

## Research Article

# Music Recognition and Classification Algorithm considering Audio Emotion

Wang Na and Fang Yong 

*School of Music, Beihua University, Jilin City 132013, Jilin Province, China*

Correspondence should be addressed to Fang Yong; [jl\\_fy2855@beihua.edu.cn](mailto:jl_fy2855@beihua.edu.cn)

Received 28 October 2021; Revised 8 December 2021; Accepted 21 December 2021; Published 20 January 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Wang Na and Fang Yong. 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.

At present, the existing music classification and recognition algorithms have the problem of low accuracy. Therefore, this paper proposes a music recognition and classification algorithm considering the characteristics of audio emotion. Firstly, the emotional features of music are extracted from the feedforward neural network and parameterized with the mean square deviation. Gradient descent learning algorithm is used to train audio emotion features. The neural network models of input layer, output layer, and hidden layer are established to realize the classification and recognition of music emotion. Experimental results show that the algorithm has good effect on music emotion classification. The data stream driven by the algorithm is higher than 55 MBs, the anti-attack ability is 91%, the data integrity is 83%, the average accuracy is 85%, and it has good effectiveness and feasibility.

## 1. Introduction

In the field of music information retrieval, affective recognition and classification have been widely concerned. Emotional classification of music is a multidisciplinary field of research. It perceives emotions from songs and classifies them into specific emotional categories. Emotional recognition depends on acoustic features, linguistic features, and metadata-based filtering [1]. Genre of music can be classified to describe the type and characteristics of music, and music characteristics can be used to construct more digital form of music, so it is of great significance to music information retrieval. In the past classification methods of audio types, music emotion classification has always been performed manually, and music types have been identified by artificial emotion. However, this method will cause subjective evaluation errors due to the variety of instrumental music equipment and music rhythm structure and form related to statistical types. Therefore, it is necessary to describe the classification of audio signals with the help of automatic emotion classification algorithm [2]. The performance of these musical features can be evaluated by training statistical pattern recognition. A multi-type user interface browsing

audio collection system based on automated classification has been developed to fully enhance human interaction. In the process of life and work, music emotion classification is mostly used in mobile music stores, radio stations, the Internet, and other terminals. Through the description of music emotion, it can effectively select the music type suitable for the current play [3, 4]. Although there may be some subjectivity and arbitrariness in the emotional classification of music, it is possible to classify music according to instrument type and rhythmic structure. However, due to the different subjective initiative and evaluation standards of individuals, there will be different classification standards for the same piece of music. Therefore, it is necessary to make use of the information processing ability of network big data to classify music types through specific templates and expand the music repertoire of the system database. In the previous analysis, it is found that automatic music type classifier can be used to improve the ability of music emotion recognition and classification by automatically indexing and retrieving audio data. The classification behavior is performed by selecting the frequency band that should be attenuated or re-attenuated according to the automatic music type classifier, through the type label assigned to the signal.

Music emotion classification can effectively make up for the lack of retrieval based on lyrics, singer names, and other ways and has become a research hotspot in recent years. Because multimedia information data cannot be directly matched with human emotion perception, it is necessary to analyze the relevance of music information data to human emotion from its own characteristics [5, 6]. At present, music emotion classification and recognition are mainly based on audio signals and lyrics, which are two different forms of data features [7]. The relationship between audio signals and music emotion is analyzed by machine learning method, and the characteristics of music rhythm and rhythm are discovered [8–10], so as to achieve the effective recognition of music emotion.

The essence of music is the latent emotion, and the recognition of music emotion is satisfied according to the mapping relationship between music characteristics and emotion, which is an effective promotion of music emotion recognition and can strengthen the effect of human-computer emotional interaction [4, 11]. At present, the classification and recognition of music emotion is mainly based on the analysis of data features in the form of music text and audio signal. Chaudhary et al. [2] proposed a spectral approach to transform musical features into visual representations using convolution neural networks, using CNN to extract music signal from the song as a specific feature, and then through the comparison of features of the song emotional classification. Quasim et al. [10], combining emotional maturity with existing Internet of Things systems, proposed an affection-based music recommendation and classification framework that allows for high precision classification of songs with memories and emotions. Chen et al. [12] aiming at the deficiency of single network classification model proposed a multi-feature combination network classifier based on CNN-LSTM to improve the effect of music emotion classification. By using convolution kernels to extract 2D musical information and using serialization to output single modal emotion classification, the heterogeneous features are effectively combined to improve the classification performance. Rajesh et al. [13] believed that music is one of the effective media to convey emotion, so the emotion recognition in music fragment is of great significance to the understanding of music. Therefore, a classification recognition method of music emotion based on deep learning technology is proposed. The Mel frequency cepstrum coefficient, chrominance energy normalization statistic, and chrominance short time Fourier transform are extracted from the musical instrument data set, and the neural network is trained to recognize emotion. Wang D et al. [1] using intelligent recognition and classification algorithm of human emotion studied different emotion classifications under different audio signals. In Haridas AV et al. [14] through the Taylor Series Deep Belief Network emotion recognition system, the noise in audio signal is removed, and then emotional information feature is extracted. Based on the above research, this paper presents a new algorithm of music emotion recognition and classification based on forward neural network, which is used to classify the lyrics and rhythm. The parameterized audio

emotion feature is trained by mean square error and gradient descent algorithm, and the neural network is constructed by input layer, output layer, and hidden layer to recognize and classify music emotion feature. Experimental results show that the proposed method has strong classification accuracy and can effectively identify the multiple classification criteria and emotional complexity of musical emotion.

