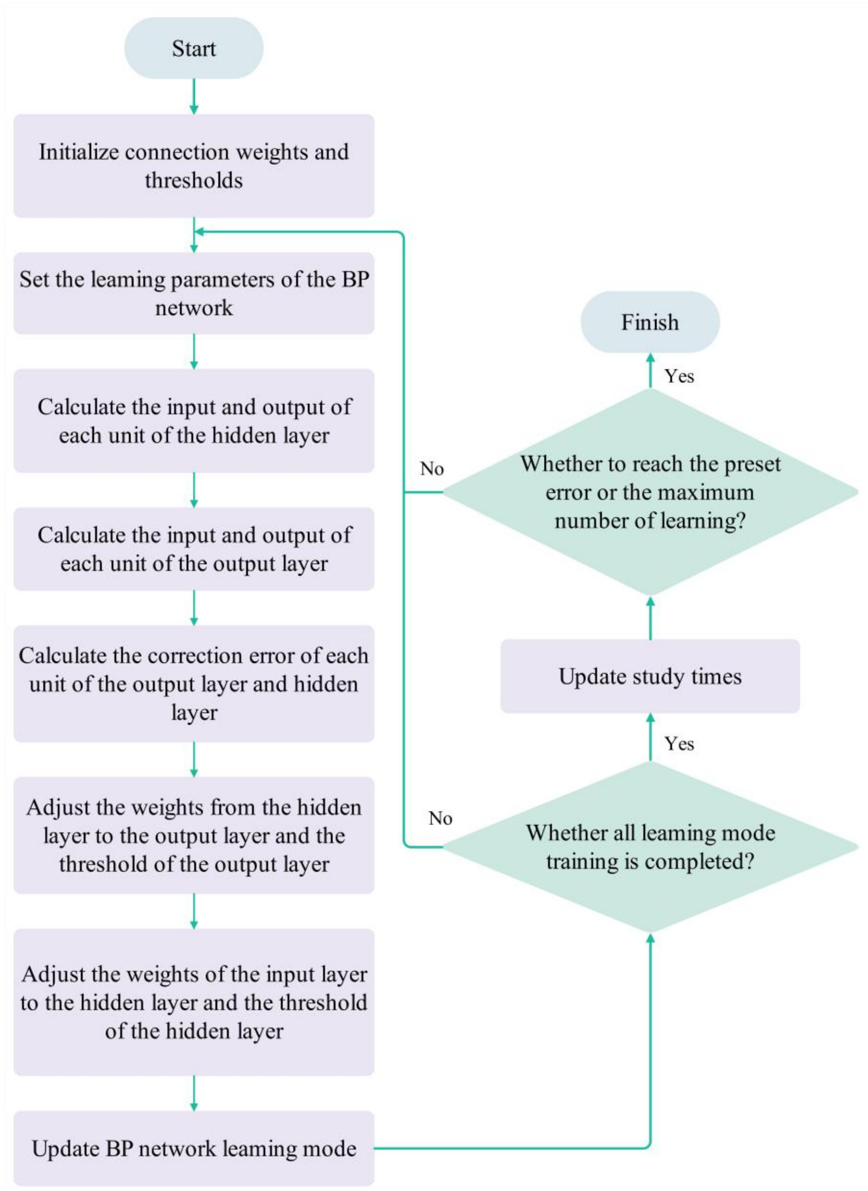

Fig. 4. BP neural network learning steps.

Step five involves calculating the derivative of the error function for each neuron in the hidden layer concerning the connection weight between the hidden and output layers, the derivative of the output layer, and the output value of the hidden layer, following the Formula (11).

$$\frac{\partial E}{\partial o_j} \quad (j = 1,2,\cdots,l) \tag{11}$$

Step six involves updating the connection weights based on the partial derivatives of each neuron in the output layer and the output of each neuron in the hidden layer [21], [22]

In the seventh step, the connection weights are updated according to the partial derivatives of each neuron in the hidden layer and the input of each neuron in the input layer.

