

## *Retraction*

# **Retracted: Music Classification and Detection of Location Factors of Feature Words in Complex Noise Environment**

### **Complexity**

Received 23 January 2024; Accepted 23 January 2024; Published 24 January 2024

Copyright © 2024 Complexity. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Manipulated or compromised peer review

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### **References**

- [1] Y. Xu and Q. Li, "Music Classification and Detection of Location Factors of Feature Words in Complex Noise Environment," *Complexity*, vol. 2021, Article ID 5518967, 12 pages, 2021.

## Research Article

# Music Classification and Detection of Location Factors of Feature Words in Complex Noise Environment

Yulan Xu<sup>1</sup> and Qiaowei Li <sup>2</sup>

<sup>1</sup>Academy of Film and Television Arts, Hunan Mass Media Vocational and Technical College, Changsha, Hunan 410012, China

<sup>2</sup>Academy of Music, Hengyang Normal University, Hengyang, Hunan 421002, China

Correspondence should be addressed to Qiaowei Li; [liqiaowei1129@hynu.edu.cn](mailto:liqiaowei1129@hynu.edu.cn)

Received 26 February 2021; Revised 6 April 2021; Accepted 12 April 2021; Published 21 April 2021

Academic Editor: Zhihan Lv

Copyright © 2021 Yulan Xu and Qiaowei Li. 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 order to solve the problem of the influence of feature word position in lyrics on music emotion classification, this paper designs a music classification and detection model in complex noise environment. Firstly, an intelligent detection algorithm for electronic music signals under complex noise scenes is proposed, which can solve the limitations existing in the current electronic music signal detection process. At the same time, denoising technology is introduced to eliminate the noise and extract the features from the signal. Secondly, from the perspective of audio and lyrics of song sentiment analysis and the unique characteristics of lyrics text, a lyric sentiment analysis method based on text title and position weight is proposed. Finally, considering the influence of the weight of feature words in different positions on the classification of lyrics, the analytic hierarchy process is used to calculate the weight of feature words in different positions of text title and lyrics before, in, and after the text. The results show that in the complex noise environment, the accuracy of music classification and detection of the proposed model is more than 90%, which is far beyond the control range of the actual application of music processing. The effect of music classification and detection is better than that of the contrast model, which has a certain practical application value.

## 1. Introduction

Music is a form of expression of culture. Under different musical instruments, rhythms, and arrangements, people can feel joy, anger, sorrow, and joy in music. Anger, sorrow, and joy have become the main form of people enjoying music now [1–3]. With the popularization of smart phones, the scale of mobile online music users has further expanded. In order to improve user loyalty, most music software has successively introduced music recommendation, classification, and other functions to enhance user experience. How to effectively search for music of a specific emotion classification has become one of the hot research directions [4–6].

Due to the massive investment of scholars and the attention of all occupations, there are many current music detection algorithms [7–12]. Compared with other types of algorithms, music detection algorithms based on artificial neural networks and support vector machines are the most commonly used.

Moreover, they are all new technologies in the field of artificial intelligence [13–15]. In the actual application of music detection, both artificial neural networks and support vector machines have some insurmountable problems. For example, the detection error rate of the artificial neural network is high, the music detection results are extremely unstable, and people often cannot get the music detection results that they really need. The music detection speed of the support vector machine is slow. For large-scale music signals, it is usually impossible to obtain the music detection result within the effective time. At the same time, they do not consider the interference of noise on the music detection results, and the robustness to noise is poor, which affects the classification and subsequent processing of music [16–18].

In order to improve the effect of music classification and detection, this paper proposes a music classification and detection algorithm under complex noise scene and tests its superiority and versatility. Contributions of this paper are as follows:

- (1) An intelligent detection algorithm for electronic music signals under complex noise scenes is proposed, which can solve some limitations existing in the current electronic music signal detection process.
- (2) Denoising technology is introduced to eliminate noise and extract features from signals.
- (3) Based on the research of song sentiment analysis from the perspective of audio and lyrics and the unique characteristics of lyrics text, a lyric sentiment analysis method based on the combination of text title and position weight is proposed.
- (4) Considering the influence of the weight of feature words in different positions on the classification of lyrics, the analytic hierarchy process (AHP) was used to calculate the weight of feature words in different positions of text titles and lyrics before, in, and after the text.

## 2. Related Works

Music classification is to divide existing music into different categories according to certain standards. The classification standard is generally determined according to human subjective perception. The knowledge of music psychology, music signal processing, pattern recognition, and other disciplines is involved in the whole system framework of music classification. At present, many researchers have proposed many effective algorithms to realize music classification. The new era of the Internet is in the ascendant. People are also pursuing new development while focusing on intelligence. Music is infiltrating people's lives in various forms and has quietly become people's spiritual food after dinner. At the same time, the Internet has become the main carrier of music transmission.

Taking the audio angle as the starting point, the researcher conducts emotional research on music by extracting relevant acoustic emotional features from the audio data of music rhythm and melody and analysing the speed and intensity of the music. Chen et al. [19] proposed an information cell hybrid to classify music emotions. This method is based on ternary semantic structure to model the high-dimensional concepts and estimate the emotions in each piece of music. Hamada et al. [20] proposed a music classification method based on music highlight detection. In this paper, Gaussian mixture model and AdaBoost algorithm are used to combine the proposed rhythm characteristics with timbre characteristics and improve the performance of music emotion classification based on the detected music highlight segments. The music is divided into three categories: calm, pleasant, and excited. The experimental results confirm that the proposed method achieves preliminary and promising results in terms of accuracy. The final accuracy of this method is as high as 97.6%. Ke et al. [21] constructed a Chinese music classification model based on support vector machines by extracting music acoustic features. It is found that the characteristics of the combination of specific time-domain factors and three frequency-domain factors are the best for

music sentiment classification using the SVM classifier in Chinese, with an accuracy rate of 83.85%. It can be seen that temporal features are still very important in audio-based music emotion classification. Zhu et al. [22] took the multitrack perspective as a starting point to extract five aspects of acoustic features. The BP neural network model is used for more than 100 trainings, and finally the music emotion is divided into 8 categories. The accuracy rate exceeds 90% within the given error range, and good results have been achieved. Su et al. [23] proposed a music recognition system based on a two-level support vector machine. In 301 music works, 10-fold cross-validation is used to evaluate the classification performance. Different from the traditional classification method, this classification method assigns an emotional category to each 30s music clip in a dichotomy. The emotional classification of pop, rock, jazz, and blues into four categories of happiness, sadness, calm, and anger has also achieved good results.

From the perspective of lyrics, Ahuja et al. [24] focused on how to extract useful and meaningful language features from the lyrics text to assist in the emotional classification of music. Under the framework of the n-gram language model, they studied three preprocessing methods and a series of language models with different orders to extract more semantic features. Then, three different learning ideas are used to test the classification performance. The experimental results show that the feature extraction method improves the accuracy of music emotion classification. In the lyrics-based sentiment analysis method proposed by Sun et al. [25], the model is integrated into the sentiment vector space model to consider the sentiment unit in the feature definition, and a more discriminative support vector model is used for song sentiment classification research. Song et al. [26] used Chinese word segmentation, feature extraction, text vectorization, and weighting to perform emotional classification in popular music based on the perspective of lyrics. The KNN algorithm was used to perform emotional classification with an accuracy rate of over 90%. In order to verify the impact of different data sets on classification performance, four different data sets were put into the naive Bayes model to train and compare the results [27]. In the end, the accuracy rate after sentiment classification is about 68%, which verifies that the impact of different data sets is different.

