ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Emotional Response Analysis and Teaching Improvement Strategies in Music Education Based on Pattern Recognition

Yanyan Wang[1,*]

[1] School of Education, Shanghai Donghai Vocational and Technical College, Shanghai, 20072, China

SUMMARY: *This study utilizes artificial intelligence technology to design a music education model based on pattern recognition. Music signals are preprocessed to extract emotional feature vectors from the signals. A music emotion recognition mapping framework is established, and a backpropagation (BP) neural network is employed for automatic music emotion recognition. Combining emotional teaching measures in music education, the application scheme for the emotion recognition teaching model is determined. Through teaching practice, the effectiveness of the proposed scheme in improving music education is evaluated. 96% of students believe that the model-assisted approach is helpful for classroom learning, and 87% of students actively use the model for learning after class. All students have improved their interest and confidence in music learning through teaching practice. Regarding students' attitudes toward teaching methods, 90% of students approve of the AI-assisted teaching model, and after learning, 96.5% of students feel that learning music is an enjoyable experience.*

KEYWORDS: *music education; BP neural network; emotional feature vector; music emotion recognition; teaching practice*

1 Introduction

Music is a non-verbal form of expression that conveys emotions and information through sound and melody. Emotion, as one of the core elements of a musical piece, enables students to develop their emotional expression skills through music education, learning to use music to articulate their inner feelings and emotions [1, 2]. They can express their emotions and thoughts through playing instruments, singing, or composing music, thereby enhancing their communication and interpersonal skills [3]. Additionally, in music education, students' emotional responses in the classroom and their emotional responses to musical works are important factors for teachers to consider when providing instructional guidance [4]. However, in traditional music education, teachers rely solely on their personal experience to assess students' emotional responses, neglecting situations where students' emotional responses are subjective, hidden, not obvious, or fleeting, leading to insufficient personalized instructional guidance [5]. Furthermore, students' emotional responses are influenced by their own emotions, psychological factors, and external factors, which can reduce their enthusiasm and sustainability in music learning. However, teachers' assessments struggle to capture and identify these emotional responses [6, 7]. This situation highlights the limitations of traditional music education and the inevitable trend toward the intelligent transformation of music education.

*mmaxyxxy@163.com
https://doi.org/10.65102/is2026656
Published: 30 April, 2026

With the rapid development of information technology in today's society, the application of artificial intelligence is becoming increasingly widespread. Pattern recognition is one aspect of artificial intelligence application and plays an important role in the intelligent transformation of education [8]. Pattern recognition refers to the process of processing and analyzing various forms of information (numerical, textual, and logical relationships) that represent objects or phenomena, in order to describe, identify, classify, and interpret those objects or phenomena [9]. Moreover, the application of pattern recognition is increasingly receiving attention and support, with significant progress in various areas, such as the recognition of multiple modalities including text, speech, physiology, and facial features, which can provide services for emotional analysis [10-12].

Reference [13] constructs a novel long short-term memory network model to simulate human auditory and visual perception, which is used to analyze multimodal music emotions (emotional information and facial expressions). This method can effectively identify students' emotions and serves as an important basis for real-time teaching evaluation in distance education. Literature [14] employs constrained non-negative matrix factorization (CNMF), CNMF with external information, and model optimization algorithms to construct a mathematical model, combined with digital audio technology to analyze emotions in musical performances. Literature [15] proposes an optimized brain-emotion learning model that integrates brain neural region signals and the Taylor psychological model to identify test subjects' musical emotional feedback. Literature [16] shares an explanatory and predictive model based on a new concept of quantitative physics-physiology relationships, and combines a random forest regression model to explore the association between musical acoustic features and human emotional responses. Musical rhythm, timbre, and pitch can evoke emotional responses, while individual attributes such as gender and age also have a tendency to influence them. Literature [17] developed a classroom student emotion and engagement detection system using a convolutional neural network model, which can accurately capture changes in students' emotions in the classroom. This system facilitates teachers in providing remedial feedback and interactive teaching. Literature [18] established a new emotion recognition model using an improved frame attention network to identify facial expressions of students in online music teaching, thereby capturing their emotional states.

This paper first introduces the preprocessing workflow for music signals and provides a detailed explanation of methods for extracting time-domain and frequency-domain features. It proposes a music emotion recognition model based on a backpropagation (BP) neural network, analyzing its mapping framework and training process. By combining emotional teaching cases, it explores teaching strategies such as scenario setting and emotional factor extraction. Through simulation experiments, it demonstrates the feature extraction workflow. The BP neural network model is constructed on the MATLAB platform to determine the optimal model parameters. A 16-week teaching practice is designed to examine the impact of the proposed teaching model on students' emotional experiences and learning outcomes.

2 The application of artificial intelligence in emotional teaching in music education

Music education, as an important component of aesthetic education, has a direct impact on students' aesthetic abilities and artistic literacy through its emotional teaching effects. Traditional music teaching relies heavily on teachers' subjective experience and lacks quantitative analysis of students' emotional responses. With the development of artificial intelligence technology, pattern recognition has provided new insights into the extraction and

classification of emotional features in music.