The eighth step is calculating the global error, as shown in Formula (8).

In the ninth step, randomly select a new input sample as the BP neural network input and return to the third step until all input samples are trained.

The tenth step is to randomly re-select an input sample from the *m* input samples and return to the third step until the global error function *E* of the entire network is less than the preset precision [23]–[26].

The flowchart of the BP algorithm is shown in Fig. 4.

C. *The Software Basis of Art Teaching System*

The BP network structure of the music intelligent system has been stated as before. The intelligent system operates in a combination of software and hardware. A database server is arranged in the server environment, and the system is set to 8G memory, 500G solid-state hard disk, dual CPU processors system, and Gigabit Ethernet. The software adopts SQL Server

2000 as the database management system. SQL Server contains many tools for databases that can store, retrieve, and manage databases using the SQL language and GUI applications [27].

Ideal for mobile, wireless, and embedded applications, SQL Server CE provides important relational database functionality in a compact form factor, along with flexible data access and the familiar SQL Server experience. Extending enterprise applications to .NET Compact Framework greatly enhances the development of the user's mobile space.

SQL Server 2000 provides the same performance as many advanced database managers.

1) Rich programming interfaces and development tools: Query Analyzer is provided as a development tool for writing Transact-SQL script programs. Query Analyzer provides users a graphical working environment for writing and debugging Transact-SOL programs. Supports most commonly used database application programming interfaces, such as ADO, OLE DB, ODBC, etc. These tools allow programmers to directly control the interaction between the hemp program and the database and include APIs such as ADO that support rapid program development. These tools can develop a database application program with strong functions quickly.

2) Dynamic automatic management and configuration: SQL Server 2000 can be configured autonomously and dynamically during operation. For example, when the task of the database service increases, it will dynamically and autonomously apply for more system resources; when the work is reduced, the system resources will be released; when data is inserted or deleted in the database, the size of the database can be automatically adjusted to Adapt to new situations [28], [29].

3) Dynamic realization of database concurrency control: In SQL Server 2000, row-level blockade can be implemented on data. When dealing with the concurrency control problem of the database, the strength of data blocking will be dynamically adjusted according to different situations to achieve the best state of data blocking and sharing. For example, when a query needs to access a limited amount of data in a table, a row-level lock will be added to the data; when a query needs to access the vast majority of data in a table, a page-level lock will be added to the data, and take steps to make this query complete as quickly as possible. All the concurrency control performed by SQL Server 2000 is performed automatically in the background, and users do not need to interfere and participate [30].

V. Verification of Intelligent Art Teaching System

A. System Parameter Settings

In the BP neural network, the input layer uses 12 neurons, which are used to correspond to the 12-dimensional characteristic parameters of the musical sound signal; the output layer uses one neuron, which corresponds to the main frequency of the musical sound; the number of neurons in the hidden layer is 20.

Taking the piano music as an example to verify the system, considering that the piano keys only contain 88 keys, the data is relatively small, so choose a dedicated microphone and sound acquisition card to collect the 88 keys of the piano three times and use MIDI The format saves 264 musical sound data, forming the BP neural network input unit, normalized processing of the input data, and using the MATLAB toolbox to directly read the musical tone signal, main frequency value assigned to every note is known as the training output unit.

By setting the relevant decomposition filter bank in MATLAB, the musical sound signal is decomposed by Daubechies 4th-order wavelet, and the low-pass decomposition coefficients are retained. After that, the wavelet is reconstructed by setting the low-pass reconstruction filter bank, and the removal of high-frequency components is obtained. The perfect signal is Fourier transformed to obtain its spectrogram, as shown in Fig. 5.

