

Research Article

Research on Intelligent Recognition and Classification Algorithm of Music Emotion in Complex System of Music Performance

Daliang Wang  and Xiaowen Guo

Music Academy, Henan Polytechnic, Zhengzhou 450046, China

Correspondence should be addressed to Daliang Wang; daliangwang@hnzj.edu.cn

Received 6 May 2021; Accepted 10 June 2021; Published 24 June 2021

Academic Editor: Huihua Chen

Copyright © 2021 Daliang Wang and Xiaowen Guo. 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.

In the complex system of music performance, there are differences in the expression of music emotions by listeners, so it is of great significance to study the classification of different emotions under different audio signals. In this paper, the research of human emotional intelligence recognition and classification algorithm in the complex system of music performance is proposed. Through the recognition of SVM, KNN, ANN, and ID3 classifiers, the accuracy of a single classifier is compared, and then the four classifiers are combined to compare the classification accuracy of audio signals before and after preprocessing. The results show that the accuracy of SVM and ANN fusion is the highest. Finally, recall and *F1* are comprehensively compared in the fusion algorithm, and the fusion classification effect of SVM and ANN is better than that of the algorithm model.

1. Introduction

Music performance is a comprehensive expression of emotional characteristics of music, behavior, emotion, and other factors, and it is also an important way to express human emotions. Because there are many expressions of personal emotions related to their own experiences in music emotions, music emotions deeply rooted in people's hearts can strengthen the musical expressions of human beings' control, suppression, and encouragement of their own emotions. In human emotional expression, classifying music emotions can distinguish different human beings from classifying emotional expression in music performance. Therefore, it is of great research significance to classify human emotions intelligently from music performance; to distinguish music classification in the music performance system more effectively, music emotion recognition is the process of using computer to recognize the emotion contained in music itself, that is, to extract the basic characteristic information of music from music and classify the music emotion information through the preset recognition

mode; it is a relatively cutting-edge artificial intelligence technology; music visualization is a design activity that aims at accurately and quickly conveying music content information and effectively realizing the value of music information, transforming music content information into visual elements for rearrangement and combination, and finally presenting music information in a visual way. The combination of music emotion recognition and visualization can help people reduce the cognitive burden of music understanding, thus promoting the application and development of music in many fields such as work analysis, auxiliary teaching, visual retrieval, game entertainment, and stage performance. According to the existing research situation, researchers at home and abroad have made some achievements in music emotion recognition and music visualization and put forward some feasible models or solutions. However, looking at the overall situation of this research field, the existing research still focuses on the technical level, while the theoretical exploration is still relatively lacking. Although it is necessary to study music emotion and its visualization technology, it is quite valuable to conduct interdisciplinary

research based on the field of information science and combine musicology, psychology, and mathematics and computer science. This paper will start from the perspective of information and carry out some innovative research on the recognition and visualization of music emotional information.

In the emotional analysis of music, it is necessary to digitize different data forms for further processing. Some scholars classify emotions by decomposing audio signals such as music rhythm and timbre, extracting acoustic features, and combining relevant algorithms. In view of the music performance recommendation system, human emotions are not fully considered for classification research. The traditional music performance system is basically based on the user's playing history and playlist or aims at the characteristics of music performance, such as sound quality and rhythm. Calculating the distance of different performing music, using the basic information of music genre, singer, lyrics, emotional characteristics, beat, and so on, distinguishing the emotional distance of different music, and classifying the characteristics are the key work of this paper. Literature [1] aims at some limitations in music performance recommendation methods, such as users only listen to rock music, while hip hop or R&B with similar emotions is difficult to get corresponding recommendations. For the classification of songs in music performance (popular and unpopular), it is proposed to give corresponding weights to the emotional characteristics in each music performance based on factor analysis of the characteristics of music factors, so as to quantitatively analyze the emotional analysis in music performance. Literature [2] studies the algorithm of classification and recognition of music ranking in the music performance system and analyzes posts and comments from Facebook and corresponding music emotions from human emotion data. The music emotion in the above music performance system is ranked, and the experimental results show that the preparation rate based on music emotion analysis reaches 80%, which is more conducive to improving the competitive quality of music works. In Literature [3], aiming at the problem of music emotion expression in the music performance system in the Internet network, a fusion research on the music emotion recommendation algorithm based on user personalization and network tags is proposed. Feature users are used to measure the similarity of personalized tendency of music emotion, and the music emotion recognition recommendation method is improved. The similarity personalized recommendation is realized from three aspects: emotion acquisition, analysis, and aggregation. Literature [4] analyzed the emotional influence of human beings in music performance. Music emotion expresses human happiness, hope, emotion, and other factors. Expressing emotion through music is a complex problem in intelligent recognition. In order to express this relationship better, it is proposed to extract features (for example, random k -tag set and multitag k -nearest neighbor) and classify human emotions to different music in music expression. Literature [5] proposed a music emotion label prediction algorithm based on the music emotion vector space model to solve the problem that the prediction effect of music emotion label in music