2.1 Extraction of musical emotional feature vectors

2.1.1 Music Signal Preprocessing

Music signals not only record instrument sounds and vocals, but also contain some environmental noise, and the file formats used to record them vary greatly. Therefore, it is necessary to preprocess the music before extracting its emotional features. Music signal preprocessing mainly focuses on some commonly used short-term analysis and processing techniques to prepare for subsequent feature extraction. The music signal preprocessing process is shown in Figure 1.

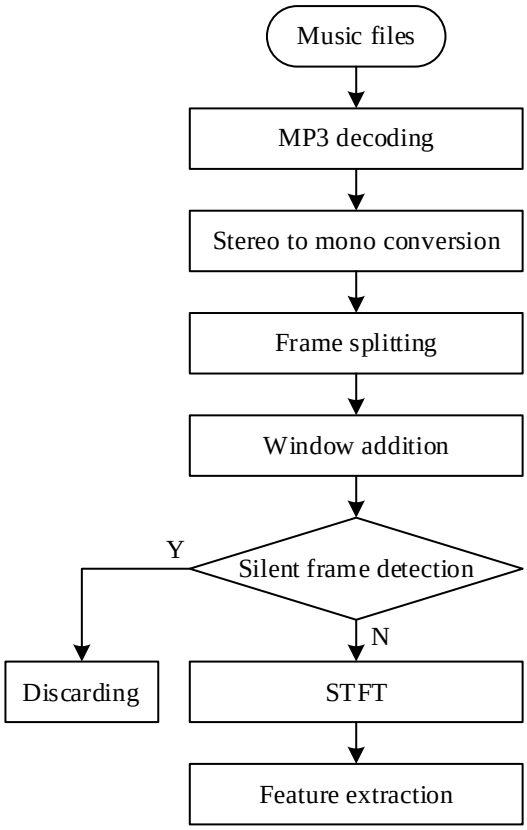

Figure 1: Music signal preprocessing process

On our personal computers or the internet, digital music is typically stored in MP3 format because it takes up less space and offers high-quality sound. For MP3 files, we first need to perform MP3 decoding, which results in data streams for the left and right channels. However, analyzing and processing stereo sound from two channels is quite complex and not the focus of this paper. Therefore, we calculate the average of the two channel data streams to obtain a single-channel data stream.

After obtaining the monaural data stream, we need to perform frame division on the data stream, with each frame defined as 1024 sample points (approximately 10.68 ms). After framing, a windowing operation is performed, with the Hamming window selected as the window function to ensure smooth transitions between adjacent frames. Next, silent frames are identified and discarded if present. Otherwise, a short-time Fourier transform is performed to

obtain the frequency spectrum information of the music signal. At this point, the preprocessing is complete, and the process proceeds to frame-level feature extraction.

2.1.2 Frame-level feature extraction

The musical emotional features extracted in this paper primarily consist of two types of features: time-domain and frequency-domain features. Time-domain features include the time-domain zero-crossing rate. The preprocessing steps for extracting this feature differ slightly and do not require windowing or Fourier transformation. Frequency-domain features include: spectral center, spectral decay value, spectral flow, MFCC cepstrum coefficients, and syllable coefficients.

The following sections provide a detailed explanation of the principles and calculation methods for these different features:

Time-domain zero-crossing rate: In the time domain, the zero-crossing rate is the simplest type of feature. For the time-domain discrete signal of digital music, when the algebraic signs of two consecutive sample points are different, it is referred to as a signal zero-crossing. The zero-crossing rate refers to the rate at which the signal crosses zero. The time-domain zero-crossing rate to some extent reflects the spectral characteristics of the signal. The specific calculation is as follows:

$$Z_r = \frac{1}{2(N-1)} \sum_{m=1}^{N-1} \left| \mathrm{sgn}[x(m+1)] - \mathrm{sgn}[x(m)] \right| \tag{1}$$

In this context, $x(m)$ represents the time-domain discrete signal of music, N denotes the number of samples in a frame of the signal, and $\mathrm{sgn}[\ ]$ is the sign function. Research on human speech production indicates that speech consists of two components: voiceless sounds and voiced sounds. However, music signals do not follow this pattern, so the zero-crossing rate of speech is higher than that of music; Additionally, the zero-crossing rate is lower in voiced segments and higher in unvoiced segments, allowing unvoiced and voiced sounds to be distinguished using the time-domain zero-crossing rate. In music containing vocals, the faster the singer's speech rate and the denser the speech, the higher the signal's zero-crossing rate.

Spectral center: The spectral center is defined as the center of the spectrum obtained by applying the short-time Fourier transform (STFT) to the audio signal, representing the balanced point of the signal across the entire spectrum.

$$C_t = \frac{\sum_{n=1}^{N} M_t(n) \times n}{\sum_{n=1}^{N} M_t(n)} \tag{2}$$

Among them, $M_t(n)$ refers to the STFT transform of the audio frame at time t. The spectral center is an indicator that describes the shape of the spectrum. The larger the spectral center value, the more high-frequency components are contained in the music signal.

Spectral attenuation value: The spectral attenuation value refers to the spectral energy less than frequency R_t, which is exactly equal to 90% of the total spectral energy.

$$Rolloff = \sum_{n=1}^{R_t} M_t(n) = 0.85 \sum_{n=1}^{N} M_t(n) \tag{3}$$

The spectral attenuation value reflects the shape of the spectrum.