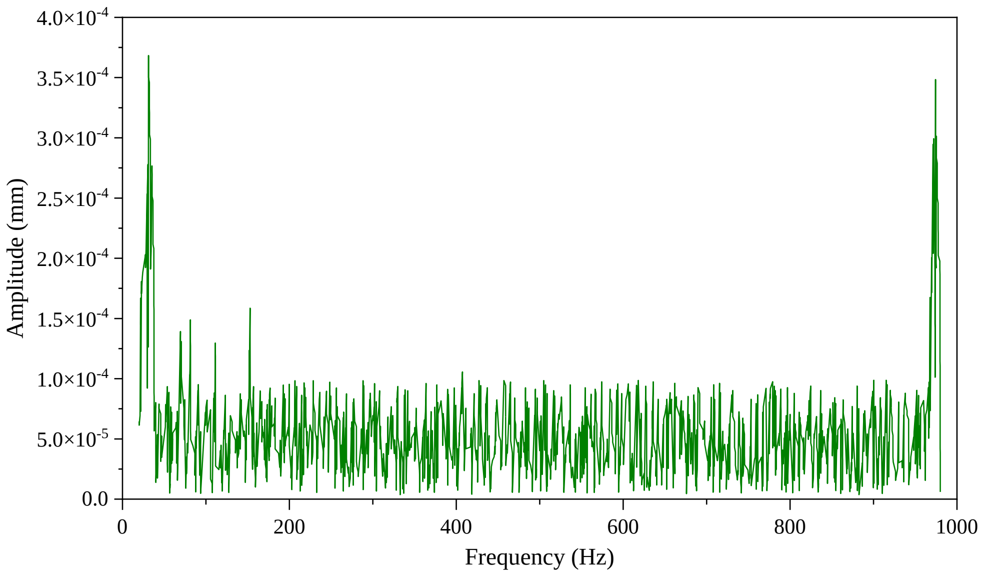

Fig. 5. Spectrogram of fourier transform.

B. Feature Parameter Extraction

After analyzing the characteristic parameters of the MFCC algorithm, a 12-order extraction algorithm has been chosen for the accuracy of musical tone recognition and to balance the amount of calculation. The length of the audio signal FFT transformation is 256, with a sampling frequency of 20500Hz and each frame having a length of 256 points. Then, the music file passed the Fourier transformation is read, and the Melbankm function normalizes the Mel filter bank coefficients. After that, the DCT parameters are solved, the pre-emphasis filter is filtered, and the music signal is re-framed. Ultimately, the musical sound signal's MFCC parameters are extracted, and the five frames of characteristic parameters are taken, as shown in Fig. 6.

C. Result Analysis

Select the characteristic parameters of 234 keys as the input of the training data and the corresponding main frequency value as the expected output result of the training, set the input neuron to 12, the output neuron to 1, the number of hidden layer neurons to 20, the maximum number of iterations is 2000, and the expected error is 10e-9. The BP neural network is trained in MATLAB, and the remaining 30 key characteristic parameters are selected as verification data. The error performance curve of the system is shown in Fig. 7. It is evident that after 1943 training times, the error reaches 10e-9.