From the perspective of audio and lyrics fusion, Wang et al. [28] conducted a comprehensive study on the role of lyrics in music emotion classification by evaluating and comparing various text features of lyrics, including language and text style features. Then, two fusion methods are used to combine the best lyrics features with the features extracted from the music audio. Wang et al. [29] proposed a multilabel k-nearest neighbour algorithm. They combined the method with the TF-IDF algorithm to solve the ambiguity of emotional words for music emotional classification research. Panda et al. [30] proposed a method to detect song emotions based on lyric and audio features. Firstly, the lyrics are divided to generate lyric features, and then the value and arousal value are calculated. In addition, the language association rules are applied to properly handle the ambiguity.

When performing emotional analysis of music, different data forms need to be digitized for further processing [31, 32]. Some scholars decompose audio signals such as music rhythm and timbre, extract acoustic features, and combine related algorithms to classify emotions. Noroozi et al. [33] imported acoustic information into a support vector machine to construct a Chinese music classification model. Chaudhary et al. [34] extracted information from different audio tracks as the sample data of the BP neural network model for music emotion classification. Huang et al. [35] used support vector machines to analyse music emotions by summarizing words with specific meanings.

In recent years, machine learning has developed rapidly in the analysis of the formation factors of complex problems [36, 37]. Some scholars also use machine learning to analyse the corresponding relationship between music emotion and audio signal. With the deepening of the research, the research method has gradually evolved from the original book analysis of single mode to the joint analysis of multidimensional data, and the multimodel feature data analysis has also achieved good applications in the music emotion analysis.

### 3. Music Classification and Detection Algorithm

**3.1. Noise Suppression of Music Signal and Feature Extraction of Music Signal.** When music contains noise, the music signal curve will change. Suppose the noise is  $n(t)$  and the useful and clean music signal is  $nr(t)$ , then the music signal in the complex noise environment is as in the following equation:

$$n(t)' = nr(t) - \alpha^* n(t). \quad (1)$$

In formula (1),  $\alpha$  represents the type of noise, such as white noise and Gaussian noise.

The noise  $n(t)$  usually increases the storage space of the music signal, and the variation curve mutates, which makes it impossible to correctly identify the type of the music signal. Because of the need to eliminate the negative impact of noise, this article chooses wavelet transform to remove the noise of music signal. Suppose a music signal containing noise before and after the change curve is shown in Figure 1. It can be clearly seen that the change characteristics of the noise-free music signal and the music signal with noise are very different.

In order to suppress the noise in the music signal to the greatest extent, the wavelet analysis algorithm is used to remove the noise in the music signal. The continuous wavelet should satisfy the following conditions:

$$C(t) = \left(\frac{1}{a}\right)^{1/2} \int n(t) \left(\frac{t-q}{a}\right) dt. \quad (2)$$

In formula (2), the variable  $q$  is the displacement.

A computer processes the music signal, so it is a discrete signal. Set  $a = 2^j/k$  and  $q = 2^{-j}/k$ , and the discrete wavelet is obtained as follows:

$$C(t) = \int n(t) * 2^{1/j} \left( \frac{2^{1/j}t - (2 - j/k)}{2/j} \right) dt. \quad (3)$$

The steps to suppress electronic music signals are as follows:

*Step 1.* Use wavelet analysis to decompose the original music signal to obtain multiple music subsignals. Each music subsignal contains a wavelet coefficient, and the music subsignals are arranged according to the frequency. The wavelet coefficient vector of the first music subsignal is expressed as  $n_1$ , and then the wavelet coefficient vector set of all subsignals is  $N = [n_1, n_2, \dots]$ .

*Step 2.* The music signal vector set can be divided into two parts: useful music signal ( $a_m$ ) and noise ( $n_m$ ), namely:

$$N = a_m + n_m. \quad (4)$$

*Step 3.* Calculate the standard deviation of the music signal noise ( $g$ ). Then, set a threshold ( $thr$ ), and the relationship between the two can be described as follows:

$$thr = g(\lg(n))^{1/2}. \quad (5)$$

Set the wavelet coefficient value of the subsignal vector greater than  $thr$ , and the others remain unchanged. The new vector of the wavelet coefficient of the music signal is as follows:

$$N_g = \begin{pmatrix} N, |N| \geq thr, \\ 0, |N| < thr. \end{pmatrix} \quad (6)$$

*Step 4.* Use wavelet analysis to reconstruct the wavelet coefficients of the music signal to obtain a music signal without noise.

**3.2. Extraction of Location Factors of Feature Words in Lyrics.** The overall structure of lyrics sentiment analysis is shown in Figure 2. In order to fuse lyrics and audio data, it is necessary to extract feature vectors of the two forms of data, respectively. The extraction of the feature vector of the lyrics text adopts a vector space model. The idea is to treat different music lyrics as a vector in an N-dimensional space and form a feature word-lyric matrix by sorting the feature words in the lyrics text.

In order to characterize the influence of feature words with different frequencies on the ability of musical emotion expression, it is necessary to evaluate the frequency and location of feature words. The term frequency-inverse document is usually used to judge the importance of a specific word to a text in the training data set. TF-IDF mainly involves two aspects of calculation. TF represents the ratio of the frequency of a certain characteristic vocabulary in a text to the number of occurrences of all the characteristic words in the text. IDF is used to characterize the degree of classification of a characteristic word, and its calculation formula is as follows:

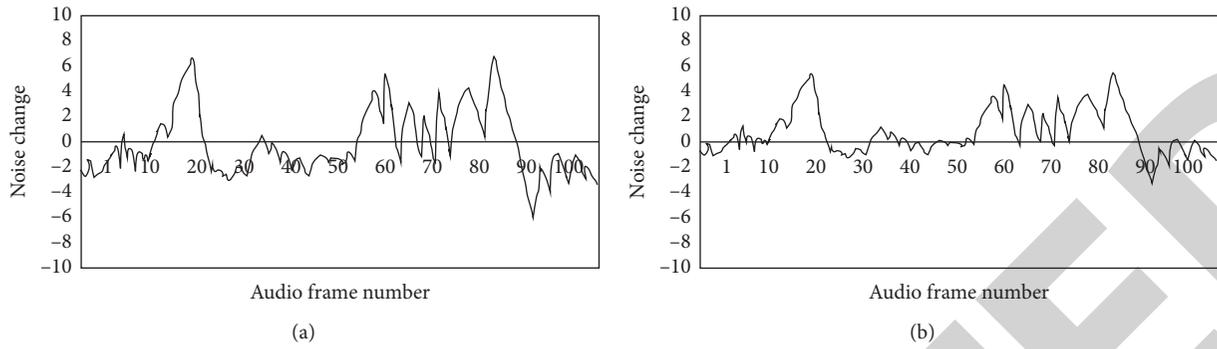


FIGURE 1: Music signal noise removal effect test: (a) noisy music signal; (b) noisy-removed music signal.