performance is not ideal. Firstly, the emotion vector in the music performance system is extracted, the corresponding spatial model is established, and then it is classified and recognized by SVM. The results show that the music emotion classification method based on the vector space model has better advantages in speech recognition. Since music melody comes from human auditory perception, Paiva et al. [6] proposed a pitch tracking method based on the auditory model to calculate auditory spectrogram of music signal and to calculate pitch significance according to correlation graph to obtain melody pitch candidates. After quantifying the pitch candidates into music notes, the minimum pitch interval is adopted for them. Xia et al. [7] used the continuous emotional psychological model and regression prediction model to generate robot dance movements driven by constructing emotional changes and music beat sequences. Schmidt et al. [8] chose a continuous emotion model, linked the music emotion content with the acoustic feature model, and established a regression, and the model studied the music emotion changes with time. Sordo et al. [9] studied a variety of acoustic features, including bottom-level features and melodies, tones, high-level genres, styles, and other features, then reduced these features to D -dimensional space, linked them with semantic features, and used the K -nearest neighbor algorithm for automatic recognition of music emotions. Anders [10] invited 20 music experts to express the different emotions of happiness, sadness, fear, and calm by controlling the numerical combination of 7 feature quantities such as rhythm and timbre in the equipment and to obtain the relationship between feature quantities and music emotions. Shan et al. [11] established a recommendation model based on music emotion, mainly studying the emotion conveyed by movie music. Chen and Li [12] used a continuous emotional psychological model and regression model to predict the emotional value of music and used two fuzzy classifiers to measure the emotional intensity to identify the emotional content of music. Rajib Sarkar et al. [13] proposed to use convolution neural network to identify music models and compared it with commonly used classifiers such as BP neural network. Rao and Rao [13] proposed a two-way mismatch (TWM) pitch saliency calculation method based on the fact that the harmonic amplitude attenuation speed of singing sounds is lower than that of musical instrument sounds and used the dynamic programming (DP) algorithm to track vocal melodies. Huang et al. [14] proposed a multimodal deep learning method, which uses double convolution neural network to classify emotions. Depth model 370-Boltzmann machine is used to reveal the correlation between audio and lyrics. In the second part of this paper, the emotion classification methods in music performance system are introduced, especially the related physical characteristics and music signal processing. In the third part, the emotion recognition algorithm is analyzed, and eight kinds of emotion expressions are put forward. Then, SVM, KNN, and other models in the intelligent algorithm are explained. In the third part, the intelligent recognition method proposed above is applied to music emotion recognition. From the experimental results, the proposed method has recognition rate and preparation rate.

2. Emotion Classification Method in Music Performance System

Emotion in music performance is based on the resonance of sound in musical instrument performance. The sounds emitted by different equipment will produce different emotional expressions to the audience, and they can be intelligently classified through different emotions.

2.1. Physical Properties of Music. In music performance, several instruments are usually played at the same time.

$$y(x, t) = \sum_{n=1}^{\infty} \left\{ \frac{2}{l} \left[\int_0^1 f_1(x) \sin \frac{n\pi}{l} x dx \right] \cos p_n t + \frac{2}{\pi c n} \left[\int_0^1 f_2(x) \sin \frac{n\pi}{l} x dx \right] \sin p_n t \right\} \sin \frac{p_n}{c} x. \quad (1)$$

Definition of the original frequency of the performance system is as follows:

$$p_n = \frac{n\pi}{l} \sqrt{\frac{T_0}{\rho A}}. \quad (2)$$

When $n=1$, p_1 is described as the fundamental frequency, representing the pitch in the performance music. When n takes 2, 3, 4, ..., n , it is called n harmonic in physical music characteristics, which will have an important influence on timbre in music performance. The length, density, and tension of strings in performance equipment have obvious influence on frequency. If the length of the string is changed, the pitch will also change. The tightness of the string also determines the natural frequency. If the string is tighter, the pitch will be higher, and vice versa.

2.2. Music Feature Extraction. In the music performance system, the combination structure of music is the systematic structure relationship of notes (pitch, length, and intensity), bars (notes, notes, ..., notes), segments (bars, bars, ..., bars), and music (segments, segments, ..., segments). This structure is described as follows:

$$X_m(k) = \sum_{n=0}^{M-1} \omega(n) x(n+mH) e^{-j \frac{2\pi}{N} kn}, \quad m = 0, 1, \dots, N; k = 0, 1, \dots, N-1, \quad (4)$$

where $x(n)$ is a window signal and $\omega(n)$ is a window function. N , m , n , and H , respectively, represent the number of frames, window length, Fourier transform length, and window movement size.

In order to facilitate the processing of the information of the above signals, the transformation in equation (4) needs to be normalized. The processing process is as follows:

$$\bar{X}_m(k) = 2 \frac{|X(k)|}{\sum_{n=0}^{M-1} \omega(n)}. \quad (5)$$

Different musical instruments will produce different timbres, pitches, and other characteristics. In order to better study the basic nature of sound, the next study is the sound principle of musical instruments.

The tension of strings in music equipment reflects the intensity of sound expressed by T_0 , the density is P , the length is L , the strings are divided into $N+1$ segments, and the number of separation points is N , and then the intensity of sound can be described as follows:

$$\begin{cases} \text{Note} = (\text{Pitch}, \text{SL}, \text{SI}), \\ \text{Syll} = (\text{Note}_1, \text{Note}_2, \dots, \text{Note}_n), \\ \text{Per} = (\text{Syll}_1, \text{Syll}_2, \dots, \text{Syll}_n), \\ \text{Music} = (\text{Per}_1, \text{Per}_2, \dots, \text{Per}_n). \end{cases} \quad (3)$$

The smallest unit in music is notes, which are combined into bars, bars are combined into segments, and segments are combined into the whole music. This classification is reasonable. From the perspective of granularity, notes, bars, or passages can be selected as important indicators to identify emotions. Segment size is determined by the characteristics of music itself, which is similar to the whole emotion classification.