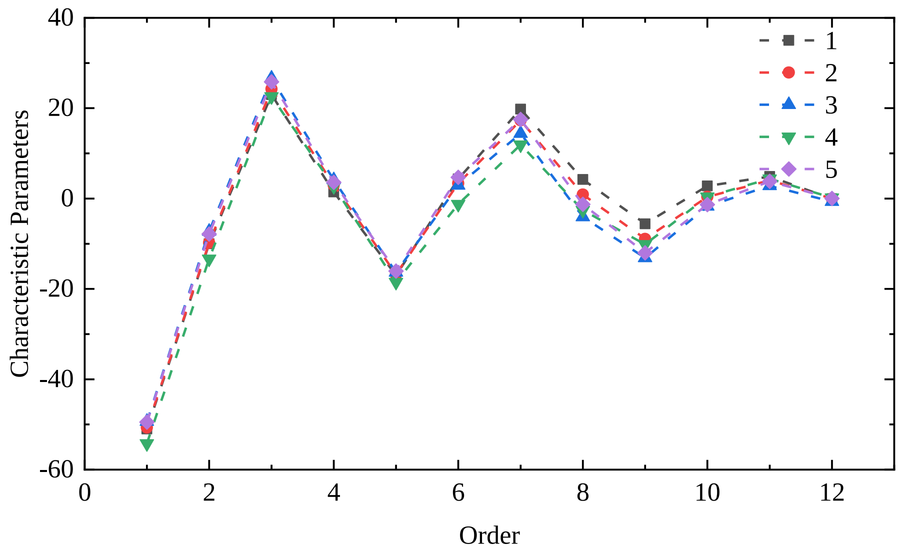

Fig. 6. Characteristic MFCC parameters.

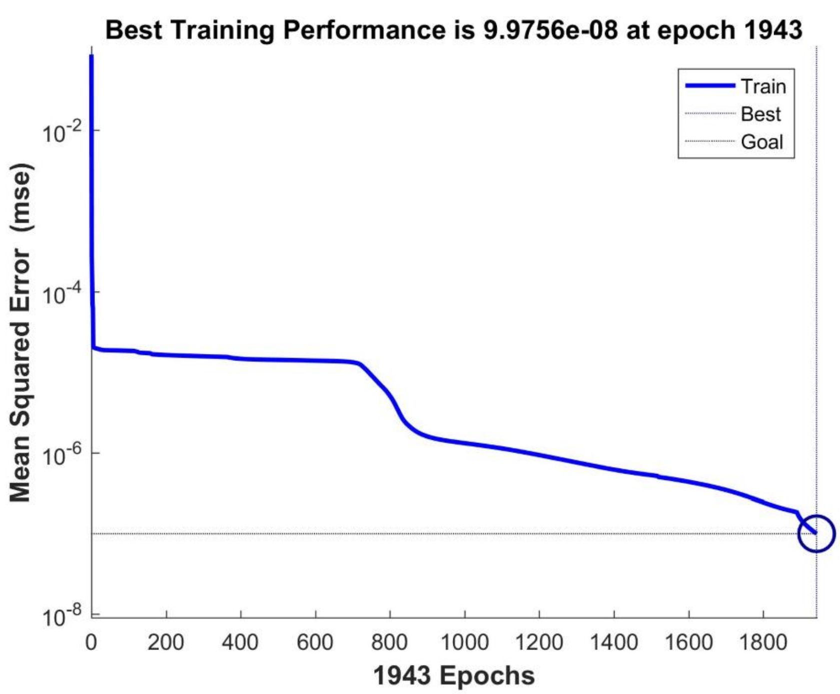

Fig. 7. BP network training performance.

The output result of the music sound teaching system using MFCC parameters as characteristic parameters for BP network matching is shown in Fig 8. It is observed that the expected value of the main frequency signal and the predicted value are consistent, with a relatively high recognition rate. It verifies the correct feasibility of the route proposed in the current paper using the BP neural network algorithm for musical tone recognition. However, from the identification results of musical tones in Table I and the relative error values in Fig. 9, it is clear that due to the compromise in wavelet filtering, identification errors in different situations will occur in low-frequency signals outside the selected area, while the relative error values of individual samples larger, which requires further investigation in subsequent studies.

TABLE I. TONE RECOGNITION RESULTS

Number	Expected value	Expected pitch names	Predictive value	Predictive pitch names
1	2637	E7	2586.67	E7
2	30.87	B0	29.2	A#0
3	220	A3	218.66	A3
4	523.2	C5	521.78	C5
5	49	G1	50.65	G#1
6	740	F#5	738.12	F#5
7	493.9	B4	497.43	B4
8	587.3	D5	585.98	D5
9	1175	D6	1172.56	D6
10	87.31	F2	85.55	F2
11	110	A2	108.82	A2
12	207.6	G#3	207.3	G#3
13	277.2	C#4	276.78	C#4
14	880	A5	878.28	A5
15	698.5	F5	697.99	F5
16	1661	G#6	1559.78	G#6
17	1245	D#6	1243.91	D#6
18	1760	A6	1756.5	A6
19	32.7	C1	31.34	B0
20	233.1	A#3	231.76	A#3
21	2794	F7	2789.66	F7
22	1661	G#6	1657.87	G#6
23	1109	C#6	1105.96	C#6
24	622.2	D#5	621.4	D#5
25	98	G2	97.6	G2
26	123.5	B2	124.1	B2
27	196	G3	194.85	G3
28	830.6	G#5	831.5	G#5
29	58.27	A#1	57.91	A#1
30	155.6	D#3	154.22	D#3