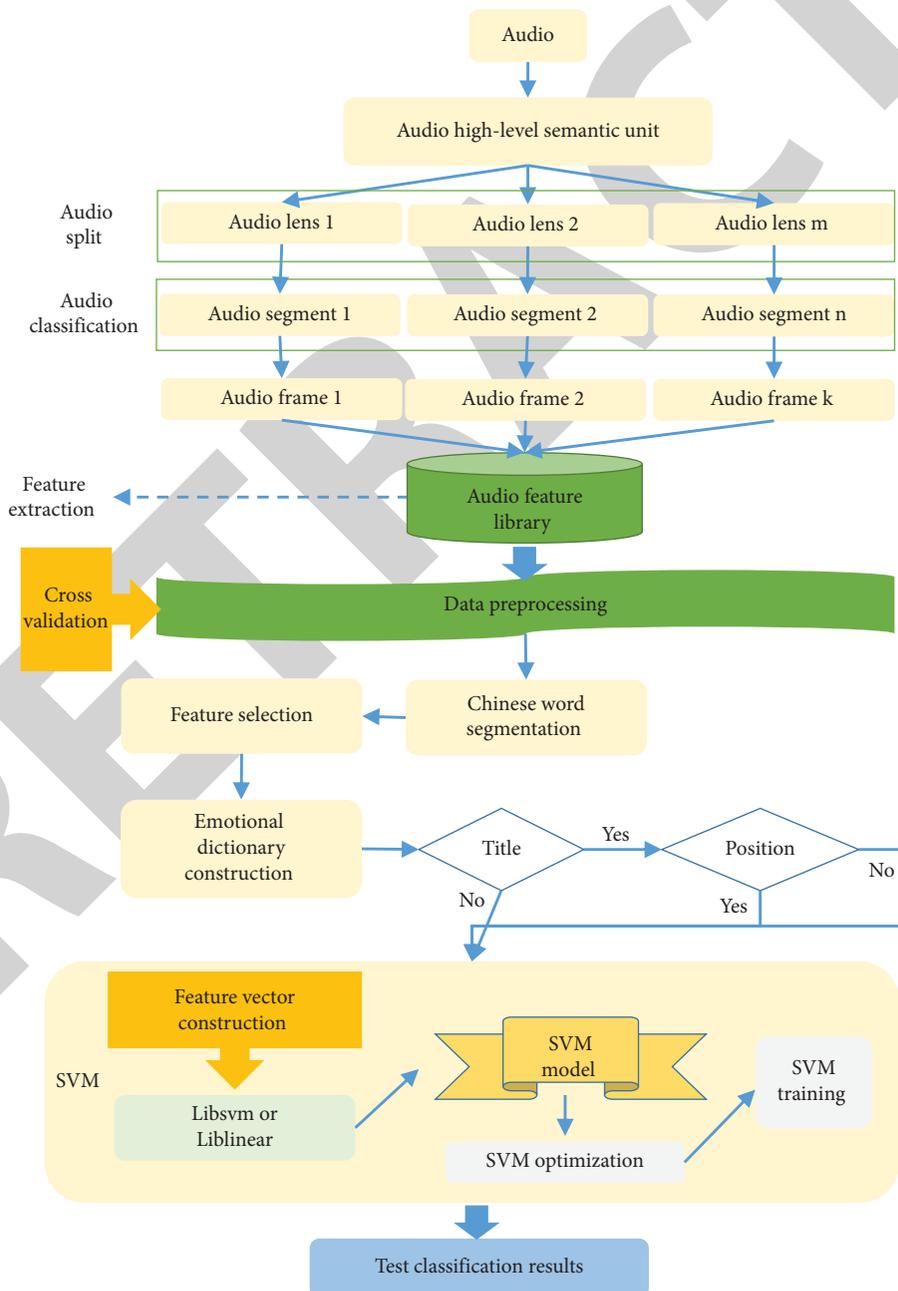


FIGURE 2: Lyrics sentiment analysis process.

$$\text{IDF}_t = \ln\left(\frac{N}{\text{num}(t)}\right). \quad (7)$$

In formula (7),  $\text{num}(t)$  represents the number of texts containing the characteristic word  $x$  and  $N$  represents the total number of texts. TF-IDF has played a good effect in many situations. However, the algorithm is not sensitive to the location of feature words and cannot evaluate the category impact brought by the location of feature words. It can be seen that the position where the characteristic word appears has a great influence on the emotion displayed by the entire lyrics. Therefore, this article divides the lyrics text into 4 parts: verse, chorus, sublimation, and ending, and calculates the corresponding weights of position factors and then performs more accurate emotion classification results. In this paper, TF-IDF is improved, and the following expression is used for calculation:

$$\text{TF-IDF} = t^* [d + \lg(w + 1)]^* k^* h. \quad (8)$$

In formula (8),  $w$  is consistent with the definition in TF-IDF, which means the frequency of a certain characteristic word in the current sample. The variable  $h$  represents the position factors of the verse, chorus, sublimation, and ending. When calculating the word frequency TF, considering that the number of words in the lyrics itself is small, the number of feature words is likely to be 1, which makes the  $\lg$  function value 0, which affects the accuracy of the experiment. Variable  $d$  and variable  $k$  are the adjustment factors of word frequency and position factor, respectively, mainly to reduce the influence of the difference between the structures of different music lyrics on music sentiment analysis.

For the location factor  $h$ , analytic hierarchy process (AHP) is used in the paper to calculate. The main process is as follows: constructing a hierarchical structure model, constructing a paired comparison matrix, and checking consistency. The lyrics in this article are divided into 4 parts, and the weights are compared by constructing a judgment matrix. When there are  $n$  factors  $S = \{s_1, s_2, \dots, s_n\}$ , the judgment matrix  $M$  can be expressed as follows:

$$M = \begin{pmatrix} m_{11} & m_{12} & \dots & m_{1n} \\ \dots & \dots & \dots & \dots \\ m_{n1} & m_{n2} & \dots & m_{nn} \end{pmatrix}. \quad (9)$$

The element  $m_{ij}$  of the judgment matrix represents the ratio of the influence of the factor  $s_i$  on music emotion. Therefore, the elements on the main diagonal of the judgment matrix are all ones. In addition,

$$m_{ji} = \frac{1}{m_{ij}}. \quad (10)$$

Since the text of the lyrics is divided into four parts: main song, chorus, sublimation and ending, the value of  $n$  is four. Use a scale of 1–9 to describe the weight of a factor's influence on the judgment of music emotion. The larger the odd number is, the more important it is. Even numbers represent the middle value of adjacent judgments. Therefore,

the judgment matrix for the position factor of the feature words of lyrics in the article is as follows:

$$M = \begin{pmatrix} 1 & 9 & 8 & 7 \\ \frac{1}{9} & 1 & \frac{1}{3} & 1 \\ \frac{1}{8} & 3 & 1 & 3 \\ \frac{1}{7} & 1 & \frac{1}{3} & 1 \end{pmatrix}. \quad (11)$$

The value of the position factor can be obtained by calculating the maximum Eigen root of the judgment matrix and its corresponding normalized eigenvector. The specific process is as follows:

- (1) Normalize each column of the judgment matrix to get a new matrix
- (2) Find the sum of each row of the matrix, respectively
- (3) Calculate the proportion of each in the total, namely, the weight of the four positions of the main song, the chorus, the sublimation, and the ending in the lyrics to the classification of music emotion.