## 2. Music Emotion Feature Extraction

A piece of music mainly includes audio information and lyrics text information, and audio information includes time domain features, frequency domain features, and cepstrum features. Time-domain features usually refer to the time-domain parameters calculated in a fixed length of music, mainly including short-time energy, short-time average amplitude, and average amplitude difference function. Frequency-domain features refer to the characteristic parameters obtained by converting time-domain signals into frequency-domain signals through Fourier transform, mainly including spectral centroid, spectral rollover, and spectral flux. Cepstrum features include audio information and lyrics text information, and audio information includes time domain, frequency domain, and cepstrum features by using different loudness and tones of human ears. Time-domain features usually refer to the time-domain parameters calculated in a fixed length of music, mainly including short-time energy, short-time average amplitude, and average amplitude difference function. Frequency-domain features refer to the characteristic parameters obtained by converting time-domain signals into frequency-domain signals through Fourier transform, mainly including spectral centroid, spectral rollover, and spectral flux. Cepstrum features, which use the human ear on the sound of different loudness, tone, and timbre, have different perceptual effects. Therefore, it can be used to classify different musical emotions. The specific operation methods are as follows:

- (1) The music is divided into specific lengths and transformed into frequency domain signals by Fourier transform.
- (2) Calculating the logarithm of the signal in frequency domain and then inverse Fourier transform, the cepstrum characteristic of the music can be extracted.

In addition to the rhythmic characteristics of music, audio messages also imply certain emotions.

The characteristics of music can be divided into basic features, complex features, and overall features [15], as shown in Figure 1. The basic characteristics of music include the following: the complex characteristics of music include rhythm, melody, harmony, etc. The overall characteristics include musical form structure, musical style, emotional connotation, etc. [16]. Accordingly, the recognition of musical features can be divided into three levels: the first is to extract the basic characteristics of music, the second is to get the complex characteristics based on the analysis of basic characteristics, and the last is to identify the overall characteristics of music according to the basic and complex

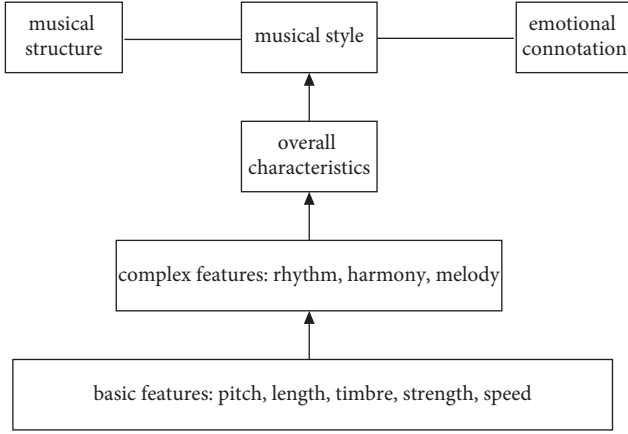


FIGURE 1: Composition of musical emotional characteristics.

characteristics. Firstly, the basic features of music are extracted and the complex features of music are calculated by using the feedforward neural network. Then, the emotion space is established according to the identified audio emotion features, and the emotion recognition model is constructed by using the emotion feature space to complete the classification of music emotion.

### 2.1. Overall Steps of Music Emotion Classification.

Through constructing the music emotion classifier, complete the music emotion concrete classification. Using FNN [17] as the music intelligent classifier, the FNN can realize the multi-feature fusion of music emotion appreciation effectively by forming multi-modal data from different modal data.

The process of music emotion classification is as follows: a record is randomly selected from the database to find the similar information in the database, and the result is fed back to the system in the form of annotated record. Using these annotation records, the system trains the music emotion classifier until the music emotion segment search error reaches a controllable range. Then, the music emotion is labeled in the classifier.

The steps of music emotion classification are as follows:

Step 1. Extract feature vectors  $d_1, d_2, \dots, d_n$  from  $n$  music fragments in the library, where  $d_i = \{d_{i1}, d_{i2}, \dots, d_{im}\}$ .

Because the physical meanings of the  $m$  components in  $d_i$  are different, we use Gaussian function [18] to normalize them:

$$d_i = \frac{1}{2} \times \frac{d_i - u_i}{e_i}, \quad 1 \leq i \leq m. \quad (1)$$

In the formula,  $u_i$  is the mean and  $e_i$  is the standard deviation,  $m = 25$ .

Step 2. The  $d_1$  of one of the  $n$  music segments is randomly selected and its characteristic is vector  $d_{i1}$ . Calculate other music fragments in the library from  $d_1$ :

$$d = \sqrt{(d_i - e_i)}. \quad (2)$$

Return the distance calculation result to the system in turn.

Step 3. The system judges these results, marks the information with similar feelings to the system, and marks them in the database.

Step 4. In order to train music emotion classifier, the feature of music segment marked as similar emotion information in database is used as training sample.

Step 5. The trained music emotion classifier is used to classify the unlabeled music segments in the database and transmit the same type of emotion information to the user. The user judges the same type of emotion segment. If all of these segments are similar emotion segments, then after marking the segment in the database, the segment will jump to Step 6, and if the segment is different emotion segment, the segment will jump to Step 3.

Step 6. Update the tagged music fragment information to the new music library.

Step 7. Return to Step 1 and loop through the steps of Step 1 and Step 6 until all the songs in the database are categorized.