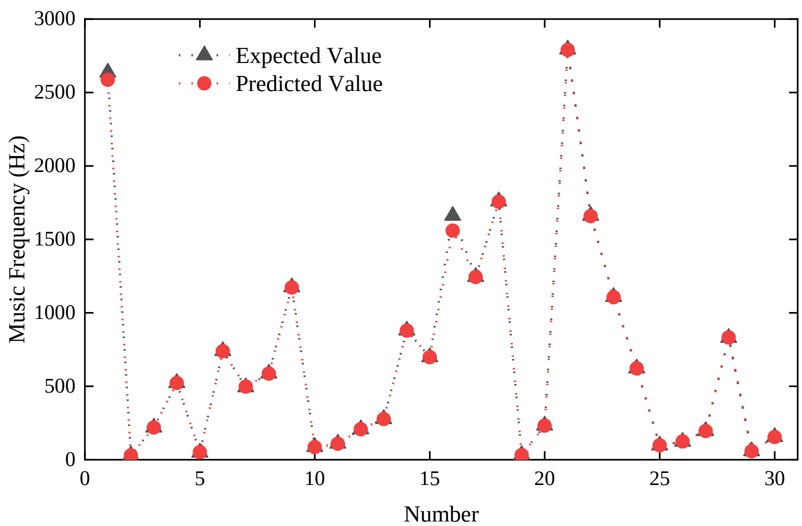

Fig. 8. Expected value and predicted value of main frequency.

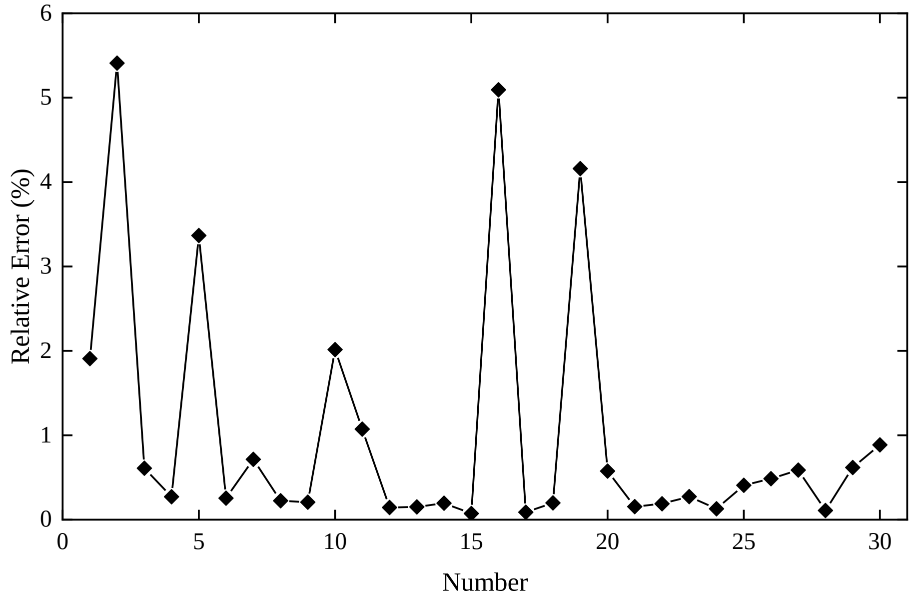

Fig. 9. Relative error between expected value and predicted value.