**3.3. Intelligent Classification Algorithm of Music Emotion.** The music emotion model means that we need to find a suitable classification model to identify the emotional category of music. This process mainly includes two stages: training stage and testing stage.

The main task of the so-called training phase is to learn the training set with known music text and its emotional label and find the model parameters until the model is built.

What needs to be done in the testing phase is to verify the learned model. The specific method is to separate the music text in the test set from the corresponding emotional labels, input the separated music samples with unknown emotional labels into the learned classification model for prediction, and compare the predicted emotional labels with the original emotional labels. The accuracy of the prediction is calculated to evaluate the classification performance of the model.

There are endless research methods for music sentiment analysis. The most common one is to use machine-learning algorithms combined with statistical knowledge for research. This article chooses support vector machines as our music classification model.

The structured risk minimization theory of support vector machines based on statistical principles has the following three advantages:

- (1) Supported by the structured risk minimization theory, it still has good generalization ability under a limited small sample.
- (2) As a classification-learning machine, it generally transforms actual problems into convex quadratic linear programming problems, which can obtain more accuracy.

- (3) Introduce the kernel function and linear relaxation to transform the nonlinear problem into a higher-dimensional space into a linearly separable problem. The calculation formula is only related to the number of samples and has nothing to do with the sample dimension, to avoid the increase in sample dimension.

The support vector machine intends to find a straight line  $L1$  to separate the two types of data. Assuming that the data set  $D = \{(a1, b1), \dots, (am, bm)\}$  of  $m$  samples, corresponding to the label  $\text{label} = i-1 (i = 1, \dots, m)$ , define the classification hyperplane expression as follows:

$$f(a) = \lambda^T a + b. \quad (12)$$

Among them,  $\lambda = (\lambda_1, \dots, \lambda_m)$  is the normal vector of the hyperplane and  $b$  is the intercept of the hyperplane. Let the straight line  $L1$  be  $\lambda^T a + b = 0$ , and move the straight line to both sides. If there are two types of samples that can be divided, the two types of samples must meet  $\lambda^T a + b \geq y$  after normalization. At this time, the samples on the two classification straight lines  $L2$  and  $L3$  are called support vectors. The distance between the straight lines  $d = 1/|\lambda|$ .

In summary, according to the structured risk minimization theory, the straight line  $L1$  separates the two types of samples, which satisfies the experience risk minimization of the learning model. The straight lines  $L2$  and  $L3$  maximize the interval between the two types of samples and meet the minimum confidence risk. The final problem is transformed into solve the problem of minimum  $\lambda$ :

$$\min \left( \frac{\|\lambda\|^2}{2} \right) \text{ s.t. } (\lambda^T a + b)^* y \geq 1. \quad (13)$$

In real scenes, there are still most linear inseparable problems. First, the original low-dimensional space is mapped to a higher latitude space through an appropriate mapping kernel function  $\text{kernel}(a)$  so that it is linearly separable at high latitudes, as shown in Figure 3.

Modify formula (13) at this time as follows:

$$\begin{cases} f(a) = \text{sigmoid} \left( \sum \lambda_i y_i (\text{kernel}(a) \cdot \text{kernel}(a) + b) \right), \\ b = y_i - \sum \lambda_i y_i (\text{kernel}(a) \cdot \text{kernel}(a)), \end{cases} \quad (14)$$

and by combining the relaxation coefficient and the kernel function, the soft-edge kernel function SVM effectively solves the nonlinear problem and greatly improves the generalization ability.

## 4. Results and Discussion

**4.1. Experimental Data Source and Preprocessing.** Since the research object of this paper is the multiemotion classification of songs, in order to ensure the credibility of the data, the text of lyrics under four categories in four music apps is captured. Among them, after removing the lyrics texts that are mixed in Chinese and English or mostly in English, 500 pieces of each category are, respectively, counted and 2,000 lyrics texts are used as the final data set of the experiment.

Chinese word segmentation adopts the precise pattern in stammering word segmentation. Segment the most accurate part of the lyrics text, and remove stop words, and eliminate ambiguous words. In terms of feature extraction, use CHI to calculate feature words and category chi-square values, sort, and construct a fixed-dimensional emotional dictionary to convert each lyrics text into a unified-dimensional word vector.

### 4.2. Performance Test of Intelligent Detection of Music Signal.

In order to analyse the effectiveness of the intelligent detection of music signals in the complex noise scenarios designed in this paper, simulation tests are carried out. The test platform is shown in Table 1.

In order to test the superiority of the music signal intelligent detection algorithm under the complex noise scene designed in this paper, the music signal intelligent detection algorithm in literature [11] and literature [12] is selected for comparison experiments, and their test platforms are the same. Using popular music as the object, the test data collect two types of music signals: one is the music signal that does not contain noise; the other is the music signal that contains noise. Each method carries out five simulation experiments, and the test data of electronic music signal are shown in Table 2.

Three kinds of music signal intelligent detection algorithms are used to detect noise-free and noisy music signals, respectively, and their accuracy results are shown in Figures 4 and 5. Comparing and analysing the detection accuracy of noise-free and noisy music signals in Figures 4 and 5, we can see the following:

- (1) For noise-free and noisy music signals, the average value of the detection accuracy of the music signal in literature [11] is 86.83% and 81.0%, respectively. The detection accuracy of noise-free music signals is slightly higher than that of music signals that contain noise. This is mainly because it does not introduce noise cancellation technology and cannot suppress the interference of noise on the detection result of the music signal. Therefore, the accuracy of music signal detection needs to be further improved.
- (2) For noise-free and noisy music signals, the average value of the detection accuracy of the music signal in literature [12] is 91.39% and 86.09%, respectively. The detection accuracy of the music signal is difficult to meet the requirements of the actual processing of the music signal, resulting in a relatively high error rate of the music signal detection.
- (3) Regardless of noise-free or noisy music signals, the average detection accuracy of the algorithm in this paper is higher than that of the algorithm in literature [11, 12]. This is because the algorithm in this paper introduces a noise elimination algorithm, which removes the noise in the music signal and obtains a high-quality music signal. At the same time, the problem of parameter optimization in the music signal process is solved, the detection error of

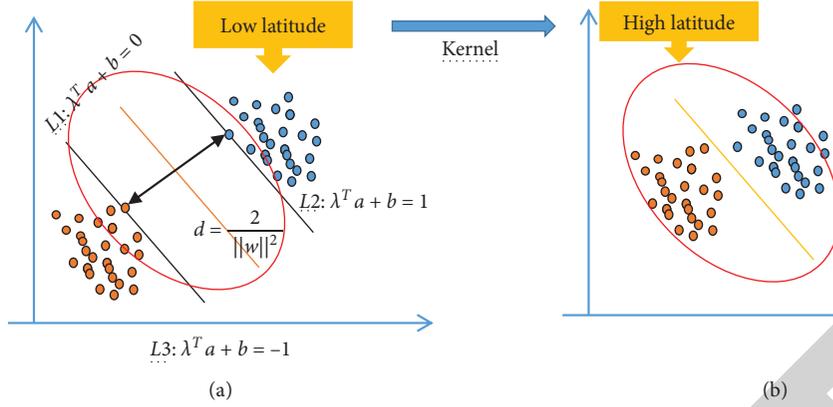


FIGURE 3: Linearly inseparable support vector machine model.