Spectral flow: Spectral flow is defined as the standard deviation of the continuous spectral distribution.

$$F_t = \sum_{n=1}^{N} (N_t[n] - N_{t-1}[n])^2 \tag{4}$$

Among these, $N_t[n]$ and $N_{t-1}[n]$ represent the Fourier transforms of the audio frames at time t and $t-1$, respectively. This metric quantifies the total amount of local spectral changes and, to some extent, reflects the shape of the spectrum.

Mel-frequency cepstral coefficients (MFCC): Mel-frequency cepstral coefficients are frequently used in speech recognition and music classification. This feature reflects the auditory characteristics of the human ear, so many speech recognition systems currently use this feature to achieve excellent speech recognition results.

As we all know, in noisy environments, we can not only distinguish between various sounds but also filter out certain sounds. Further research has found that this is due to the role of the cochlea, which acts as a set of filters on a logarithmic frequency scale, performing a similar filtering function on sounds. Therefore, the human ear's ability to distinguish between multiple sounds is not linear; it is linear below 1 kHz and logarithmic above 1 kHz.

$$Mel(f) = 2595 \log\left(1 + \frac{f}{700}\right) \tag{5}$$

The spectrum obtained after mapping using Equation (5) is referred to as the Mel-spectrogram domain, and Mel-spectrogram coefficients refer to a set of parameters extracted according to a fixed theoretical framework after mapping a digital music signal to the spectrogram domain, which reflect the emotional characteristics of the music. The specific calculation is as follows: first, the music signal is subjected to a Fourier transform, then mapped to the cepstrum domain according to Equation (5), followed by filtering using several (typically 13) equally spaced triangular filters, and finally, the output signals from these filters are summed in terms of energy, i.e., calculated using the following formula:

$$C_n = \sqrt{\frac{2}{K}} \sum_{k=1}^{K} \left\{ \lg[x(k)] \times \cos\frac{\pi(k-0.5)}{K} \right\} (n = 1, 2, \ldots, 13) \tag{6}$$

Among these, $x[k]$ represents the output energy of the k th filter.

Syllable coefficients: A pure octave pitch is divided into twelve equal parts, and the pitch distance between each part is defined as a semitone. This method is called the twelve-tone equal temperament. In some recent studies, the structural analysis of music is increasingly determined by syllable coefficients, which are used to describe the power spectrum energy distribution among 12 similar pitch classes, achieving good classification results. Therefore, this paper follows this method to extract twelve syllable coefficients from the cepstrum domain, where $v_c(t)$ represents the c th pitch class $(c = 1, 2, \ldots, 12)$ on a given octave:

$$v_c(t) = \sum_{Oct_L}^{Oct_H} \int_{-\infty}^{+\infty} B_{BPF_{c,h}}(f)\psi(f,t)df \tag{7}$$

Among these, $\psi(f,t)$ represents the frequency f in the spectrogram domain at time t, obtained through the STFT transformation; Oct_L and Oct_H denote the octave ranges, which are 3 and 8, respectively, with a frequency coverage range of 130 Hz to 8000 Hz; $B_{\mathrm{BPF}_{c,h}}(f)$ is a Hanning window bandpass filter that allows the logarithmic scale frequency $F_{c,h}$ to pass through.

$$B_{BPF_{c,h}}(f) = \frac{1}{2}\left(1 - \cos\frac{2\pi(f - (F_{c,h}(f) - 100))}{200}\right) \tag{8}$$

$F_{c,h}$ denotes the logarithmic scale frequency on the cth pitch class at octave position h, expressed as:

$$F_{c,h} = 1200h + 100(c-1) \tag{9}$$

After obtaining the 12-dimensional syllable coefficients, in order to reflect the differences between the maximum, minimum, and average values of the syllables, the peak-to-mean ratio and peak-to-trough ratio are defined as follows:

$$R_{\text{max-}avg} = \frac{v_{\max}}{v_{avg}} \tag{10}$$

$$R_{\text{max-min}} = \frac{v_{\max}}{v_{\min}} \tag{11}$$

Among these, $v_{\max}$, $v_{\min}$, and v_{avg} represent the maximum, minimum, and average values among the twelve syllable coefficients, respectively. Ultimately, the syllable coefficients will be 14-dimensional.

2.2 Automatic recognition model for musical emotions

2.2.1 Music Emotion Recognition Mapping Framework

Natural phenomena and objects can generally be categorized into groups or individuals that are similar yet not entirely identical. Such categories are referred to as pattern classes or patterns, while each individual phenomenon or object within a category is termed a sample of the pattern. Samples within the same category share similarities and common characteristics, whereas samples from different categories are distinct from one another. Pattern recognition essentially involves the accurate mapping from the pattern space to the category membership space.

From a mathematical perspective, the recognition of musical emotions is equivalent to pattern recognition, requiring a correct mapping process from a high-dimensional musical feature space to a low-dimensional emotional space. The mapping framework is shown in Figure 2. A music emotion recognition system must, based on a set of music samples annotated with emotions, use some method to identify the regularities of emotion recognition and establish cognitive discrimination formulas and rules. Thus, when new music samples are input, they can

be effectively discriminated and their emotional categories determined. Therefore, the key to this system lies in constructing a recognition model with high accuracy.