VI. Conclusion

The intelligent music teaching system can provide better learning concepts for art learners. Feature parameter extraction is done via MFCC, which is based on the BP neural network intelligent algorithm, and SQL Server is used as the music database management system to build a network matching of music signals to realize better the effect of the interactive teaching music intelligent system. This paper uses the piano sound signal to verify the system, and the BP learning algorithm is built in MATLAB. The verification results show that the predicted value of the network has achieved a high recognition effect, but there are certain errors in individual areas. The next research phase will involve using larger-scale model samples to realize artificial intelligence, resulting in improved accuracy across a wider range of applications. This will ensure greater stability in terms of usability and the ability to integrate various artificial intelligence algorithms to enhance the training algorithm of neural networks and create a more universal music intelligence system.

Funding

General project of art science of National Social Science Foundation、A study on the pattern drawing and construction thought of "Different people wear different uniforms" of men's clothing in the Northern Song Dynasty (No.18BG112).

Competing of interests

The authors declare no competing of interests.

Authorship Contribution Statement

Xianyu Wang: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Xiaoguang Sun: Methodology, Software, Validation.

Data Availability

On Request

Declarations

Not applicable

References

[1] W. Zhang, A. Shankar, and A. Antonidoss, "Modern art education and teaching based on artificial intelligence," Journal of Interconnection Networks, vol. 22, no. Supp01, p. 2141005, 2022.

[2] Q. Cao, "Curriculum design of art higher vocational education based on artificial intelligence assisted virtual reality technology," Security and Communication Networks, vol. 2022, pp. 1–9, 2022.

[3] C. He and B. Sun, "Application of artificial intelligence technology in computer aided art teaching," Comput Aided Des Appl, vol. 18, no. S4, pp. 118–129, 2021.

[4] D. Lin, Z. Naiyao, and Z. Hancheng, "A review on the research of music features recognition," Computer Engineering and Applications (Beijing, China), vol. 38, no. 24, pp. 74–77, 2002.

[5] N. Li and M. J. Bin Ismail, "Application of artificial intelligence technology in the teaching of complex situations of folk music under the vision of new media art," Wirel Commun Mob Comput, vol. 2022, pp. 1–10, 2022.

[6] Y. Shen, "RETRACTED: Analysis of the Possibility of Art Theory Construction from the Perspective of Art Image based on Computer-aided Analysis," in Journal of Physics: Conference Series, IOP Publishing, 2020, p. 032146.

[7] X. Wang and T. Wang, "Voice Recognition and Evaluation of Vocal Music Based on Neural Network," Comput Intell Neurosci, vol. 2022, 2022.

[8] Q. Mei, M. Gül, and M. Boay, "Indirect health monitoring of bridges using Mel-frequency cepstral coefficients and principal component analysis," Mech Syst Signal Process, vol. 119, pp. 523–546, 2019.

[9] M. G. de Pinto, M. Polignano, P. Lops, and G. Semeraro, "Emotions understanding model from spoken language using deep neural networks and mel-frequency cepstral coefficients," in 2020 IEEE conference on evolving and adaptive intelligent systems (EAIS), IEEE, 2020, pp. 1–5.

[10] N. Yang, N. Dey, R. S. Sherratt, and F. Shi, "Recognize basic emotional statesin speech by machine learning techniques using mel-frequency cepstral coefficient features," Journal of Intelligent & Fuzzy Systems, vol. 39, no. 2, pp. 1925–1936, 2020.

[11] S. Song, X. Xiong, X. Wu, and Z. Xue, "Modeling the SOFC by BP neural network algorithm," Int J Hydrogen Energy, vol. 46, no. 38, pp. 20065–20077, 2021.

[12] Q. Liu et al., “Comparative analysis of BP neural network and RBF neural network in seismic performance evaluation of pier columns,” Mech Syst Signal Process, vol. 141, p. 106707, 2020.