2.3. Signal Processing of Music Performance

2.3.1. Spectrum Transformation. The purpose of spectrum transformation is to divide music equipment in music performance into frames and windows electronically and to express the spectrum of each frame information number of received music equipment information. Fourier transformation is used here to realize spectrum transformation. The formula is as follows:

The above signals are averaged and normalized in different windows, and the signals tend to be stable.

2.3.2. Peak Processing. After the above processing, the signal can obtain the corresponding peak frequency and amplitude. The simplest method is to directly use the position of the frequency point and the amplitude of the Fourier transform, but this method is limited by the frequency resolution of the Fourier transform. In order to solve this limitation, various

correction methods have been formed to achieve higher frequency accuracy, and better amplitude estimation can also be obtained.

$$\widehat{f} = (k_m + \kappa(k_m)) \frac{f_s}{N}, \quad (6)$$

where \widehat{f} is the evaluated peak frequency, f_s is the sampling rate, and κ is the offset of the window, and the calculation formula is as follows:

$$\kappa(k) = \frac{N}{2\pi H} \psi \left(\phi_m(k) - \phi_{m-1}(k) - \frac{2\pi H}{N} k \right). \quad (7)$$

However, in the high-pitched region, it is significant and is expressed by the following formula:

$$S(b) = \sum_{h=1}^{N_k} \sum_{i=1}^N e(\widehat{\alpha}_i) \cdot g(b, h, \widehat{f}_i) \cdot (\widehat{\alpha}_i)^\beta, \quad (8)$$

where $\widehat{\alpha} = (X_m(k_m)N(1-z^2))/(\sin c(z)(\kappa(k_m)))$, N and h denote the maximum number of harmonics, and i denotes the number of peaks.

$$B^H(t) = \frac{1}{\Gamma(H+0.5)} \left\{ \int_{-\infty}^0 [(t-\tau)^{H-0.5} - (-\tau)^{H-0.5}] dB(\tau) + \int_0^t (t-\tau)^{H-0.5} dB(\tau) \right\}, \quad (9)$$

$$\nabla B^H(t) = B^H(t) - B^H(t-1) = \int_{-\infty}^1 K_H(t-u) dB(u), \quad (10)$$

where m is the average arrival rate, $\Gamma(*)$ is the γ function, H is the related parameter ($0 < H < 1$), B^H is the standard fractal Brownian motion, and $\nabla B^H(t)$ is a discrete fractal Gaussian noise model, and the function $K_H(u)$ is as follows:

$$K_H(u) = \begin{cases} u^{H-0.5}, & 0 < u < 1, \\ u^{H-0.5} - (u-1)^{H-0.5} & u > 1. \end{cases} \quad (11)$$

From equation (9), it can be seen that $B^H(at) = a^H B^H(t)$, the mean value is 0, the variance is $\sigma^2 t^{2H}$, and the autocorrelation function is shown in the following equation:

$$r(t) = \frac{\sigma^2 (t^{2H} + s^{2H} - |s-t|^{2H})}{2}. \quad (12)$$

The probability density function of FBM is

$$f_{B^H}(x, t) = \frac{1}{\sqrt{2\pi t^{2H}}} e^{-(x^2/2t^{2H})}. \quad (13)$$

In equation (10), $\nabla B^H(t)$ obeys $N(0, \sigma^2 |\tau|^2)$ the Gaussian distribution, where σ^2 is as follows:

$$\sigma^2 = E \left\{ |B^H(1)|^2 \right\} = \frac{\Gamma(1-2H) \cos(\pi H)}{\pi H}. \quad (14)$$

The wavelet is reconstructed by scale contraction and related scale function, and the appropriate wavelet function and scale function are selected as shown in the following equation:

$$\begin{cases} \psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k), \\ \delta_{j,k}(t) = 2^{j/2} \delta(2^j t - k). \end{cases} \quad (15)$$

2.3.3. Music Signal Filtering. In this paper, fractal wavelet is used to filter the equipment signals in music performance. In the complex environment of different equipment sounds, there are M sound source nodes [15], in which node O is the source point that controls all equipment, node An is the music sound source, and the traffic flow is sent from node An to node O and then forwarded through node O . Traditional methods to describe the performance of traffic flow include Poisson model and Gaussian distribution. The research basis is that music signals show short correlation characteristics. However, with the deepening of the research, the results show that the actual music signals show self-similarity and long-correlation characteristics. Based on the fractal Brown motion (FBM) model, the following description model is proposed:

Equation (15) becomes a pair of orthogonal bases, and the signal $X(t)$ is shown in the following equation:

$$X(t) = \sum_K U_{j,k} \delta_{j,k}(t) + \sum_{j=j_0}^{\infty} \sum_k W_{j,k} \delta_{j,k}(t). \quad (16)$$

The scale coefficient $U_{j,k}$ and wavelet coefficient $W_{j,k}$ are shown in the following equation:

$$\begin{cases} U_{j,k} = \int X(t) \delta_{j,k}(t) dt, \\ W_{j,k} = \int X(t) \psi_{j,k}(t) dt, \end{cases} \quad (17)$$

where j represents the scale and K represents different audio sources.