TABLE 1: Simulation test platform parameter setting.

| Platform parameters | Set up             |
|---------------------|--------------------|
| CPU                 | AMD R7 3700X       |
| Motherboard         | GIGABYTE H310M-s2  |
| SDD                 | Kingston 240 GB    |
| RAM                 | Kingston DDR4 8 GB |
| OS                  | Linux              |

TABLE 2: Test data for music signal detection.

| Number | Number of samples without noise | Number of samples with noise |
|--------|---------------------------------|------------------------------|
| 1      | 200                             | 50                           |
| 2      | 400                             | 80                           |
| 3      | 300                             | 100                          |
| 4      | 200                             | 80                           |
| 5      | 200                             | 50                           |

the music signal is reduced, and the ideal music signal detection result is obtained, which can meet the actual requirements of music signal processing.

For modern mass music signals, detection efficiency is very important. Therefore, the signal detection time of each experiment is counted, and the results are shown in Figures 6 and 7. From Figures 6 and 7, it can be found that for noise-free and noisy music signals, the average value of music signal detection time in [11] is 2.97 s and 6.20 s, respectively. The average value of signal detection time in literature [12] is 4.93 s and 7.02 s, respectively. The average detection time of the algorithm in this paper is 2.63 s and 4.63 s, respectively. The detection time of the algorithm in this paper is significantly reduced, and the detection efficiency of music signal is improved correspondingly, which is adapted to the large-scale development of modern music signal.

**4.3. Optimization of Support Vector Machine Parameter Model.** The optimization algorithm mainly tunes the  $c$  and  $g$  initialization parameters of SVM and strives to find the global optimal solution through the best initial value. In this paper, three-optimization algorithms of genetic, particle

swarm, and grid search are used to compare the characteristic parameters under 5-fold cross-validation.

According to experience, due to the existence of the cross-mutation process, the optimization model based on the genetic algorithm (GA) consumes the most time, while the particle swarm optimization (PSO) operation is simpler and has fast convergence speed, which is suitable for parameter tuning. Therefore, this paper selects an optimization algorithm based on particle swarm optimization (PSO) to conduct comparative experiments on the two parameters of population size and termination algebra. The experimental results obtained by default are also applicable to genetic algorithms within a certain range. First, do a comparison experiment on the termination algebra, set the population size to 20, and select 50, 100, and 200 generations for comparison experiments. The results are shown in Figure 8(a).

It can be seen that with the deepening of the termination algebra, various indicators have an upward trend, but the improvement effect is only 0.7% at most. In contrast, the time cost has doubled from 556s to 2218s. Considering the recognition rate and time cost, 100 is selected as the termination algebra of PSO and GA optimization algorithms, then continue to consider the impact of the population size on the model performance, and compare the PSO optimization algorithm models with population sizes of 10, 20, and 50. The results are shown in Figure 8(b). From 10 to 50 populations, the three evaluation indicators all reach the best when the number of populations is 20, and the recognition performance shows a trend of increasing first and then decreasing. Based on the analysis of the above results, moderately deepening the termination algebra can indeed improve the performance of the model. However, if the sample size is large, the time cost must be considered. As for the population size, it does not need to increase blindly, and the most suitable parameters can be selected according to the actual comparison experiment results. Finally, the PSO and GA algorithms with 100 generations and 20 populations as the initial parameters are compared with the grid optimization algorithm, and the model optimization capabilities of the three algorithms are examined. The experimental results are shown in Figure 8(c).

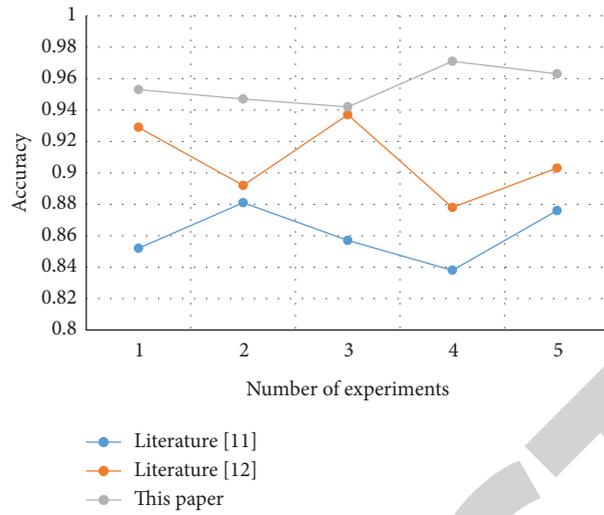


FIGURE 4: Accuracy of noise-removed music signals.

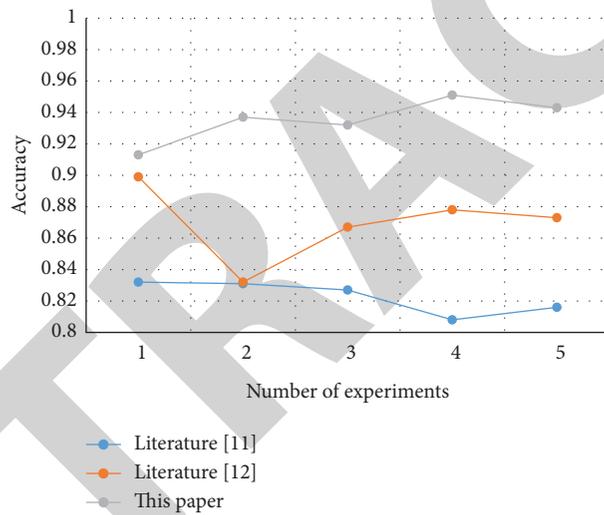


FIGURE 5: Accuracy of noisy music signals.

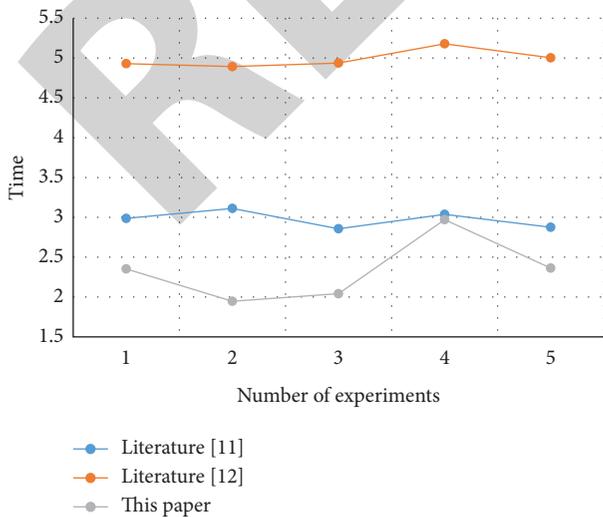


FIGURE 6: Time of noise-removed music signals.