### 2.2. Audio Emotion Feature Parameterization.

Musical features include pitch, duration and timbre, rhythm, melody and speed, as well as musical form, mode, and other semantic features. Music emotion analysis needs to consider music basic information, rhythm change, and structure form, etc. Music characteristics are divided into musical note characteristics and high-level characteristics [19]. Note features include pitch, duration, and intensity, while high-level features include speed, power, rhythm, melody, and interval information [20]. Pitch, duration, and timbre are the acoustic cues of musical emotion cognition. Therefore, they are regarded as the most basic components of musical emotion characteristics and are characterized by the mean square deviation [21]:

$$F_i = \sqrt{\frac{1}{x} \sum_{p=1}^x L_i}. \quad (3)$$

In the formula,  $L_i$  denotes the pitch of the  $i$  and  $x$  denotes the number of notes. Intensity is an important means of expressing emotion in music [22]. Different musical intensity can create different emotional experience for listeners. Therefore, the intensity can be used to express the pitch breadth of the music, and the average intensity and intensity change characteristics can be used to represent the music emotion. The intensity characteristics can be extracted as follows:

$$S_i = \left( f_i \times \frac{v}{G} \right) \times t. \quad (4)$$

In the formula,  $f_i$  represents the strength of the  $i$  note;  $G$  represents the total number of notes;  $v$  denotes pitch; and  $t$  stands for time. Through formula (4) to describe the

dynamics change in the form of music sections, we can effectively exclude the influence of light and light changes in the rhythm on the parameterization of emotional characteristics.

Melody is an organization and harmonious movement formed according to certain music rules, which can effectively reflect the organization form of music in time and space. The direction of the melody describes the variation in pitch. The direction of the melody can be expressed as follows:

$$R_i = \sum_{i=1}^m \frac{S_i}{H - h_i}. \quad (5)$$

In the formula,  $h_i$  stands for the  $i$  length and  $H$  stands for the sum of the tones. Rhythm is a regular intensity phenomenon that occurs alternately in music [23]. Different rhythms give different strains to the music. According to the tension change of rhythmic movement, the musical characteristics are extracted by using the density of pronunciation point [24] to reflect the basic state of music. The greater the density of the point of articulation, the stronger the intensity, and the more obvious the characteristics. The density of the rhythm is

$$Q_i = \frac{W_i - 1}{W_{\text{Maxi}}}. \quad (6)$$

In the formula,  $W_i$  means the energy value of the subbar of  $i$  and  $W_{\text{Maxi}}$  means the maximum energy value of the bar.

**2.3. Establishment of an Audio Emotion Feature Extraction Model.** Emotion is the essential characteristic of music. The mathematical model of music emotion must be based on the research of its psychological model. From the perspective of information theory, the whole process of music psychological feeling is actually a process of information acquisition, transformation, transmission, processing, and storage [25], while from the perspective of cybernetics, every composer, performer, or listener can be regarded as a biological steady-state system, which has emotional sensation and self-control behavior. Forward neural network is a kind of neural network model imitating human brain, so it can be constructed by imitating human brain to recognize musical emotion.

Through the feature extraction model of audio emotion, any 10-second music fragment  $j$  can be identified emotionally. Formula (7) is used to calculate the sum of the squares of the difference between the eigenvalues of  $j$  and the average value of the music features of each class, and then the qualification vector value of  $j$  in each class is calculated by formula (8), and then the  $j$  is classified into the music features with the largest qualification vector value.

$$d_{j,l} = \sum_{b=1}^{25} (j_b - t(l, b)). \quad (7)$$

$d_{j,l}$  denotes the proximity of  $j$  to  $l$ , and  $j_b$  denotes the  $b$  eigenvalue of  $j$ .

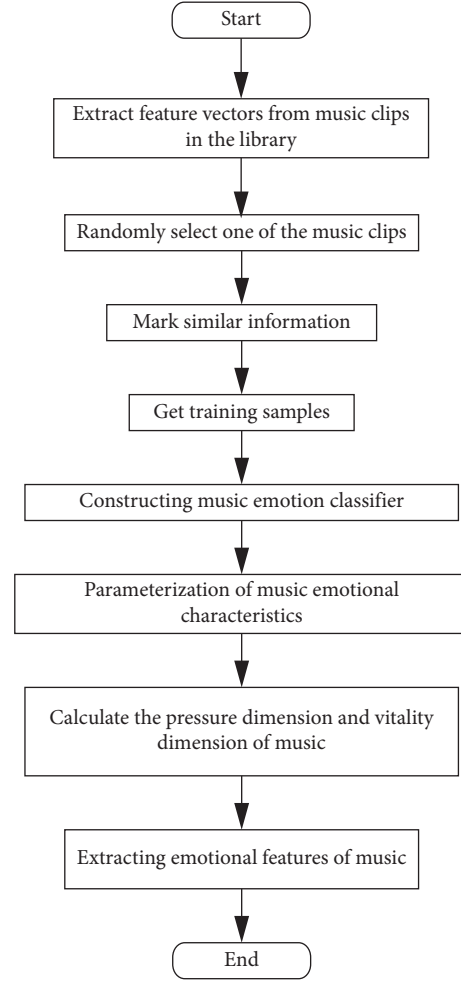


FIGURE 2: Flowchart of music emotion feature extraction.

$$g_{j,l} = \frac{d_{j,l} \times K}{\sum_{b=1}^{25} d_{j,b}}. \quad (8)$$

$g_{j,l}$  denotes the eligibility vector of  $j$  in  $l$ -like, and  $K$  denotes the ambiguity.

**2.4. Emotional Feature Extraction of Whole Music.** In order to sympathize with the listener and let the listener feel certain emotion, the composer must exaggerate certain emotion through different emotion to make certain emotion express more intensely, so there are different emotion changes in the same piece of music.

Firstly, the whole piece of music is divided into several segments, each segment is 10 seconds, and there is a 3-second repetition between two consecutive segments; then, the FNM method is used to calculate the four qualification vectors  $g_{j,l}$  for each segment, formula (9) is used to calculate the value of the stress dimension, and formula (10) is used to calculate the value of the vitality dimension, so as to complete the emotional feature extraction of the whole piece of music.

$$j_w = g_{j1} + g_{j2} + g_{j3} + g_{j4}. \quad (9)$$

$$j_i = g_{j4} - g_{j2} - g_{j3} + g_{j1}. \quad (10)$$



Flowchart of music emotion feature extraction is shown in Figure 2.