Due to environmental reasons or the influence of other bands, there is noise in the original music, so it is necessary to filter the music signal. Compared with the existing filtering algorithms, wavelet filtering is time-sensitive and fast in signal processing. Other filtering algorithms are not as efficient as wavelet filtering.

3. Emotional Intelligence Recognition Algorithm

3.1. Emotional Model. In music performance, the audience is emotional and can express a series of emotional types. The following emotional model is explained. At present, Hevner's emotional ring model [16] is widely used, which is

divided into music emotions that have progressive connections with their adjacent classes. The emotions among the 8 classes can smoothly transit to each other, forming a ring emotional model to symbolize the emotional response of music works, as shown in Figure 1.

Hevner's emotional loop model is a music emotional model composed of a series of discrete words, and the emotional attributes of these offline words are quite different, which is beneficial to the emotional identification of music works. In order to classify the emotional experience of music works, researcher Hevner asked the experimenters to compare the eight emotions they felt under the stimulation of hundreds of music works. Hevner's emotional ring model was tested by many experimenters on different tones of music works. It is found that major is easy to make people happy, elegant, and lively; minor is easy to make people sad, illusory, multifeling, and other emotional expressions. The speed of music also has a certain impact on human emotions; music with fast speed and rhythm is easy to make people happy and cheerful; slow-speed or slow-paced music tends to make people suffer and sad negative emotions.

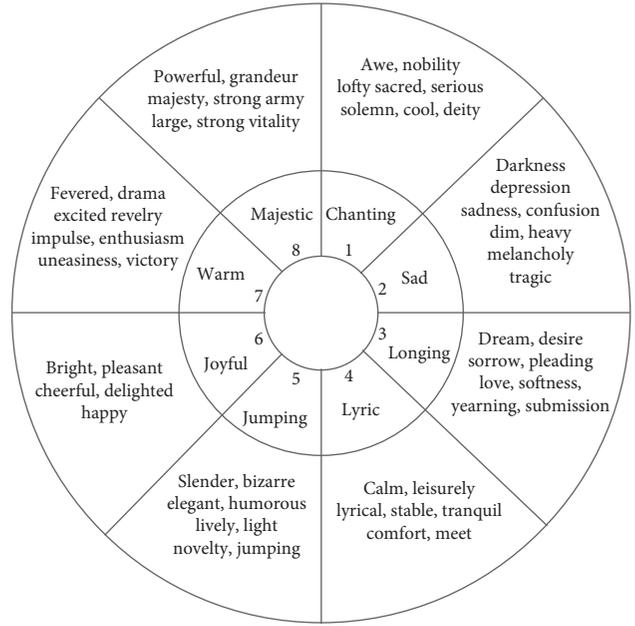


FIGURE 1: Hevner's emotional loop model.

3.2. Intelligent Identification Model

3.2.1. SVM Model. The principle of SVM is to find a hyperplane to complete the binary classification of samples. That is to say, the given training dataset is transformed into the corresponding convex quadratic programming problem by the principle of interval maximization to obtain an optimal separation hyperplane, and the dataset is divided into two categories. Figure 2 is a hyperplane schematic diagram.

For linearly separable SVM hyperplanes, the leading quantity w and the distance b are described.

$$w \cdot x + b = 0. \quad (18)$$

The classification function for equation (18) is defined as follows:

$$f(x) = \text{sign}(w \cdot x + b). \quad (19)$$

As can be seen from Figure 2, when the size of the margin value determines the effect of SVM classification, the margin is calculated as follows:

$$\text{margin} = \frac{2}{\|w\|}. \quad (20)$$

When the classification effect of the sample is more obvious, the larger the margin value is. For linear SVM classifiers, there is a minimum requirement for margin value. If the value is too small, the worse the classification effect of SVM classifiers is, and the constraint relationship is as follows:

$$\min_{w,b} \frac{1}{2} \|w\|^2, \quad (21)$$

$$\text{s.t. } y_i(w \cdot x_i + b) - 1 \geq 0, \quad i = 1, 2, \dots, N.$$

Generally, for the objective function of linear SVM, Lagrange operator is introduced to transform the original

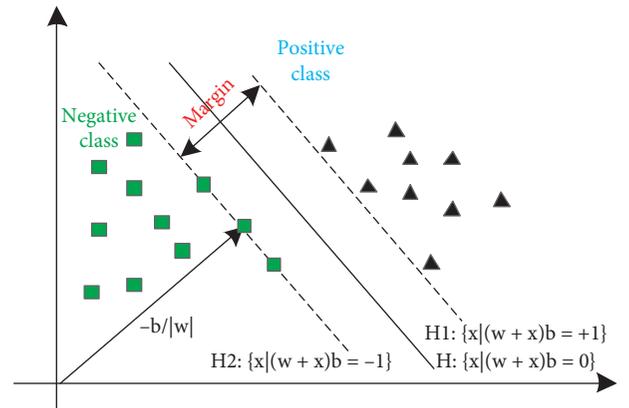


FIGURE 2: SVM model.

optimization problem into dual problem for calculation simplification. Linear separability is based on the ideal sample situation. In the dataset used in real classification problems, most samples are linearly inseparable.