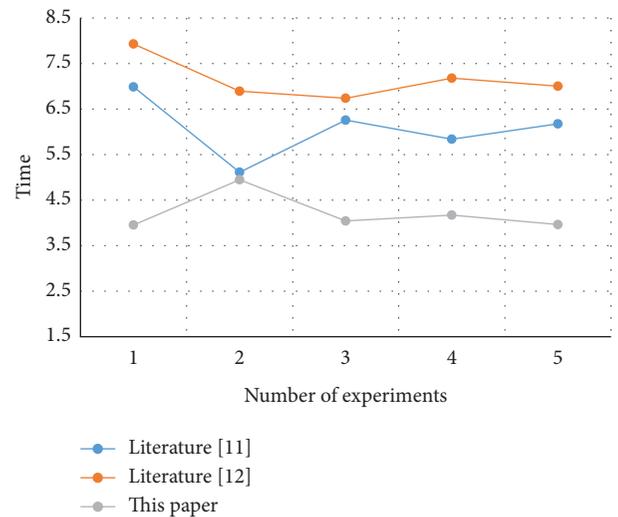


FIGURE 7: Time of noisy music signals.

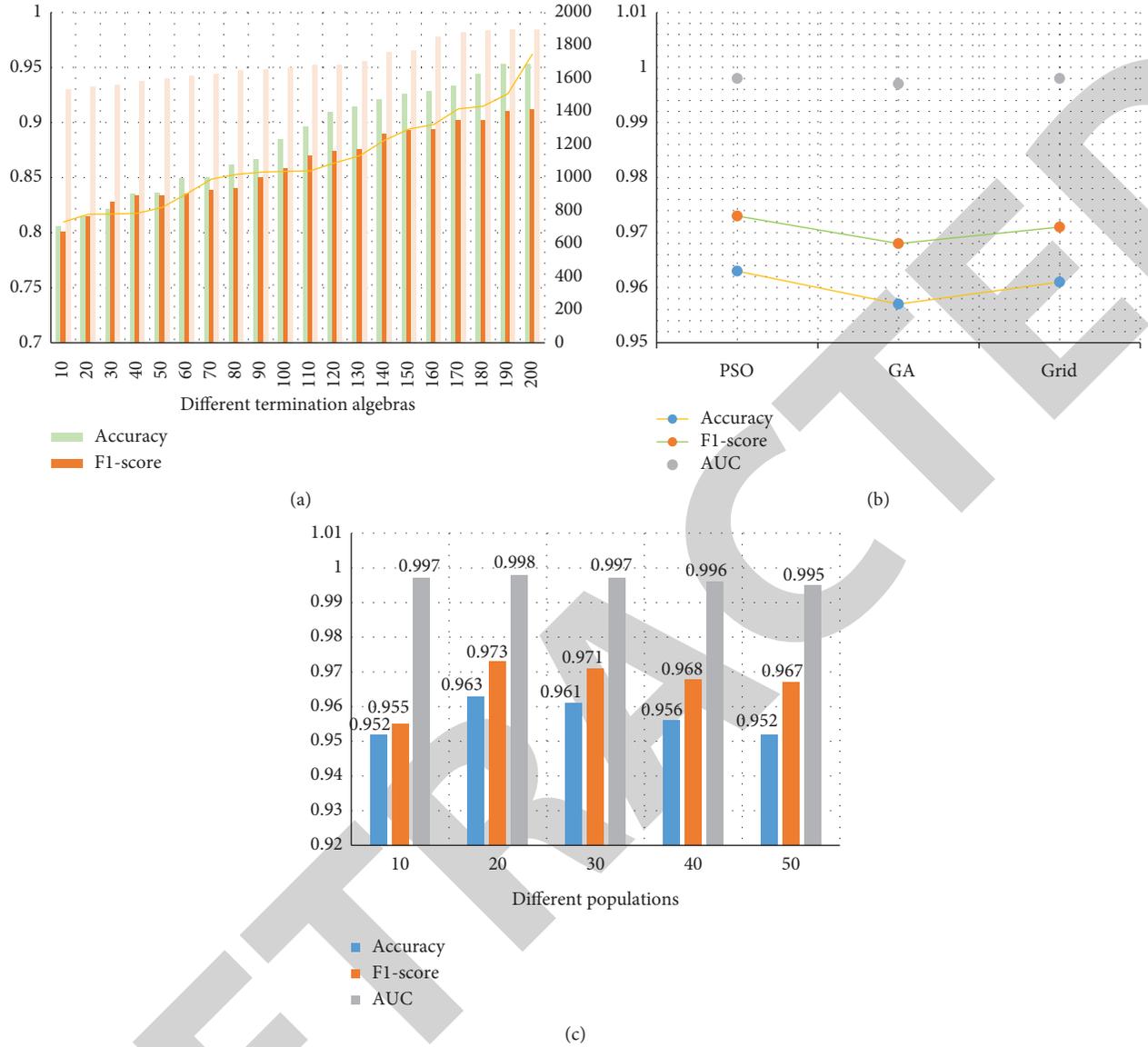


FIGURE 8: Comparison of results in different situations.

It has been verified that the classification and recognition performance of the three algorithms is not much different, and the PSO algorithm with the best recognition performance is only 0.64% higher than the worst GA algorithm. In terms of training time consumption, grid is less than 5 minutes and far less than the previous two algorithms. Therefore, in the end, this paper selects PSO and grid of 100 generations and 20 populations as the initial parameter settings for the characteristic parameter comparison experiment. Figure 9

shows the curves automatically searched by the three-optimization algorithms after a certain experiment.

#### 4.4. Performance Analysis of Feature Word Location Factors.

It is assumed that the feature word position factor of the title is the largest, the feature word position factor of the middle position of the lyrics text is the second, and the feature word position factor of the two positions at the beginning and end is the smallest and equal. The result is shown in Figure 10.