[13] W. Shen et al., “Assessment of dairy cow feed intake based on BP neural network with polynomial decay learning rate,” Information Processing in Agriculture, vol. 9, no. 2, pp. 266–275, 2022.

[14] T. Li, J. Sun, and L. Wang, “An intelligent optimization method of motion management system based on BP neural network,” Neural Comput Appl, vol. 33, pp. 707–722, 2021.

[15] Z. Qu, W. Mao, K. Zhang, W. Zhang, and Z. Li, “Multi-step wind speed forecasting based on a hybrid decomposition technique and an improved back-propagation neural network,” Renew Energy, vol. 133, pp. 919–929, 2019.

[16] B. Sang, “Innovation of enterprise technology alliance based on BP neural network,” Neural Comput Appl, vol. 33, pp. 807–820, 2021.

[17] Y. Zhang 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, pp. 1273–1291, 2021.

[18] B. Liu et al., “Prediction of rock mass parameters in the TBM tunnel based on BP neural network integrated simulated annealing algorithm,” Tunnelling and Underground Space Technology, vol. 95, p. 103103, 2020.

[19] J. Yang, Y. Hu, K. Zhang, and Y. Wu, “An improved evolution algorithm using population competition genetic algorithm and self-correction BP neural network based on fitness landscape,” Soft comput, vol. 25, pp. 1751–1776, 2021.

[20] W. Wang, R. Tang, C. Li, P. Liu, and L. Luo, “A BP neural network model optimized by mind evolutionary algorithm for predicting the ocean wave heights,” Ocean Engineering, vol. 162, pp. 98–107, 2018.

[21] S. Wang, T. H. Wu, T. Shao, and Z. X. Peng, “Integrated model of BP neural network and CNN algorithm for automatic wear debris classification,” Wear, vol. 426, pp. 1761–1770, 2019.

[22] C. Yan, M. Li, W. Liu, and M. Qi, “Improved adaptive genetic algorithm for the vehicle Insurance Fraud Identification Model based on a BP Neural Network,” Theor Comput Sci, vol. 817, pp. 12–23, 2020.

[23] T. Shen, Y. Nagai, and C. Gao, “Design of building construction safety prediction model based on optimized BP neural network algorithm,” Soft comput, vol. 24, pp. 7839–7850, 2020.

[24] C. Huang, Y. Zhao, W. Yan, Q. Liu, and J. Zhou, “A new method for predicting crosstalk of random cable bundle based on BAS-BP neural network algorithm,” IEEE Access, vol. 8, pp. 20224–20232, 2020.

[25] S. Zhou et al., “Dual-optimized adaptive Kalman filtering algorithm based on BP neural network and variance compensation for laser absorption spectroscopy,” Opt Express, vol. 27, no. 22, pp. 31874–31888, 2019.

[26] W. Ma, Q. Li, J. Li, L. Ding, and Q. Yu, “A method for weighing broiler chickens using improved amplitude-limiting filtering algorithm and BP neural networks,” Information Processing in Agriculture, vol. 8, no. 2, pp. 299–309, 2021.

[27] A. T. Raj, G. Nalinipriya, M. Shobana, D. Bharath, and S. P. Hariharasudhan, “A Vibrant GUI Based Data Handling Using Relational Database Framework,” in 2022 8th International Conference on Smart Structures and Systems (ICSSS), IEEE, 2022, pp. 1–6.

[28] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao, “A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends,” Autom Constr, vol. 101, pp. 127–139, 2019.

[29] J. S. Horsburgh, S. L. Reeder, A. S. Jones, and J. Meline, “Open source software for visualization and quality control of continuous hydrologic and water quality sensor data,” Environmental Modelling & Software, vol. 70, pp. 32–44, 2015.

[30] R. Winkelmann, J. Harrington, and K. Jänsch, “EMU-SDMS: Advanced speech database management and analysis in R,” Comput Speech Lang, vol. 45, pp. 392–410, 2017.