3.2.2. k Nearest Neighbor Algorithm. K nearest neighbor (kNN) classification algorithm [17] is a common passive supervised classification algorithm in the field of pattern recognition. The classification standard model for the KNN algorithm is defined based on the formula of distance:

$$\text{dist}(s_i, s_j) = \sqrt{\sum_{k=0}^{N-1} (a_k(s_i) - a_k(s_j))^2}. \quad (22)$$

For equation (22), the smaller the dist value, the higher the similarity between samples, and the N samples with high similarity are only classified in the same category.

3.2.3. Multiclassifier Fusion Method. In music performance, the processed audio signal is divided into audio features and text features by the above algorithm, and then multimodal fusion is carried out so as to judge the overall emotion after music performance. Model fusion is to integrate multiple learners through model combination strategy and finally achieve better prediction results than a single model. In this paper, the common average fusion strategy is selected to realize multiclassifier fusion.

The above classifiers are fused by simple average or weighted average, and the fused average value of the two classifiers is taken as the final value of the model. The corresponding average value or weighted value for SVM and CNN models is taken, and the formula is as follows:

$$P(x_0) = \frac{1}{M} \sum_{m=1}^M y_m(x). \quad (23)$$

In this paper, SVM and CNN classifiers are used for fusion research. Considering the poor classification effect and low accuracy of a single classifier, the fusion classifier can improve the classification effect of music features.

3.3. Evaluation Model. The evaluation model is to better evaluate the classification effect of the above classifiers. The following formula is adopted in this paper:

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{FN}}, \\ \text{Accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F_1 &= \frac{2\text{TP}}{2\text{TP} + \text{FN} + \text{FP}}. \end{aligned} \quad (24)$$

4. Experimental Analysis

In this paper, the dataset of the song audio emotion classification experiment comes from the PEMO pop music emotion dataset. According to the Thayer emotion model, four emotional tag song lists are extracted from the website. The list is as follows: Chanting, Sad, Longing, Lyric, Jumping, Joyful, Warm, and Majestic. Among them, 100 songs are extracted from each emotional category, and a total of 800 songs are extracted in the dataset. The audio signal is divided into 15 s, 30 s, and 45 s original fragments, 80% of which are training sets and 20% are sample test sets.

In order to better detect the proposed algorithm effectively, the same audio is tested in segments so that the segmentation goal can be reflected in different segments of

audio and whether the test results are consistent. In the experiment, the algorithms SVM, KNN, ANN, and ID3 are used as the basic classification algorithms, and the effects of different combinations are tested by adding the weights of different algorithms to find the optimal model in the classification combination algorithm. The flow of the classification algorithm adopted in this paper is shown in Figure 3:

In this group of experiments, the original audio signal was processed and compared with the unprocessed audio. The classification effects of different algorithms were analyzed from 8 indicators. The LLDs features were input, and the above four classification algorithms were used to classify and output music emotions.

As can be seen from Tables 1 and 2, when the music segment time is relatively short, the shorter the segment, the higher the accuracy. Similarly, the accuracy of emotion classification of the processed audio as a whole is higher than that of the original audio. It shows that the audio signal is filtered and the effect is better after denoising. The following is a comparison of the overall average values of the four algorithms before and after audio signal processing, as shown in Figures 4 and 5.

From the average value of the audio under the above four algorithms, the overall processed audio has a certain improvement in accuracy under the four algorithms. Among the four single classifications, SVM and ANN classifiers have ideal preparation rate, while ID3 classification has the worst effect.

After experimental comparison of four single classifiers, the fusion of the above different algorithms is studied. The classification accuracy of fusing different classifiers is shown in Figures 6 and 7.

As can be seen from Figures 6 and 7, the overall accuracy of the processed audio is relatively high, averaging over 70%. The highest is that the accuracy of audio classification after SVM and ANN fusion is 75% (15 s). The lowest is that the accuracy of ANN and ID3 fusion is only 55% (45 s). The classification effect of audio without preprocessing is better than that of a single classifier, and the accuracy rate of audio without preprocessing is still about 5% lower than that of preprocessing classifier.

Based on the above fusion algorithm, the two indexes of recall and F_1 are analyzed again to evaluate the full search efficiency and algorithm stability after classifier fusion. The specific effect is shown in Figure 8.

The fusion effect of SVM and ANN algorithm is the best, with the highest recall and stable algorithm model. SVM and ANN are used for audio signal classification, which has a good classification effect.

The single performance analysis of four common classification algorithms is carried out, and the classification algorithms with two classification effects are fused, and the fusion effect can achieve the best classification performance. If the combination is carried out again, the performance is not significantly improved, but the complexity of the algorithm increases. Therefore, the combined recognition method can achieve the best performance.

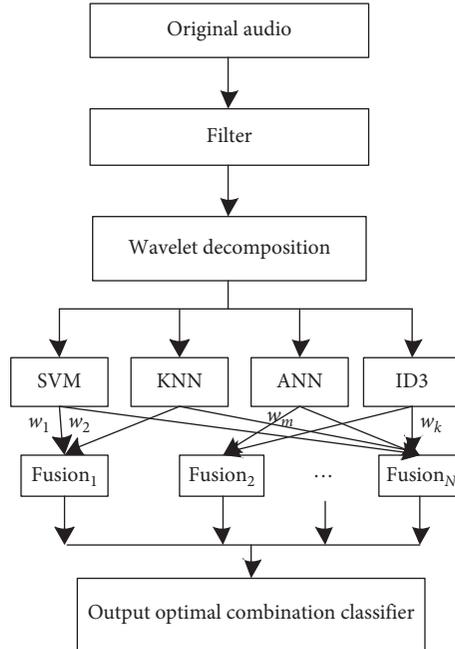


FIGURE 3: Music emotion classification algorithm flow.