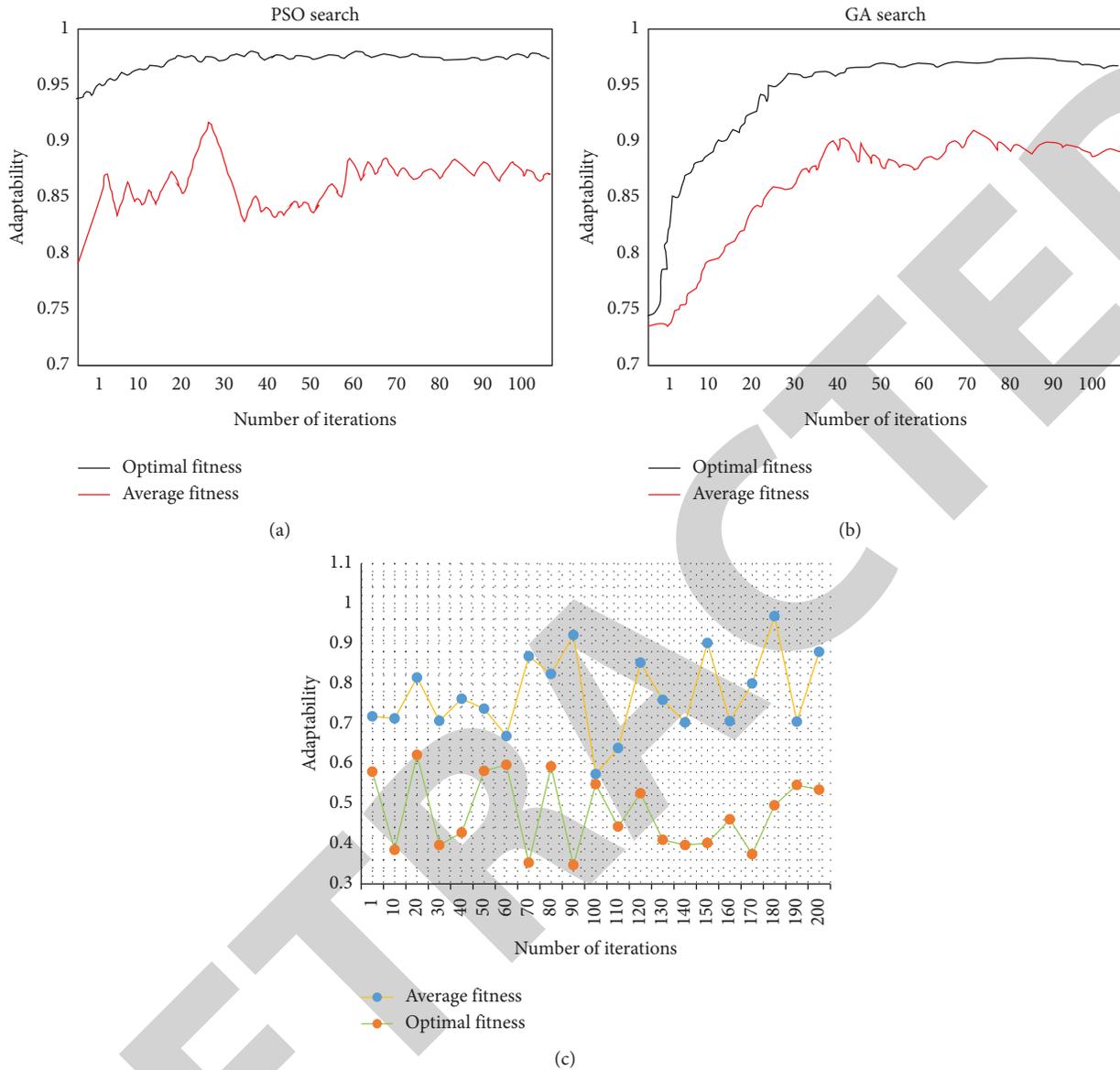


FIGURE 9: Fitness convergence curve.

After the calculation of the TTFL algorithm, it can be seen that even though the classification accuracy of the naive Bayes model is 86%, it is still less than the classification

accuracy of the support vector machine model. It is not difficult to see that location still plays an important role in the emotional classification of music based on lyrics.

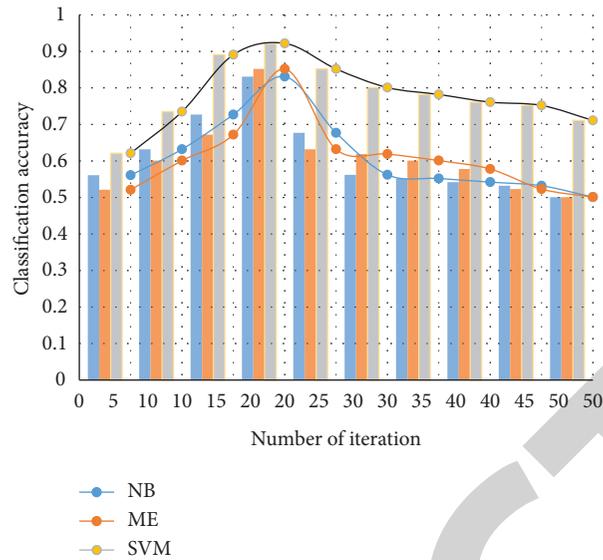


FIGURE 10: Performance comparison of lyrics-based music sentiment classification under the same sentiment dictionary.

## 5. Conclusion

Due to the interference of noise, the current music classification and detection models cannot obtain ideal music classification and detection results, resulting in a high error rate of music classification and detection. Considering the influence of the weight of feature words in different positions on the classification of lyrics, AHP is used to calculate the position weight of feature words in different positions of text title and lyrics before, in, and after the text. In order to obtain better music classification and detection results, this paper first proposes an intelligent detection algorithm of electronic music signals under complex noise scenes. Denoising technology is introduced to eliminate the noise and extract the features from the signal. Second, from the perspective of audio and lyrics of song sentiment analysis and the unique characteristics of lyrics text, a lyric sentiment analysis method based on text title and position weight is proposed. The results show that the proposed model is a high precision and robust model for music classification and detection and has a very wide application prospect.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This paper was the project of science and education joint fund of Hunan Natural Science Foundation in 2020: Research on the new curriculum development of Hunan

original folk songs from the perspective of excellent cultural heritage (project no. 2020j7016) (initial results).

## References

- [1] S. Koelsch, P. Vuust, and K. Friston, "Predictive processes and the peculiar case of music," *Trends in Cognitive Sciences*, vol. 23, no. 1, pp. 63–77, 2019.
- [2] A. J. Sihvonen, T. Särkämö, V. Leo, M. Tervaniemi, E. Altenmüller, and S. Soinila, "Music-based interventions in neurological rehabilitation," *The Lancet Neurology*, vol. 16, no. 8, pp. 648–660, 2017.
- [3] N. Ding, A. D. Patel, L. Chen, H. Butler, C. Luo, and D. Poeppel, "Temporal modulations in speech and music," *Neuroscience & Biobehavioral Reviews*, vol. 81, pp. 181–187, 2017.
- [4] Y. M. G. Costa, L. S. Oliveira, and C. N. Silla, "An evaluation of Convolutional Neural Networks for music classification using spectrograms," *Applied Soft Computing*, vol. 52, pp. 28–38, 2017.
- [5] J. Nam, K. Choi, J. Lee et al., "Deep learning for audio-based music classification and tagging: teaching computers to distinguish rock from bach," *IEEE Signal Processing Magazine*, vol. 36, no. 1, pp. 41–51, 2018.
- [6] M. Farrokhmanesh and A. Hamzeh, "Music classification as a new approach for malware detection," *Journal of Computer Virology and Hacking Techniques*, vol. 15, no. 2, pp. 77–96, 2019.
- [7] M. Bhattacharjee, S. R. M. Prasanna, and P. Guha, "Speech/music classification using features from spectral peaks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1549–1559, 2020.
- [8] G. K. Birajdar and M. D. Patil, "Speech and music classification using spectrogram based statistical descriptors and extreme learning machine," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 15141–15168, 2019.
- [9] J. Vlegels and J. Lievens, "Music classification, genres, and taste patterns: a ground-up network analysis on the clustering of artist preferences," *Poetics*, vol. 60, pp. 76–89, 2017.

