Hindawi
Wireless Communications and Mobile Computing
Volume 2021, Article ID 9298654, 7 pages
https://doi.org/10.1155/2021/9298654

Research Article

Music Style Classification Algorithm Based on Music Feature Extraction and Deep Neural Network

Kedong Zhang

Department of Vocal Music, Xi'an Conservatory of Music, Shaanxi Province, Xi'an 710061, China

Correspondence should be addressed to Kedong Zhang; 1424092523@qq.com

Received 21 July 2021; Revised 10 August 2021; Accepted 15 August 2021; Published 6 September 2021

Academic Editor: Yuanpeng Zhang

The music style classification technology can add style tags to music based on the content. When it comes to researching and implementing aspects like efficient organization, recruitment, and music resource recommendations, it is critical. Traditional music style classification methods use a wide range of acoustic characteristics. The design of characteristics necessitates musical knowledge and the characteristics of various classification tasks are not always consistent. The rapid development of neural networks and big data technology has provided a new way to better solve the problem of music-style classification. This paper proposes a novel method based on music extraction and deep neural networks to address the problem of low accuracy in traditional methods. The music style classification algorithm extracts two types of features as classification characteristics for music styles: timbre and melody features. Because the classification method based on a convolutional neural network ignores the audio's timing. As a result, we proposed a music classification module based on the one-dimensional convolution of a recurring neuronal network, which we combined with single-dimensional convolution and a two-way, recurrent neural network. To better represent the music style properties, different weights are applied to the output. The GTZAN data set was also subjected to comparison and ablation experiments. The test results outperformed a number of other well-known methods, and the rating performance was competitive.

1. Introduction

Music is an audio signal composed of a specific rhythm, melody, harmony, or musical instrument fusion according to a certain rule, and it is an art that contains and reflects human emotions [1–3]. The different characteristics formed by the unique beats, timbres, tunes, and other elements in musical works are called music styles [4–6], such as common rock music [7], classical music [8], and jazz. In recent years, with the rapid development and innovation of the Internet and multimedia technologies [9–11], digital music [12, 13] has long become the main form of people listening to music, which also promotes the increasing demand for music appreciation. Music style is now one of the most commonly used classification attributes for the management and storage of digital music databases, and it is also one of the main classification search items used by most online music websites. The efficiency of the manual labeling method used in early music information retrieval can no longer meet the needs of management in the face of massive music data, and it is very easy to consume a lot of manpower and time. As a result, studying music style classification algorithms is critical in order to achieve the goal of automatic music style classification [14–16].

Music style classification is an important branch in the field of music information retrieval that has been studied in depth because the automatic algorithm of music style classification has the abovementioned practical value. Digital signal processing random process, music theory, and other theories are primarily used in algorithmic research of music style classification to mathematically describe and express genre-related characteristics in music signals to form various types of music characteristics. The machine learning algorithm [17–19] is used to learn the feature distribution characteristics of different genres to obtain the classifier. Finally, the feature of a piece of audio signal is given as the input of the classifier, and the style of the classifier is determined according to the posterior probability. Among them, the

structure of the feature determines the upper limit of the performance of the classification algorithm, and an effective representation method can maximize the accuracy of the classification result. Therefore, a large number of scholars focus on the feature engineering link of music signals. However, there are two main difficulties in the study of music signals: on the one hand, music contains complex and abstract information such as emotions, rhythms, instruments, and chords, which are often difficult to express in artificially constructed features; on the other hand, music contains complex and abstract information. Compared with ordinary voice signals, music has more complex frequency composition and richer timbre information. Therefore, some conventional processing methods of voice signal processing cannot be simply applied, and special algorithms need to be designed according to the characteristics of music signals.

In recent years, deep learning [20–22] has achieved outstanding results in the fields of image [23], speech, and natural language processing. More and more researchers are trying to learn a good feature expression of music signals through deep neural networks, replacing the previous manual extraction. The characteristics of improving the performance of the algorithm have important theoretical value. At present, Spotify, the world's largest genuine streaming music service platform, has successfully applied deep learning to its music recommendation system. Therefore, music signal processing based on deep learning can promote the development of music platforms and provide users with a better service experience and has huge economic value and research value.