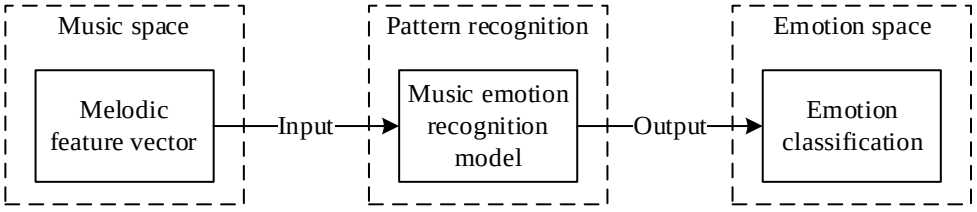

Figure 2: Mapping Framework for music emotion recognition

2.2.2 BP Network

A BP network is a multi-layer feedforward neural network named after the backpropagation learning algorithm used to adjust the network weights. The BP network is the core component of feedforward networks and embodies the most essential and perfect aspects of neural networks.

Training steps for a BP network:

(1) Initialization. Assign random values within the interval $(-1,1)$ to each connection weight w_{ij}, v_{jt}, threshold θ_j, and γ_t.

(2) Randomly select a set of input and target samples $P_k = (a_1^k, a_2^k, \ldots, a_n^k)$ and $T_k = (s_1^k, s_2^k, \ldots, s_n^k)$ to feed into the network.

(3) Use the input samples $P_k = (a_1^k, a_2^k, \ldots, a_n^k)$, connection weights w_{ij}, and thresholds θ_j to calculate the inputs s_j of each unit in the hidden layer, then use s_j to calculate the outputs b_j of each unit in the hidden layer through the transfer function.

$$s_j = \sum_{i=1}^{n} w_{ij} a_i - \theta_j \quad j = 1, 2, \ldots, p \tag{12}$$

$$b_j = f(s_j) \quad j = 1, 2, \ldots, p \tag{13}$$

(4) Calculate the output L_t of each unit in the output layer using the output b_j of the middle layer, the connection weight v_{jt}, and the threshold γ_t, and then calculate the response C_t of each unit in the output layer using the transfer function.

$$\begin{aligned} L_t &= \sum_{j=1}^{p} v_{jt} b_j - \gamma_t \quad t = 1, 2, \ldots, q \\ C_t &= f(L_t) \quad t = 1, 2, \ldots, q \end{aligned} \tag{14}$$

(5) Using the network target vector $T_k = (y_1^k, y_2^k, \ldots, y_n^k)$ and the actual output C_t of the network, calculate the generalized error d_t^k of each unit in the output layer.

$$d_t^k = (y_t^k - C_t) \cdot C_t \cdot (1 - C_t) \quad t = 1, 2, \ldots, q \tag{15}$$

(6) Calculate the generalized error e_j^k of each unit in the intermediate layer using the connection weight v_{jt}, the generalized error d_t^k of the output layer, and the output b_j of the intermediate layer.

$$e_j^k = \left[\sum_{t=1}^{q} d_t^k \cdot v_{jt}\right] b_j(1-b_j) \tag{16}$$

(7) Use the generalized error d_t^k of each unit in the output layer and the output b_j of each unit in the intermediate layer to correct the connection weight v_{jt} and threshold γ_t.

$$v_{jt}(N+1) = v_{jt}(N) + \alpha \cdot d_t^k \cdot b_j \tag{17}$$

$$\begin{gathered} \gamma_t(N+1) = \gamma_t(N) + \alpha \cdot d_t^k \\ t = 1,2,\ldots,q,\ j = 1,2,\ldots,p, 0 < \alpha < 1 \end{gathered} \tag{18}$$

(8) Use the generalized error e_j^k of each unit in the middle layer and the input $P_k = (a_1, a_2, \ldots, a_n)$ of each unit in the input layer to correct the connection weight w_{ij} and threshold θ_j.

$$w_{ij}(N+1) = w_{ij}(N) + \beta \cdot e_j^k \cdot a_i^k \tag{19}$$

$$\begin{gathered} \theta_j(N+1) = \theta_j(N) + \beta \cdot e_j^k \\ i = 1,2,\ldots,n,\ j = 1,2,\ldots,p, 0 < \beta < 1 \end{gathered} \tag{20}$$

(9) Randomly select the next learning sample vector and provide it to the network, then return to step (3) until m training samples have been trained.

(10) Randomly select a new set of input and target samples from the m learning samples, then return to step (3) until the global error E of the network is less than a pre-set minimum value, i.e., the network converges. If the number of learning iterations exceeds the predefined value, the network cannot converge.

(11) Learning ends.

Note: In the above learning steps, steps (7) to (8) constitute the “backpropagation process” for network error, while steps (9) to (10) are used to complete the training and convergence process.

2.3 Emotional teaching measures in music education

(1) Exploring emotional factors to promote aesthetic development

As students' thinking matures and their awareness of learning strengthens, they become more interested in new things. In music education, it is essential to emphasize the flexible and innovative application of teaching methods to stimulate students' active participation in learning, making the process of acquiring musical knowledge enjoyable. This approach truly enhances students' efficient learning. Integrating emotional factors into music education helps engage students' emotions and foster the development of aesthetic awareness. In actual teaching,

teachers need to analyze musical works and organize effective musical activities to allow students' emotions to flourish, thereby enhancing the quality of aesthetic education.