TABLE 1: SVM emotion classification accuracy of original audio.

Time (s)	Chanting	Sad	Longing	Lyric	Jumping	Joyful	Warm	Majestic	Avg
45	0.574	0.457	0.578	0.474	0.468	0.462	0.534	0.485	0.504
30	0.652	0.574	0.628	0.524	0.547	0.562	0.589	0.631	0.588
15	0.687	0.654	0.674	0.626	0.635	0.643	0.662	0.673	0.657

TABLE 2: That accuracy of emotion classification in SVM after processing audio

Time (s)	Chanting	Sad	Longing	Lyric	Jumping	Joyful	Warm	Majestic	Avg
45	0.624	0.527	0.635	0.624	0.587	0.545	0.582	0.526	0.581
30	0.667	0.604	0.662	0.669	0.637	0.581	0.622	0.571	0.627
15	0.703	0.643	0.686	0.694	0.684	0.634	0.659	0.634	0.667

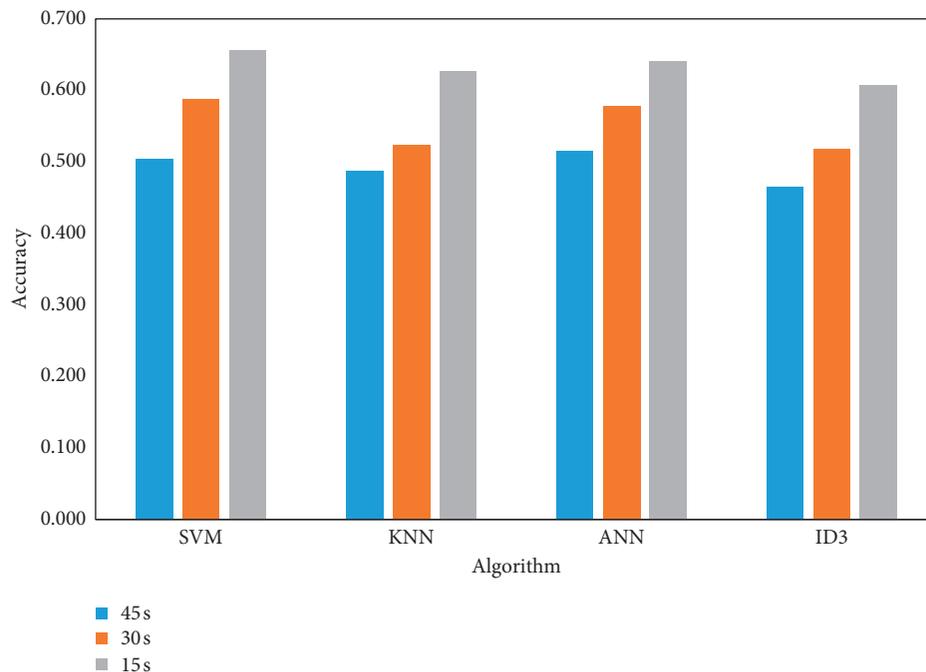


FIGURE 4: Emotion classification accuracy of original audio under four algorithms.

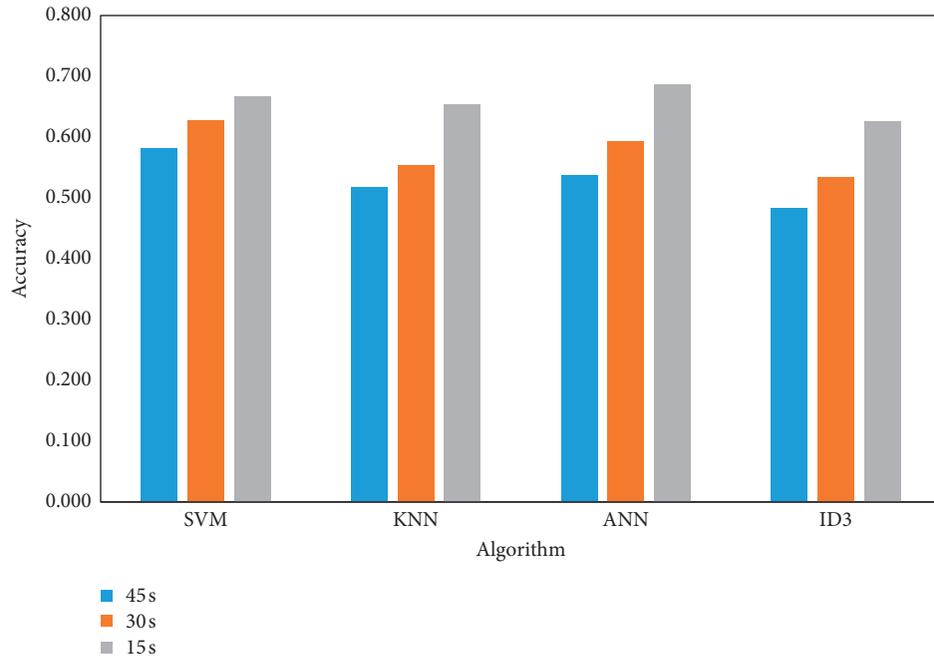


FIGURE 5: Emotion classification accuracy under four algorithms after processing audio.