The main innovations of this article are as follows:

(1) This paper proposes a novel music style classification algorithm based on music feature extraction and deep neural network, which can effectively improve the performance of music style classification

(2) This paper takes two types of features, the tonic and the melody feature, as the parameters of classification in the music style, since the method of classifying music based on convolutionary neural nets overlooks the time sequence of the audio itself. So, we combined the proposed convolution structure with the single-dimensional convolution and two-way, recurrent neural network and proposed a music classification module that would rely on the one-dimensional convolution of a recurring neuronal network. The output is given different weights of attention in order to better represent the music features

(3) Comparison and ablation experiments were also conducted on the GTZAN dataset. The experimental results surpassed other well-known methods and achieved competitive classification performance

2. Related Work

Different genres have distinct musical styles, and the identification of musical styles or musical genres has been extensively researched since its inception. People in other countries have been using artificial methods to judge music genres and styles since the 1990s. The "Music Chromosome Project" is the most well-known. The main goal of this project is for music experts to divide music into different types based on their knowledge and understanding of music technology. However, when faced with massive amounts of data, artificial methods are immature due to the limited technical conditions, and different experts have slightly different understandings of different music genres, so the project has spent a lot of financial and material resources. Driven by this situation, people began to try the research of automatic classification algorithms for music genre recognition.

Later, American researchers proposed a classification algorithm. This method mainly calculates the mean, variance, and autocorrelation coefficient from massive music data, so as to further analyze the characteristics of music, such as loudness and pitch, people can easily feel. The obtained features are then identified and classified by using some classifiers. Subsequently, this algorithm has been greatly promoted, and people have begun to try to use some improved algorithms to classify music genres based on this algorithm. In 2002, Tzanetakis and Cook [24] provided a new classification algorithm. This method first extracts acoustic features, which mainly contains three types of acoustic features, music timbre, rhythm, and pitch content. Since the extracted acoustic features are generally of higher dimensionality, the feature selection algorithm is used to reduce the dimensionality of the features to facilitate calculation, and at the same time, some insignificant redundant information is removed. Finally, some models and corresponding algorithms are used to identify and classify music genres.

With the development of computer technology, machine learning has also begun to be applied to the classification of music genres. In 2003, Xu et al. [25] studied the classification of music genres by using different music characteristics. By comparing the K-nearest neighbor method, conditional random field, and Markov model algorithms, they found that the recognition effect classification algorithm of SVM is the most effective. With the application of wavelet transform theory, Li et al. [26] used statistical methods to calculate the statistical values of wavelet coefficients, combined with classification models commonly used in machine learning, such as LDA, GMM, and KNN, to obtain good classification results. In 2011, in order to obtain more essential music features, Panagakis et al. [27] proposed an unsupervised dimensionality reduction method for the first time. Through experimental results, it was found that this method has a significant effect in extracting music features compared to previous methods.

3. Methodology

3.1. Elements of Music. Music contains three elements: pitch, rhythm, and timbre. Melody and harmony of music can be formed through the combination and transformation of pitch; tempo is related to articulation, which controls the speed and transition of music; timbre is the sound quality

of sound perception, used to distinguish different types of sounds to produce notes, and each instrument has its own unique timbre. The combination of these three elements can form other elements in music. For example, a number of different pitches played at the same time become harmony, and the coordinated effect obtained by different pitches at different times becomes a melody. These elements form a unique style of music through different combinations, conveying joy, excitement, sadness, and other emotions, thus, forming different genres.

Vocal music, which is based on vocal singing, and instrumental music, which is based on instrument performance, are the two main types of music. Chorus and solo in vocal music, as well as solo, concerto, and symphony in instrumental music, are examples of these two forms. Instrumental and vocal music can be combined to create a wide range of musical expressions.