### 3. Music Emotion Recognition Classification Algorithm Design

**3.1. Forward Neural Network Architecture.** Forward neural networks are multilayer neural networks that can be used to solve nonlinear problems [26, 27]. FNN is usually composed of input layer, output layer, and hidden layer. The structure of input layer is received by the hidden layer, and the result is transferred to the output layer as output of the whole network. When the actual output is inconsistent with the expected output, the calculated error is transmitted to the input layer as input feedback, and the weights of each layer in the network model are modified by the error gradient descent method. Aiming at the feature vector data of audio and lyrics, this paper improves the traditional feedforward neural network. Based on the special Chebyshev orthogonal polynomial clusters, a forward neural network model with special structure is constructed. In this model, a single hidden layer is used to reduce the complexity of the whole model. The stimulation function of each neuron in the hidden layer uses each function in Chebyshev's orthogonal polynomial cluster, and the stimulation function of neurons in other layers of the model uses linear stimulation function.

Assuming that the value of each neuron in the network is  $a_j$ , the layer and interlayer elements are connected with different weights, and the activation value  $b_i$  obtained by nonlinear activation function [28] is used to calculate the data of the upper layer, and the obtained results are transferred to the data of the lower layer to realize the forward calculation. Forward neural network model frame diagram is shown in Figure 3.

In the expression,  $I$  represents the input of the neural network,  $P$  represents the output of the neural network, and  $b_1$  and  $b_2$  are the weights of the input and output layers.

**3.2. Neural Network Propagation Formula.** The propagation of the neural network includes forward propagation and back-propagation [29]; forward propagation refers to the propagation of the model from the bottom up, calculated based on a given input value; back-propagation calculates losses based on forward propagation calculations and then calculates and trains each neuron parameter using a gradient descent algorithm [30].

The forward propagation formula of a neural network can be represented by a set of recurrent formulas, namely,

$$a_r^t = \sum_{i=1}^{25} q_i j_r^t + b_r^t. \quad (11)$$

In the formula, the input of the  $r$  neuron hiding layer at the time of  $t$  is marked as  $a_r^t$ , the output of the  $r$  neuron hiding layer at the time of  $t$  is marked as  $b_r^t$ , FF, the input of the  $r$  neuron is marked as  $j_r^t$ , and  $q_i$  represents the parameter weights of the input layer and the hidden layer, thus

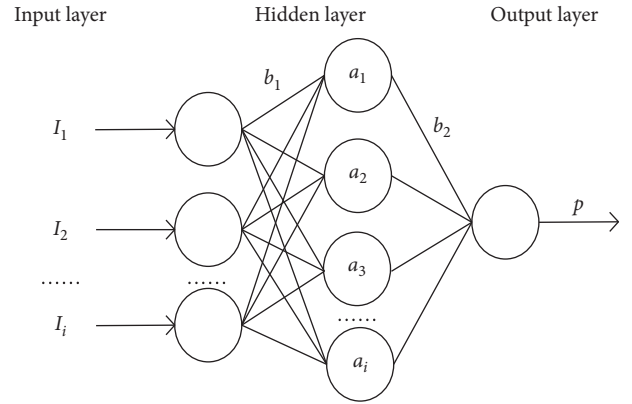


FIGURE 3: Structure diagram of forward neural network.

obtaining the specific forward propagation results of the neural network.

**3.3. Softmax Activation Function.** Neural networks are trained using Softmax as an activation function [31]. Softmax has the advantage of being computationally convenient and having a simple form of results for multiple classification problems. Because the output of each neuron of Softmax is positive and the sum is 1, the output of Softmax layer can be regarded as a probability distribution, and a more intuitive result can be obtained.

Assuming that the output of Softmax is  $f(c)$ , the input is marked  $c$ , and the  $k$  element in the array  $c$  is represented by an  $c_k$ , the formula for determining Softmax is

$$f(c) = \frac{e^c}{\sum_{k=1}^{25} e^{c_k}}. \quad (12)$$

Softmax's loss function is

$$\text{loss} = -\lg f(P_k). \quad (13)$$

In the expression,  $P_k$  is the correct input to the sample, but since there is no single data input at the time of training; assuming the input is  $f(c) = e^c / \sum_{k=1}^{25} e^{c_k}$ , there is  $\text{loss} = \log \sum_{i=1}^{25} e^c - c_k$ . In the case of applying the chain rule [32], the main steps of solving the partial derivative of loss from the weight matrix are as follows:

- (1) Calculating the activation value of each node in the network.
- (2) The gradient is propagated through the direction propagation algorithm to obtain the gradient values of each parameter.
- (3) Adopt gradient descent algorithm to update model parameters.
- (4) Iterate the process until it converges.

**3.4. Overall Affective Recognition Framework.** Music emotion recognition based on forward neural networks includes the following flow, as shown in Figure 4.

- (1) Characteristics of learning samples: through controlling the forward neural model, the thresholds of

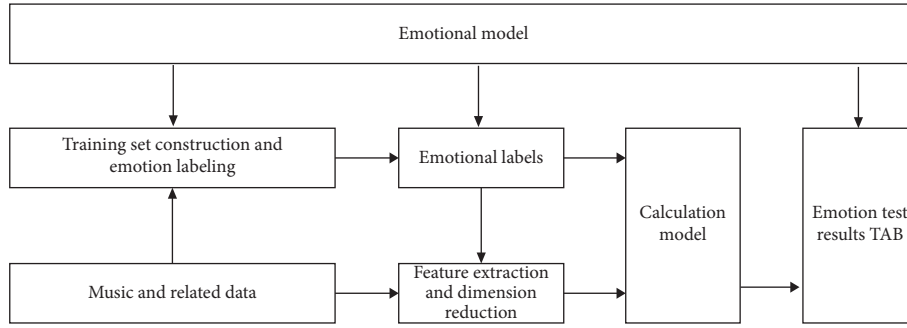


FIGURE 4: Music emotion recognition system framework.

neurons and the connection weights between neurons in the forward neural network model are iterated under the standard of sample data. Under the constraint of objective function minimization, the FNN can learn the features of sample data by a large margin.