(2) Setting the scene to stimulate emotions

When learning music knowledge, students' thinking is not yet fully developed. Therefore, in actual music teaching, teachers should focus on scientifically guiding students in learning music knowledge, and the application of emotions can be demonstrated through diverse forms. For example, by creating emotional contexts for students, they can engage their emotions within these contexts, which positively promotes their learning of music knowledge with emotional involvement. The scientific application of emotional teaching methods can effectively promote the implementation of music aesthetic education. In song composition, constructing scene structures is essential, and contextual teaching methods are particularly effective in stimulating students' musical emotions and promoting efficient learning.

3 Analysis of music teaching practices based on pattern recognition

3.1 Simulation experiment analysis

Emotions are subjective, and different music genres, forms, and styles evoke different emotional responses. In the simulation experiment section of this paper, 300 pieces of music were carefully selected from a large MIDI music library to form a dataset that sufficiently covers all emotions. Over the course of one month, 20 domain experts listened to each piece of music and classified it. The results were recorded using a survey questionnaire to determine the emotional tone of each piece as perceived by the majority of people. The emotional classification of each piece of music was determined based on the principle of majority rule. If less than half of the experts agreed on a particular classification, it was discarded. The final emotional classification of each piece of music was obtained, and emotional recognition was based on statistical learning using the labeled real emotional data.

3.1.1 Music Signal Preprocessing

Audio signals exhibit the characteristic of being “short-term stationary and long-term non-stationary,” meaning that over a short period of time, their characteristics can be approximated as stable and unchanged, while beyond this timeframe, they become typical non-stationary signals. Research has shown that these short time intervals are referred to as audio frames, which serve as the smallest unit in audio processing. Short-term framing is typically employed for framing processing. Selecting one sample from the dataset and applying framing and windowing processing, the windowed framing results for this music segment are shown in Figure 3. It can be seen that audio frames serve as the smallest unit for audio analysis and processing, with audio features first extracted at the frame level to obtain low-level audio features.

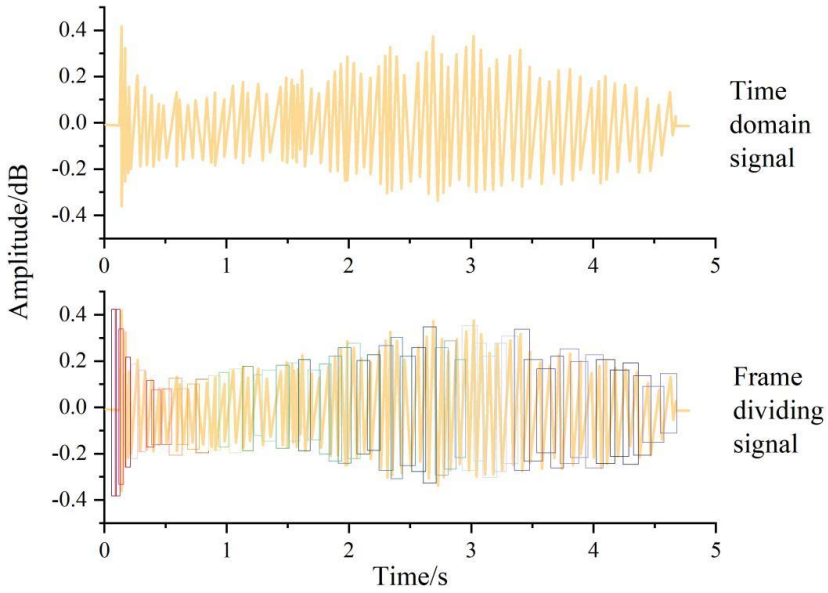

Figure 3: Windowed framing results

This paper uses the zero-crossing rate to capture the temporal characteristics of music, employing an acoustic description toolbox. The zero-crossing rate is the number of times the discrete sample signal values change from positive to negative and from negative to positive within a short-time frame. This quantity roughly reflects the average frequency of the signal within the short-time frame. The zero-crossing rate of the sample music segment is shown in Figure 4, where it is observed that the zero-crossing rate changes over time. When $x(n) \geq 0$, sign[x(n)] = 1; otherwise, sign[x(n)] = 0.

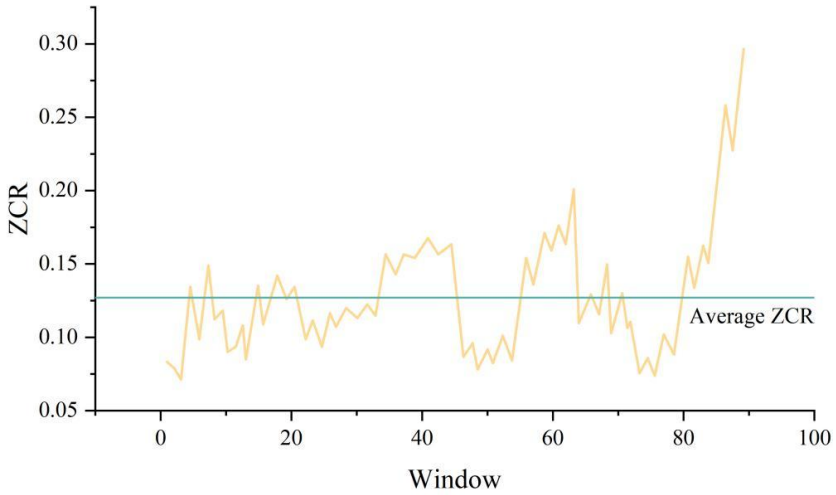

Figure 4: Zero-crossing rate of sample music segments

3.1.2 Extraction of emotional feature vectors

The method proposed in this paper was used to extract feature vectors for each MIDI file, and six feature vectors were calculated for each piece of music. Feature vectors were calculated for the selected group of pieces, and some of the results are shown in Table 1. The sample music showed significant differences in six emotional feature dimensions, with music tempo and average dynamics showing a large range of fluctuations, while the average pitch change was

relatively smooth.