- [10] J.-H. Wu, W. Wei, L. Zhang et al., "Risk assessment of hypertension in steel workers based on LVQ and Fisher-SVM deep excavation," *IEEE Access*, vol. 7, pp. 23109–23119, 2019.
- [11] L. Zhang, P. Shen, X. Peng et al., "Simultaneous enhancement and noise reduction of a single low-light image," *IET Image Processing*, vol. 10, no. 11, pp. 840–847, 2016.
- [12] B. Zhou, C.-L. Mu, J. Feng, and W. Wei, "Continuous level anisotropic diffusion for noise removal," *Applied Mathematical Modelling*, vol. 36, no. 8, pp. 3779–3786, 2012.
- [13] T. Emoto, U. R. Abeyratne, K. Kawano, T. Okada, O. Jinnouchi, and I. Kawata, "Detection of sleep breathing sound based on artificial neural network analysis," *Biomedical Signal Processing and Control*, vol. 41, pp. 81–89, 2018.
- [14] K. Guo, R. Chai, H. Candra et al., "A hybrid fuzzy cognitive map/support vector machine approach for EEG-based emotion classification using compressed sensing," *International Journal of Fuzzy Systems*, vol. 21, no. 1, pp. 263–273, 2019.
- [15] L.-K. Zhang, S.-Q. Sun, B.-X. Xing, R.-M. Luo, and K.-J. Zhang, "Using psychophysiological measures to recognize personal music emotional experience," *Frontiers of Information Technology & Electronic Engineering*, vol. 20, no. 7, pp. 964–974, 2019.
- [16] B. Y. Jang, W. H. Heo, J. H. Kim et al., "Music detection from broadcast contents using convolutional neural networks with a Mel-scale kernel," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 1, pp. 1–12, 2019.
- [17] R. Sarkar, S. Choudhury, S. Dutta, A. Roy, and S. K. Saha, "Recognition of emotion in music based on deep convolutional neural network," *Multimedia Tools and Applications*, vol. 79, no. 1-2, pp. 765–783, 2020.
- [18] S.-W. Hsiao, S.-K. Chen, and C.-H. Lee, "Methodology for stage lighting control based on music emotions," *Information Sciences*, vol. 412-413, pp. 14–35, 2017.
- [19] J. Chen, D. Jiang, Y. Zhang, and P. Zhang, "Emotion recognition from spatiotemporal EEG representations with hybrid convolutional recurrent neural networks via wearable multi-channel headset," *Computer Communications*, vol. 154, pp. 58–65, 2020.
- [20] M. Hamada, B. B. Zaidan, and A. A. Zaidan, "A systematic review for human EEG brain signals based emotion classification, feature extraction, brain condition, group comparison," *Journal of Medical Systems*, vol. 42, no. 9, pp. 1–25, 2018.
- [21] X. Ke, Y. Zhu, L. Wen, and W. Zhang, "Speech emotion recognition based on SVM and ANN," *International Journal of Machine Learning and Computing*, vol. 8, no. 3, pp. 198–202, 2018.
- [22] H. Zhu, Q. Liu, N. J. Yuan, K. Zhang, G. Zhou, and E. Chen, "Pop music generation," *ACM Transactions on Knowledge Discovery from Data*, vol. 14, no. 5, pp. 1–31, 2020.
- [23] J. H. Su, T. P. Hong, Y. H. Hsieh et al., "Effective music emotion recognition by segment-based progressive learning," in *Proceedings of the 2020 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 3072–3076, Toronto, Canada, October 2020.
- [24] M. Ahuja and A. L. Sangal, "Opinion mining and classification of music lyrics using supervised learning algorithms," in *Proceedings of the 2018 first international conference on secure cyber computing and communication (ICSCCC)*, pp. 223–227, Mumbai, India, December 2018.
- [25] S. H. Sun and M. S. Cuthbert, "Emotion painting: lyric, affect, and musical relationships in a large lead-sheet corpus," *Empirical Musicology Review*, vol. 12, no. 3-4, pp. 327–348, 2018.
- [26] P. Song, C. Geng, and Z. Li, "Research on text classification based on convolutional neural network," in *Proceedings of the 2019 international conference on computer network, electronic and automation (ICCNEA)*, pp. 229–232, Xi'an, China, July 2019.
- [27] H. Gao, X. Zeng, and C. Yao, "Application of improved distributed naive Bayesian algorithms in text classification," *The Journal of Supercomputing*, vol. 75, no. 9, pp. 5831–5847, 2019.
- [28] H. C. Wang, S. W. Syu, and P. Wongchaisuwat, "A method of music autotagging based on audio and lyrics," *Multimedia Tools and Applications*, vol. 21, pp. 1–29, 2021.
- [29] D. Wang, J. Wang, F. Hu, L. Li, and X. Zhang, "A locally adaptive multi-label  $k$ -nearest neighbor algorithm," in *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining*, vol. 45, pp. 81–93, Delhi, India, May 2018.
- [30] R. Panda, R. Malheiro, and R. P. Paiva, "Novel audio features for music emotion recognition," *IEEE Transactions on Affective Computing*, vol. 11, no. 4, pp. 614–626, 2018.
- [31] M. Skoro and A. Roncevic, "The music industry in the context of digitization," *Economic and Social Development*, vol. 23, pp. 279–288, 2019.
- [32] N. Mapuweyi and G. Kemunto-Orina, "An investigation into the prospects and challenges faced by Zimbabwean and Kenyan musicians and music production houses at the advent of digitization and digitalization in Africa," *Journal of Emerging Trends in Educational Research and Policy Studies*, vol. 9, no. 5, pp. 207–214, 2018.
- [33] F. Noroozi, D. Kaminska, T. Sapinski, and G. Anbarjafari, "Supervised vocal-based emotion recognition using multiclass support vector machine, random forests, and adaboost," *Journal of the Audio Engineering Society*, vol. 65, no. 7/8, pp. 562–572, 2017.
- [34] D. Chaudhary, N. P. Singh, and S. Singh, "Development of music emotion classification system using convolution neural network," *International Journal of Speech Technology*, vol. 18, pp. 1–10, 2020.
- [35] C. Huang, "Recognition of psychological emotion by EEG features," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 10, no. 1, pp. 1–11, 2021.
- [36] Y. Ni, M. McVicar, R. Santos-Rodriguez, and T. De Bie, "An end-to-end machine learning system for harmonic analysis of music," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1771–1783, 2012.
- [37] B. L. Sturm, O. Ben-Tal, Ú. Monaghan et al., "Machine learning research that matters for music creation: a case study," *Journal of New Music Research*, vol. 48, no. 1, pp. 36–55, 2019.