- (2) Music emotion is classified in combination with relevant music data and information, and a type data set is constructed to label emotion. The emotion classification process proposed in this paper is divided into two parts: model generation and emotion classification. The music emotion classifier is obtained by training the classifier according to the features of training samples, and the emotion classifier classifies the input music segments according to the emotion classifier.
- (2) The preprocessing of audio information mainly includes signal filtering, etc. The preprocessing of lyrics text is the operation of the vocabulary of word emotion recognition and the addition of similar words. The feature extraction and dimensionality reduction are carried out for the data set, and the neural network model is established.
- (3) Input training sample set, training output value, calculation error, and reverse update model parameters until the error is less than the expected error.
- (4) Enter the test samples, conduct the test, and obtain the label of the test results.
- (5) Judge whether the requirements for precision and iteration times are met. If they are satisfied, the model can be directly used to test the discrimination of musical emotion types; if not, it can increase the number of iterations and adjust the parameters of the network model.

## 4. Experimental Section

This experiment is a simulation experiment, the experimental software platform is MATLAB 2020b; the proposed method will be input into the simulation software to achieve operation, set the parameters of the experimental environment, and get the test results of the proposed method. The experimental indices include training by sampling,

training time analysis of FNN model, accuracy analysis of emotion classification by FNN model, transmission rate analysis driven by FNN model, anti-attack analysis of FNN model, data integrity analysis of FNN model, and accuracy analysis of different music emotion recognition.

**4.1. Experimental Environment.** Collect 300 multitrack MIDI files on YouTube through Web Crawler [33]. Perform with a variety of instruments. Play music in a variety of styles with accurate emotional descriptions. Randomly selected music files were divided into four categories according to the emotional labels at the time of download: happy, anxious, calm, and depressed. Data were divided into five categories according to the specific needs of the experiment: provocative, happy and painful, happy, humorous, and fanatical, as shown in Tables 1 and 2.

### 4.2. Simulation Results

**4.2.1. Sample Training.** Then, 1,000 volunteers of different ages, genders, and professions were randomly selected to subjectively categorize the music in the sample library. If more than 50% of people categorized the same piece of music into a certain category, the music was used as a sample. If 40% of people chose one type of emotion and 30% chose another, the piece was considered an invalid sample. The average of all the features in each category is calculated as follows:

$$R(o, e) = \frac{1}{N_o} \sum_{e=1}^{25} H_{o,e,n}. \quad (14)$$

Among them,  $R(o, e)$  represents the average value of  $e$  eigenvalue in  $o = 1, 2, 3, 4, 5$  in  $o$ -like,  $N_o$  represents the number of fragments in  $o$ -like, and  $H_{o,e,n}$  represents the  $e$  eigenvalue in  $o$ -like  $n$  fragment.

**4.2.2. Training Time Analysis of Forward Neural Network Model.** The training time of FNN model is not only related to the supervised algorithm, but also related to the number of layers of the model and the number of neurons in each layer. In this paper, 8 hidden layers and 60 neurons are used as experimental objects to compare and analyze the training

TABLE 1: Data emotions divided into four categories.

	Category 1(exuberance)	Category 2(anxious)	Category 3(contentment)	Category 4(depression)
Quantity	76	69	83	72
Total	300			

TABLE 2: Data emotions divided into five categories.

	Category 1(aggressive)	Category 2(bittersweet)	Category 3(happy)	Category 4(humorous)	Category 5(passionate)
Quantity	67	62	54	58	59
Total	300				

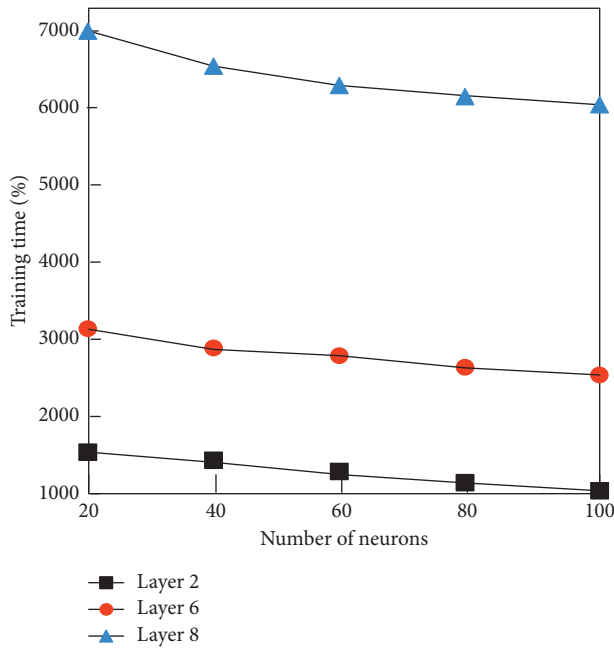


FIGURE 5: Training time of forward neural network model.

time of FNN model. The simulation results are shown in Figure 5.

Figure 5 shows that the training time is inversely proportional to the number of layers and the number of neurons. The average training time is 6400s when the number of layers increases to 8. But with the increase of the number of neurons, the training time will be reduced gradually, the number of neural network layers is 8, the number of neurons is 100, and the training time is 6000 s. Therefore, in order to achieve a better training effect, we can reduce the training time and improve the efficiency by increasing the number of neurons in the forward neural network.

4.2.3. Accuracy Analysis of Emotion Classification Using Forward Neural Network Model. In the calculation of FNN model, the complexity of emotion classification is different due to the difference of layers and neurons, and the accuracy is different. The simulation result is shown in Figure 6.