Table 1: Feature vector results of the sample music

Music Piece Number	The directionality of the melody	Average pitch	Musical tempo	Average force	Note density	Pitch stability
1	1.46	5.38	0.65	76.48	4.78	1.46
2	2.15	5.17	0.45	68.37	5.18	1.75
3	0.93	4.93	0.75	71.66	4.67	2.64
4	1.56	4.72	1	57.32	4.22	2.11
5	1.78	5.26	0.85	68.15	5.84	1.84
6	-0.47	5.39	1	74.38	6.52	1.15
7	1.36	4.88	1	68.37	5.78	2.75
8	1.64	4.97	0.85	79.43	5.37	1.64
9	1.22	5.06	0.35	66.47	5.66	2.42
10	-0.56	5.13	1	70.11	6.04	1.77
...	...	...	...	...	...	...

The six feature vectors are numbered X1 to X6, respectively. The change curves of different emotional features are shown in Figure 5. It can be seen from the figure that the change amplitude of the average pitch curve is very small, indicating that the average pitch feature does not have a significant impact on emotion recognition. However, the change amplitudes of music tempo and average intensity are relatively large, indicating that they play a significant role in emotion recognition.

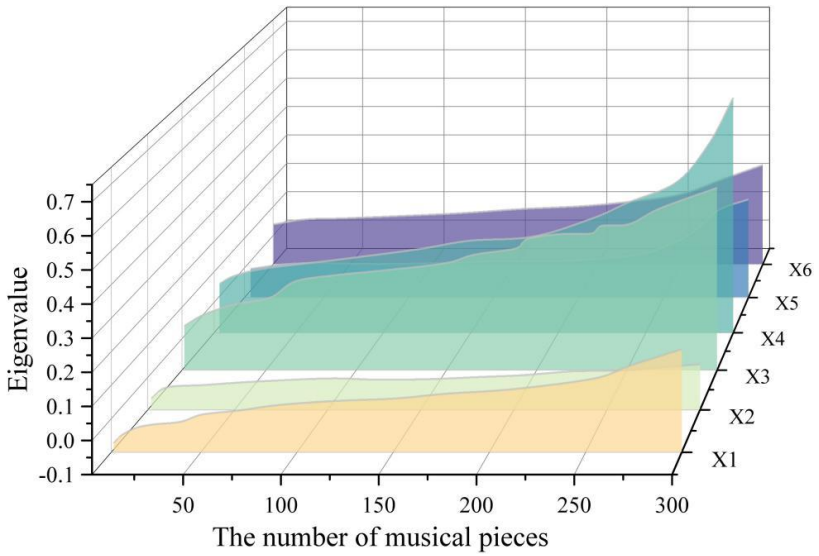

Figure 5: The variation curves of different emotional characteristics

Subsequently, one type corresponds to six emotional feature vectors, which constitute the sample space for training the BP network. In the following design, the six emotional feature vectors are the inputs to the neural network, and the one type is the output of the neural network. Among the 300 songs, 100 songs that cover the emotional space of music are selected as training samples for the BP neural network, and the remaining 200 songs are reserved for testing.

3.1.3 Training and Testing of the BP Network Model

After constructing the BP neural network model on the MATLAB platform, the input sample set began training, and the training results demonstrated that the network could converge. The training time and classification accuracy of the BP neural network model are not only related to the model's supervision algorithm but also to the number of layers in the model and the number of neurons in each layer. The specific simulation results of BP neural network models with different depths are shown in Figure 6. Increasing the number of layers and neurons both enhance the network model's “learning” capability, i.e., improve classification accuracy. However, their effects on the overall training time are diametrically opposed: increasing the number of layers prolongs training time, while increasing the number of neurons gradually reduces training time. By balancing training time and accuracy, this paper adopts a model with 7 hidden layers and 70 neurons.