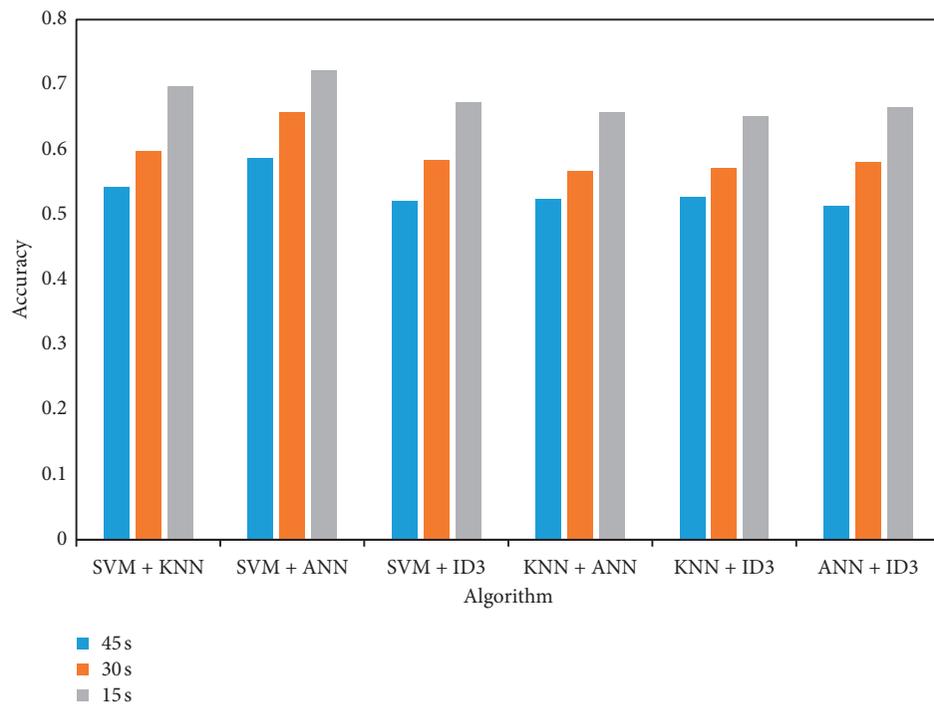


FIGURE 6: Comparison of accuracy under multiclassification fusion of original audio.

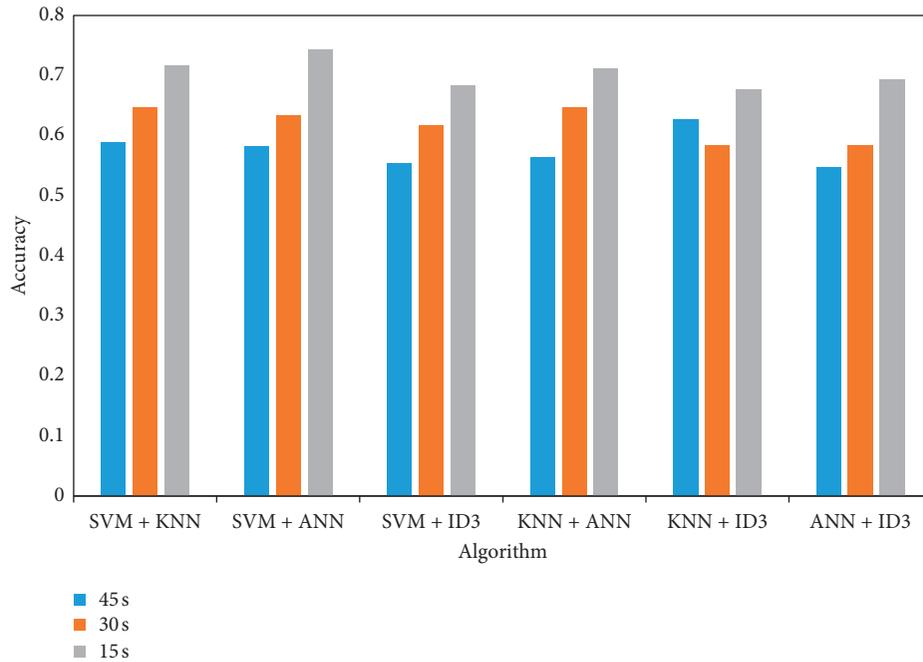


FIGURE 7: Comparison of accuracy under multiclassification fusion after processing audio.