Figure 6 shows that the increase in the number of layers and neurons will improve the network model’s learning

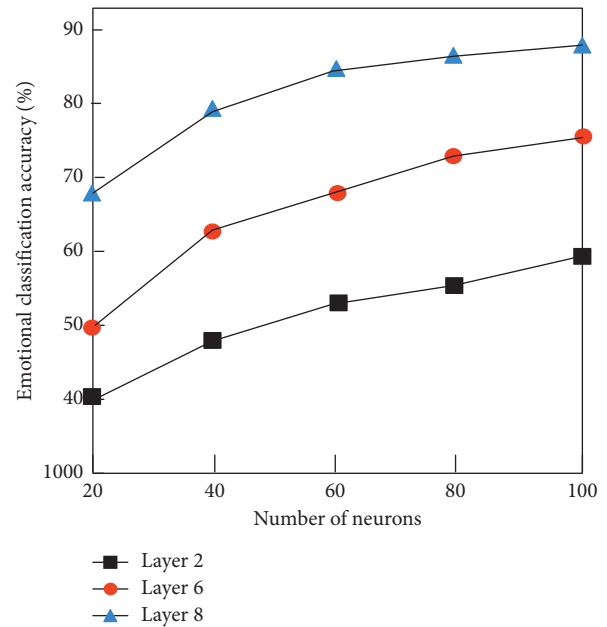


FIGURE 6: Emotional classification accuracy of forward neural network model.

ability, that is, improve the accuracy of classification. When the number of neural networks is 8 and the number of neurons is 100, the accuracy of emotion classification can reach 89%. This is because the neural parameters are trained and calculated by gradient descent algorithm, and the probability distribution is obtained by activation function, which improves the accuracy of emotion classification effectively.

4.2.4. Forward Neural Network Model Driven Transmission Rate Analysis. Data transmission rate is the number of data transmitted by data path in unit time. By analyzing data transmission rate, special bandwidth limitation can be prevented during high load, and the model can run safely and effectively. Therefore, the forward neural network model driven transmission rate is measured using 32 inputs and outputs as the standard. The result is shown in Figure 7.

Figure 7 shows that the data flow of each input and output of the application object is higher than 55MBS in the process of driving debugging, which shows that the

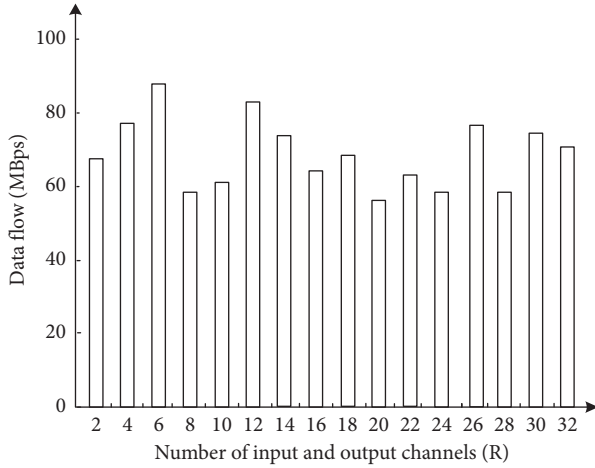


FIGURE 7: Forward neural network model driven transmission rate results.

proposed method has good performance in practical application and obvious advantages in application.

**4.2.5. Attack Resistance Analysis of Forward Neural Network Model.** In order to test the resistance of the forward neural network model, the attacker is set to perform brute force cracking and wormhole attack on the forward neural network model; the result is shown in Figure 8.

Figure 8 shows that the FNN model has the highest anti-attack ability, 91%, at the time of 20s, and the anti-attack ability is in the high anti-attack ability in different music emotion classification period. The results show that the FNN model has better security performance in the data transmission.

**4.2.6. Data Integrity Analysis of Forward Neural Network Model.** The larger the value of  $F$  is, the more complete the music segment feature data is. The formula is as follows:

$$F = \frac{d}{L} \times 100\%. \quad (15)$$

In the formula,  $d$  is the feature data extracted from the music segment in formula (1) and  $L$  is the overall feature data of the music. Using data integrity as a test metric, the results are shown in Figure 9.

Figure 9 shows that the  $F$  value is above 83 when using this method to extract the music fragment feature data. According to the test results, Gaussian function is used to normalize the feature vectors of music fragments, so the integrity of feature extraction data is improved.

**4.2.7. Analysis of the Accuracy of Different Music Emotion Recognition.** In order to verify the superiority of the proposed algorithm, the following simulation experiments focus on the accuracy of 300 music emotion segments; the specific simulation results are shown in Figure 10.

Figure 10 shows that the accuracy of music emotion recognition is influenced by the number of categories and

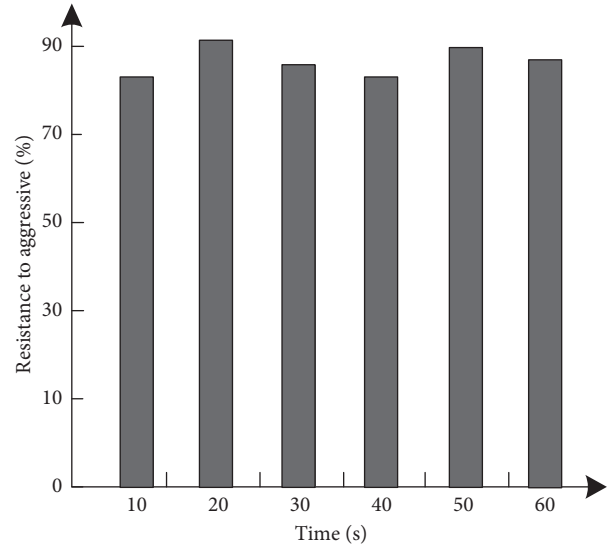


FIGURE 8: Results of forward-neural network model attack resistance.