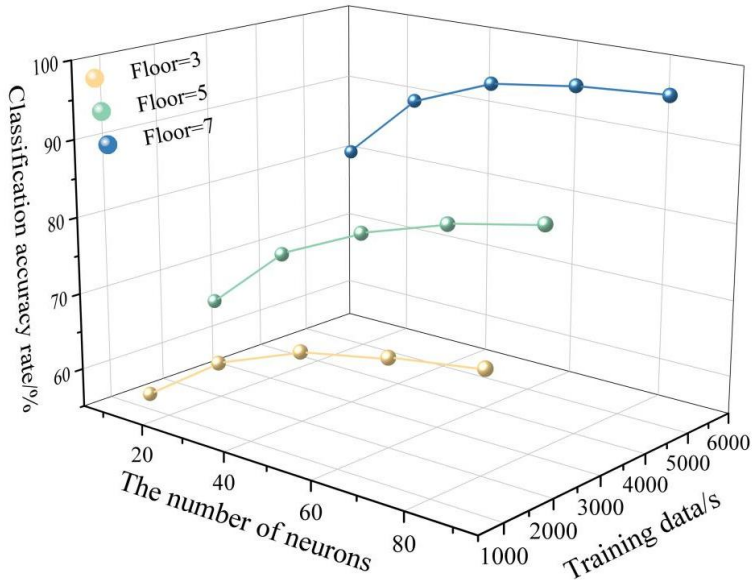

Figure 6: Specific simulation results

3.2 Results of Teaching Practice

Based on the optimal model parameters determined through prior simulation experiments, a music course teaching practice based on pattern recognition was designed. The objective of the implementation was to explore whether the application of artificial intelligence technology in music education could effectively enhance the effectiveness of music classroom teaching, stimulate students' classroom participation and their ability to actively learn music knowledge, and strengthen students' core music literacy. First-year students from a key university in a certain city were selected as the research subjects, with 200 students randomly selected (male-to-female ratio of 1:1.2), and the experimental period was 16 weeks. Prior to the implementation of the teaching program, the research team, in conjunction with expert consultation, determined the application scheme for the emotion recognition teaching model based on the BP neural network. This scheme organically integrates music emotion feature analysis technology with traditional teaching methods. Data collection was conducted using a double-blind method, with a third-party educational assessment institution completing standardized questionnaires within one week after the course concluded. The questionnaires included seven Likert five-point scale questions and four attitude measurement items, with a 100% response rate.

The results of the student evaluation survey on teaching methods are shown in Table 2. The

analysis results indicate that the music teaching model based on pattern recognition has gained widespread recognition among students. 94.5% of students expressed interest in the teaching content, 96% of students believed that model-assisted learning was beneficial for classroom learning, and 98% of students were attracted to this teaching model. In terms of classroom engagement, 99% of students demonstrated a willingness to engage in active thinking, and 94% participated in classroom activities. Additionally, 87% of students actively used the model for learning after class, and all students improved their interest and confidence in music learning through the teaching practice. These data fully demonstrate the effectiveness of AI-assisted teaching in enhancing student engagement and motivating learning.

Table 2: Students' Evaluations of the teaching links

Question	Option	Frequency	Percentage/%	Cumulative percentage/%
Q1.I am more likely to be interested in the teaching content of the teacher.	A.Quite agree	68	34	34
	B.Agree	121	60.5	94.5
	C.General	11	5.5	100
	D.Disagree	0	0	100
Q2.Adopting model assistance is very helpful for my learning in class.	A.Quite agree	65	32.5	32.5
	B.Agree	127	63.5	96
	C.General	7	3.5	99.5
	D.Disagree	1	0.5	100
Q3.This teaching mode attracts me very much.	A.Quite agree	63	31.5	31.5
	B.Agree	133	66.5	98
	C.General	4	2	100
	D.Disagree	0	0	100
Q4.I am willing to think actively in class.	A.Quite agree	58	29	29
	B.Agree	140	70	99
	C.General	2	1	100
	D.Disagree	0	0	0
Q5.I actively participated in the activities in class.	A.Quite agree	33	16.5	16.5
	B.Agree	155	77.5	94
	C.General	8	4	98
	D.Disagree	4	2	100
Q6.After class,I will take the initiative to use the model for learning	A.Quite agree	49	24.5	24.5
	B.Agree	125	62.5	87
	C.General	24	12	99
	D.Disagree	2	1	100
Q7.Through teaching practice, my interest and confidence in learning music have been enhanced.	A.Quite agree	87	43.5	43.5
	B.Agree	113	56.5	100
	C.General	0	0	100
	D.Disagree	0	0	100

The results of the survey on students' attitudes toward teaching methods are shown in Table 3. Regarding students' attitudes toward teaching methods, 90% of students approve of AI-assisted teaching models. After learning, 96.5% of students feel that learning music is an enjoyable experience, 88.5% believe that music can enrich their emotions, and 99.5% can discern the emotional feelings evoked by music. From the above four questions, it is evident

that students have undergone a significant shift in their attitudes toward music learning. In the classroom, the focus is not merely on imparting knowledge but on cultivating the habit of learning music, which is the most crucial aspect of aesthetic education. The purpose of music courses is to teach students to perceive, experience, and practice artistic language in their daily lives. This also demonstrates that AI-assisted teaching models can be effectively integrated into music education.

Table 3: Students' Attitudes towards teaching methods

Question	Option	Frequency	Percentage/%	Cumulative percentage/%
Q8.Compared with the previous classroom teaching mode, I prefer the current teaching mode.	A.Quite agree	67	33.5	33.5
	B.Agree	113	56.5	90
	C.General	17	8.5	98.5
	D.Disagree	3	1.5	100
Q9.I think learning music is an easy thing.	A.Quite agree	78	39	39
	B.Agree	115	57.5	96.5
	C.General	7	3.5	100
	D.Disagree	0	0	0
Q10.Learning music can comprehensively enrich my emotions.	A.Quite agree	49	24.5	24.5
	B.Agree	128	64	88.5
	C.General	20	10	98.5
	D.Disagree	3	1.5	100
Q11.When I hear music, I can better distinguish and feel the emotions expressed by the music.	A.Quite agree	83	41.5	41.5
	B.Agree	116	58	99.5
	C.General	1	0.5	100
	D.Disagree	0	0	0

4 Conclusion

This paper designs an automatic music emotion recognition model based on a BP neural network and conducts music teaching practices based on pattern recognition.