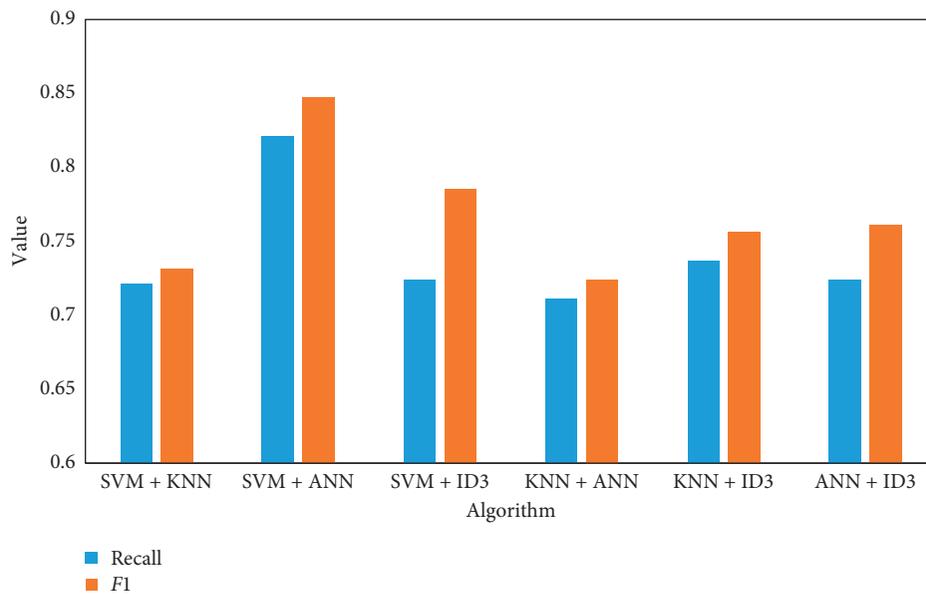


FIGURE 8: Comparison of recall and $F1$ under four algorithms.

5. Conclusion

In music performance, the audience's performance of audio signals is a lot of emotional expression. How to classify audio information emotionally is to study different music performances and bring the best audio-visual effect to listeners. In this paper, a single classifier of SVM, KNN, ANN, and ID3 is used to classify audio signals, and the classification accuracy of audio signals before and after processing is

compared. The processed audio has the best classification effect. Further research on the fusion of different single classifiers to classify audio signals has obvious advantages in accuracy, recall rate, and $F1$.

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.

References

- [1] S. J. Lee, B. G. Seo, and D. H. Park, "Development of music recommendation system based on customer sentiment analysis," *Journal of Intelligence and Information Systems*, vol. 24, no. 12, pp. 197–217, 2018.
- [2] S. Khoeurn and S. K. Yun, "Sentiment analysis engine for Canadian music industry re-building," *Journal of the Korea Society for Simulation*, vol. 26, no. 4, pp. 23–34, 2017.
- [3] M. Wang and Y. B. Wang, "One web music aggregation algorithm based on social tags sentiment," in *Proceedings of International Conference on Advanced Control, Automation and Robotics*, pp. 653–661, Hong Kong, China, 2015.
- [4] L. M. Gómez and M. N. CáCeres, "Applying data mining for sentiment analysis in music," in *Proceedings of International Conference on Practical Applications of Agents and Multi-Agent Systems*, pp. 198–205, Cham, Switzerland, 2018.
- [5] L. I. Jing, H. Lin, and L. I. Ruimin, "Sentiment vector space model based music emotion tag prediction," *Journal of Chinese Information Processing*, vol. 26, no. 6, pp. 45–44, 2012.
- [6] R. P. Paiva, T. Mendes, and A. Cardoso, "Melody detection in polyphonic musical signals: exploiting perceptual rules, note salience, and melodic smoothness," *Computer Music Journal*, vol. 30, no. 4, pp. 80–98, 2006.
- [7] G. Xia, J. Tay, R. Dannenberg, and M. Veloso, "Autonomous robot dancing driven by beats and emotions of music," in *Proceedings of 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1 AAMAS*, pp. 205–212, Valencia, Spain, June 2012.
- [8] E. M. Schmidt, T. Douglas, and E. K. Youngmoo, "Feature selection for content-based, time-varying musical emotion regression," in *Proceedings of International Conference on Multimedia Information Retrieval*, pp. 267–274, Philadelphia, PA, USA, March 2010.
- [9] M. Sordo, O. Celma, and D. Bogdanov, "MIREX 2011: Audio tag classification using weighted-vote nearest neighbor classification," in *Proceedings of Music Information Retrieval Evaluation Exchange*, Miami, FL, USA, January 2011.
- [10] R. Anders, "Emolion rendering in music range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.
- [11] M.-K. Shan, F.-F. Kuo, M.-F. Chiang, and S.-Y. Lee, "Emotion-based music recommendation by affinity discovery from film music," *Expert Systems with Applications*, vol. 36, no. 4, pp. 7666–7674, 2009.
- [12] C. Chen and Q. Li, "A multimodal music emotion classification method based on multifeature combined network classifier," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4606027, 11 pages, 2020.
- [13] V. Rao and P. Rao, "Vocal melody extraction in the presence of pitched partnership in polyphonic music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 8, pp. 2145–2154, 2010.
- [14] M. Huang, W. Rong, T. Arjannikov, J. Nan, and Z. Xiong, "Bimodal deep Boltzmann machine based musical emotion classification," in *Proceedings of International Conference on Artificial Neural Networks*, pp. 199–207, Barcelona, Spain, September 2016.
- [15] T. Liu, B. Zhu, S. Q. Sun, and C. Yang, "Music's effective computing model based on fuzzy logic," in *Proceedings of 2006 6th World Congress on Intelligent Control and Automation*, pp. 9477–9481, Dalian, China, June 2006.
- [16] G. Chen, Z. Sun, and L. Zhang, "Road identification algorithm for remote sensing images based on wavelet transform and recursive operator," *IEEE Access*, vol. 8, 2020.
- [17] G. Chen, J. Qi, C. Tang, Y. Wang, Y. Z. Wu, and X. Shi, "Analysis and research of key genes in gene expression network based on complex network," *Complexity*, vol. 2020, Article ID 8816633, 12 pages, 2020.