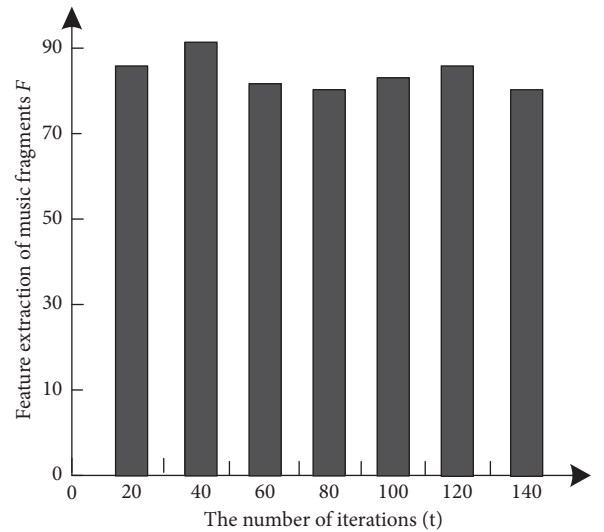


FIGURE 9: Data integrity test results of forward neural network model.

the complexity of music emotion. The accuracy of the tests is above 85%.

The above experimental results show that the training time of the feedforward neural network applied in this study is inversely proportional to the number of layers and neurons. When the number of layers increased to 8, the average training time was 6400 s. But as the number of neurons increases, the training time will gradually decrease. The number of neural network layers is 8, the number of neurons is 100, and the training time is 6000 seconds. Therefore, in order to achieve better training effect, we can reduce training time and improve training efficiency by increasing the number of neurons in the forward neural network. When the number of neural networks is 8 and the number of neurons is 100, the accuracy of emotion classification can reach 89%. The data stream of each input and



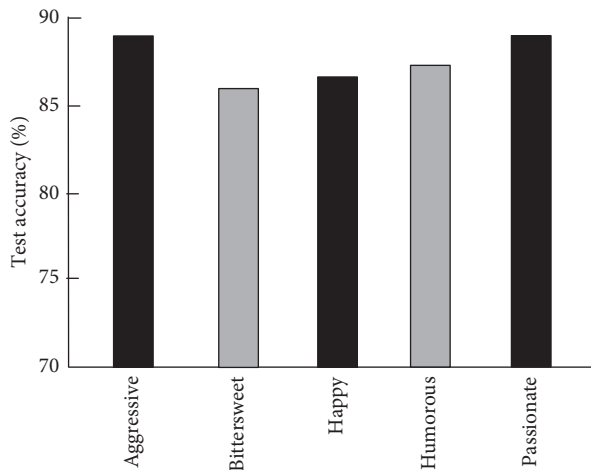


FIGURE 10: Musical emotion fragment test accuracy.

output of the application object is higher than 55MBS, which shows that the method has good performance and obvious application advantages in practical application. The FNN model has the highest anti-attack ability, up to 91%, and has the higher anti-attack ability in the different music emotion classification stage. The accuracy of music emotion recognition is affected by the number of music emotion categories and the complexity of music emotion, but the accuracy of this method is above 85%. The experimental data show that the proposed method has ideal application effect and provides a reliable basis for further research in related fields.

## 5. Conclusion

Music emotion cognition is an important part of multimedia content understanding and intelligent human-computer interaction, which has been widely used in human-computer interaction, entertainment robot, game industry, and music education. With the development of computer technology, the application of music emotion in stage performance art has become a new trend of the development of multimedia technology. Emotion recognition plays an unshakable role in music retrieval and data service, so it is of great practical significance to study emotion recognition. In this paper, we construct a forward neural network model to extract the musical emotion features and identify and classify them according to the characteristics of intensity, melody, and rhythm density. Through the experimental data, we can conclude that the forward neural network model has better use value, has higher accuracy for music emotion recognition, and has important significance for music emotion recognition classification.

## Data Availability

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

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding this work.

## Acknowledgments

This work was supported by Jilin Province Department of Education 13th Five-Year Plan, “A research on the Application of Music Therapy in Rehabilitation Education of Special Children” (JJKH20180407SK) and Jilin Province Department of Education 13th Five-Year Plan, “The Exploration and Research on Vocal Music Teaching Mode in Normal Universities in the New Era” (JJKH20170090SK).