In the simulation experiments, the amplitude of changes in the average pitch curve was small, while the amplitude of changes in music tempo and average dynamics was significant. This indicates that average pitch features have a limited impact on emotion recognition, whereas music tempo and average dynamics play a significant role in emotion recognition. By balancing training time and accuracy, the optimal model parameters were determined to be 7 hidden layers and 70 neurons.

Through teaching practice, 94.5% of students expressed interest in the teaching content, 96% of students believed that model-assisted learning was helpful for classroom learning, and 98% of students were attracted to this teaching model. In terms of classroom participation, 99% of students demonstrated a willingness to think actively, and 94% of students participated in classroom activities. Additionally, 87% of students actively used the model for learning after class, and all students improved their interest and confidence in music learning through teaching practice. Regarding students' attitudes toward teaching methods, 90% of students approved of the AI-assisted teaching model. After learning, 96.5% of students felt that learning music was an enjoyable experience, 88.5% believed music could enrich their emotions, and 99.5% could discern the emotional feelings evoked by music. This demonstrates that a music education model based on pattern recognition is feasible.

References

[1] Feng, Y. (2025). Music Performance and Emotional Expression: A Philosophical Exploration of Emotion in Music Education. Mediterranean Archaeology and Archaeometry.

[2] Ngo, T., & Spreadborough, K. (2022). Exploring a systemic functional semiotics approach to understanding emotional expression in singing performance: Implications for music education. Research studies in music education, 44(3), 451-474.

[3] Zorkeply, N. S., & Zulkifli, T. E. T. (2022). Pop Music and its Role as a Communicative Medium to Express Emotions among Youth. International Journal of Academic Research in Business and Social Sciences, 12(11), 1199-1214.

[4] Imbir, K., & Gołąb, M. (2017). Affective reactions to music: Norms for 120 excerpts of modern and classical music. Psychology of Music, 45(3), 432-449.

[5] Váradi, J., Szűcs, T., Kerekes, R., Kiss, J., & Radócz, J. M. (2024). A Systematic Review on the Emotional Dimensions of Music Education. Harmonia: Journal of Arts Research and Education, 24(2), 236-246.

[6] Nakamura, A. (2020). Evaluation of Emotional Aspects of Students in Music Lessons Based on the TAS Model. International Journal of Creativity in Music Education, 7, 31-47.

[7] Commodari, E., & Sole, J. (2020). Music education in junior high school: Perception of emotions conveyed by music and mental imagery in students who attend the standard or musical curriculum. Psychology of Music, 48(6), 824-835.

[8] Peng, P., & Fu, W. (2022). A pattern recognition method of personalized adaptive learning in online education. Mobile Networks and Applications, 27(3), 1186-1198.

[9] Bharadwaj, Prakash, K. B., & Kanagachidambaresan, G. R. (2021). Pattern recognition and machine learning. Programming with tensorflow: Solution for edge computing applications, 105-144.

[10] Alluri, K. R., Achanta, S., Prasath, R., Gangashetty, S. V., & Vuppala, A. K. (2017). A study on text-independent speaker recognition systems in emotional conditions using different pattern recognition models. In Mining Intelligence and Knowledge Exploration: 4th International Conference, MIKE 2016, Mexico City, Mexico, November 13-19, 2016, Revised Selected Papers 4 (pp. 66-73). Springer International Publishing.

[11] Fagan, J. F. (2017). The origins of facial pattern recognition. Psychological development from infancy, 83-113.

[12] Shah, S. J. H., Albishri, A., Kang, S. S., Lee, Y., Sponheim, S. R., & Shim, M. (2023). ETSNet: A deep neural network for EEG-based temporal–spatial pattern recognition in psychiatric disorder and emotional distress classification. Computers in Biology and Medicine, 158, 106857.

[13] Chen, W. (2022). A novel long short-term memory network model for multimodal music

emotion analysis in affective computing. Journal of Applied Science and Engineering, 26(3), 367-376.

[14] Sun, B. (2022). Emotional analysis and personalized recommendation analysis in music performance. Scientific Programming, 2022(1), 9548486.

[15] Jandaghian, M., Setayeshi, S., Razzazi, F., & Sharifi, A. (2023). Music emotion recognition based on a modified brain emotional learning model. Multimedia Tools and Applications, 82(17), 26037-26061.

[16] Su, J., & Zhou, P. (2024). Quantitative physics–physiology relationship modeling of human emotional response to Shu music. Frontiers in Psychology, 15, 1351058.

[17] Vishnumolakala, S. K., Vallamkonda, V. S., Subheesh, N. P., & Ali, J. (2023, May). In-class student emotion and engagement detection system (iSEEDS): an AI-based approach for responsive teaching. In 2023 IEEE Global Engineering Education Conference (EDUCON) (pp. 1-5). IEEE.

[18] Yi, X. (2025). Analysis of students' emotion based on visual clues in online music teaching. Journal of Computational Methods in Sciences and Engineering, 14727978251341493.