## References

- [1] D. Wang and X. Guo, “Research on intelligent recognition and classification algorithm of music emotion in complex system of music performance,” *Complexity*, vol. 2021, no. 6, Article ID 4251827, 10 pages, 2021.
- [2] D. Chaudhary, N. P. Singh, and S. Singh, “Development of music emotion classification system using convolution neural network,” *International Journal of Speech Technology*, vol. 24, no. 2, pp. 1–10, 2020.
- [3] D. Zheng, “Music emotion recognition classification algorithm based on forward neural network,” *Journal of Information Technology*, vol. 43, no. 12, pp. 57–61, 2019.
- [4] T. Li, “Visual classification of music style transfer based on PSO-BP rating prediction model,” *Complexity*, vol. 2021, no. 6, Article ID 9959082, 9 pages, 2021.
- [5] S. Werner and G. N. Petrenko, “Speech emotion recognition: humans vs machines,” *Discourse*, vol. 5, no. 5, pp. 136–152, 2019.
- [6] A. Fh, B. Ma, and E. Smcd, “Continuous emotion recognition during music listening using EEG signals: a fuzzy parallel cascades model,” *Applied Soft Computing*, vol. 101, Article ID 107028, 2020.
- [7] S. Panwar, P. Rad, K.-K. R. Choo, and M. Roopaei, “Are you emotional or depressed? Learning about your emotional state from your music using machine learning,” *The Journal of Supercomputing*, vol. 75, no. 6, pp. 2986–3009, 2019.
- [8] T. Saha, D. Gupta, S. Saha, and P. Bhattacharyya, “Emotion aided dialogue act classification for task-independent conversations in a multi-modal framework,” *Cognitive Computation*, vol. 13, no. 3, pp. 1–13, 2021.
- [9] X. Liu, “An improved particle swarm optimization-powered adaptive classification and migration visualization for music style,” *Complexity*, vol. 24, no. 7, p. 0872, 2020.
- [10] M. T. Quasim, E. H. Alkhamash, M. A. Khan, and M. Hadjouni, “Emotion-based music recommendation and classification using machine learning with IoT Framework,” *Soft Computing*, vol. 25, no. 7, Article ID 12249, 2021.
- [11] S. Demircan and H. K. Örnek, “Comparison of the effects of mel coefficients and spectrogram images via deep learning in emotion classification,” *Traitement du Signal*, vol. 37, no. 1, pp. 51–57, 2020.
- [12] C. Chen and Q. Li, “A multimodal music emotion classification method based on multifeature combined network classifier,” *Mathematical Problems in Engineering*, vol. 2020, no. 8, 11 pages, Article ID 4606027, 2020.
- [13] S. Rajesh and N. J. Nalini, “Musical instrument emotion recognition using deep recurrent neural network,” *Procedia Computer Science*, vol. 167, pp. 16–25, 2020.
- [14] A. V. Haridas, R. Marimuthu, V. G. Sivakumar, and B. Chakraborty, “Emotion recognition of speech signal using Taylor series and deep belief network based classification,” *Evolutionary Intelligence*, no. 4, pp. 11–14, 2020.

- [15] S. Hizlisoy, S. Yildirim, and Z. Tufekci, "Music emotion recognition using convolutional long short term memory deep neural networks," *Engineering Science and Technology an International Journal*, vol. 24, no. 3, pp. 760–767, 2020.
- [16] D. Lee, L. Robinson, and D. Bawden, "Modeling the relationship between scientific and bibliographic classification for music," *Journal of the Association for Information Science and Technology*, vol. 70, no. 3, pp. 230–241, 2019.
- [17] S. Mohammed, C. Ramos, W. C. Fang, and T. H. Kim, "Emerging higher-level artificial neural network-based intelligent systems," *Neural Computing and Applications*, vol. 33, no. 3, pp. 4595–4597, 2021.
- [18] G. Wang, C. Lopez-Molina, and B. Debaets, "Multiscale edge detection using first-order derivative of anisotropic Gaussian kernels," *Journal of Mathematical Imaging and Vision*, vol. 61, no. 5, pp. 1096–1111, 2019.
- [19] M. Kaliakatsos-Papakostas and E. Cambouropoulos, "Conceptual blending of high-level features and data-driven salience computation in melodic generation," *Cognitive Systems Research*, vol. 58, pp. 55–70, 2019.
- [20] A. Vilorio, O. B. Pineda Lezama, and D. Cabrera, "Segmentation process and spectral characteristics in the determination of musical genres," *Procedia Computer Science*, vol. 175, pp. 96–101, 2020.
- [21] D. d. F. P. Melo, I. D. S. Fadigas, and H. B. d. B. Pereira, "Graph-based feature extraction: a new proposal to study the classification of music signals outside the time-frequency domain," *PLoS One*, vol. 15, no. 11, Article ID e0240915, 2020.
- [22] M. Jabreel and A. Moreno, "A deep learning-based approach for multi-label emotion classification in tweets," *Applied Sciences*, vol. 9, no. 6, p. 1123, 2019.
- [23] S. Ma, "Music rhythm detection algorithm based on multipath search and cluster Analysis," *Complexity*, vol. 2021, no. 4, pp. 1–9, Article ID 5627626, 2021.
- [24] K. Ohara, K. Saito, M. Kimura, and H. Motoda, "Critical node identification based on articulation point detection for uncertain network," *International Journal of Networks and Communications*, vol. 9, no. 2, pp. 201–216, 2019.
- [25] M. A. Khan and V. S. Patel, "Knowledge and approximations: a formal study under the perspective of information systems and rough set theory," *Information Sciences*, vol. 524, no. 5, pp. 97–115, 2020.
- [26] A. Benavoli, A. Facchini, D. Piga, and M. Zaffalon, "Sum-of-squares for bounded rationality," *International Journal of Approximate Reasoning*, vol. 105, pp. 130–152, 2019.
- [27] A. González-Espinoza, G. Martínez-Mekler, and L. Lacasa, "Arrow of time across five centuries of classical music," *Physical Review Research*, vol. 2, no. 3, Article ID 33166, 2020.
- [28] S. Bharadwaj, S. Hegde, N. Dutt, and A. P. Rajan, "Nonlinear signal processing method detects emotional changes induced by Indian classical music," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 6200–6206, 2019.
- [29] Y. Zheng, "The use of deep learning algorithm and digital media art in all-media intelligent electronic music system," *PLoS One*, vol. 15, Article ID 0240492, 2020.
- [30] V. Bondarev, "Training a digital model of a deep spiking neural network using backpropagation," *E3S Web of Conferences*, vol. 224, no. 38, Article ID 01026, 2020.
- [31] S. Totaro, A. Hussain, and S. Scardapane, "A non-parametric softmax for improving neural attention in time-series forecasting," *Neurocomputing*, vol. 381, pp. 177–185, 2020.
- [32] R. Cushman and J. Niatycki, "On subcartesian spaces Leibniz' rule implies the chain rule[J]," *Canadian mathematical bulletin = Bulletin canadien de mathématiques*, vol. 63, no. 2, pp. 1–9, 2019.
- [33] D.-H. Han and Y.-K. Lee, "Design of action-based Web crawler structural configuration for multi-website management," *KIISE Transactions on Computing Practices*, vol. 27, no. 2, pp. 98–103, 2021.