3.2. Music Feature Extraction. This paper uses multiple feature extraction methods to extract the timbre features of the bottom music features (Mel cepstrum coefficients) and the melody features of the middle music features (pitch frequency, formant, and band energy) from the original audio signal and then will be composed of these features. The training set explains how to use the training classification system to improve the classification system's accuracy.

3.2.1. Mel Cepstral Coefficient. The Mel cepstrum coefficient simulates the characteristics of human hearing and conforms to the characteristics of human hearing. It has good antinoise ability and high recognition rate. In the current speech signal research, it has become a widely used characteristic parameter [28]. First, import the audio signal, perform frame processing and windowing on the signal, and use Fourier transform to transform the time domain signal into the frequency domain signal:

$$x(k) = \sum_{k=0}^{N-1} x(n) e^{-j2\pi nk/N}, 0 \le k \le N-1. \quad (1)$$

The input signal is represented by x, and the signal input strength at n is represented by $x(n)$, $n = 0, 1, \cdots, J$, where J is the signal length. In the discrete Fourier transform, the number of points it performs is denoted by N.

The energy spectrum should be calculated and then transferred. A set of Mel scale triangle filters are used to implement the transfer method. A key parameter of this form of filter is the center frequency, which is denoted as $f(m)$. For each triangle filter, the output energy of the group is calculated and expressed in logarithm, then,

$$S(m) = \ln\left(\sum_{k=1}^{N-1} |x(k)|^2 H_m(k)\right), 0 \le m \le M-1. \quad (2)$$

Second, to calculate the MFCC parameters, the way to achieve it is to perform a discrete cosine transform (DCT):

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos\left(n\pi\left(m - \frac{0.5}{m}\right)\right), n = 1, 2, \cdots, L. \quad (3)$$

3.2.2. Pitch Frequency. An audio signal consisting of various tones may be seen as an audio sequence. The tone fluctuation contains the composer's emotion when creating this piece of music, and the tone is determined by the tone frequency. The pitch frequency is the voice; therefore, this is a very important signal processing parameter. The pitch frequency extraction must take into account the short-term stability of the speaker signal. Currently, the most common methods are the autocorrelation detection (ACF), the average amplitude difference (AMDF), and peak removal, etc. In this paper, the autocorrelation function detection method is chosen to extract the frequency of pitch in view of the stability and smoothness of the pitch signal. The short-term autocorrelation function $Rn(k)$ of the speech signal $s(m)$ is defined as:

$$R_n(k) = \sum_{m=0}^{N-k-1} S_n(m) S_n(m+k), \quad (4)$$

where N is the length of the window added by the speech signal; $s_n(m)$ is a segmented window speech signal intercepted by speech signal $s(m)$ through a window with window length N and is defined as:

$$S_n(m) = s(m) w(n-m). \quad (5)$$

The autocorrelation function of the fundamental part of the audio clip will have obvious peaks, and the high-frequency tones are not obvious compared to the fundamental. Therefore, judging whether it is a fundamental tone or a high-frequency tone can be determined by detecting whether there is an obvious peak, and the pitch frequency can be extracted by detecting the distance between adjacent peaks.

3.2.3. Resonance Peak. Resonance frequency is another name for formant. It refers to the phenomenon in which the energy contained in a particular sound channel is increased as a result of the audio signal's resonance phenomenon. The vocal tract can usually be regarded as a uniformly distributed sound tube, and the resonance of the sound tube vibration in different positions is the sound process. The shape of the formant is usually related to the structure of the vocal tract. As the structure of the vocal tract changes, the shape of the formant will also change. For a segment of speech signal, different emotions correspond to different channel shapes. Therefore, the formant frequency can be used as an important parameter of speech signal emotion recognition.

3.2.4. Frequency Band Energy Distribution. Band energy distribution refers to the distribution of energy possessed by a segment of audio signal, which contains information such as the strength and frequency of the audio signal. It has a strong correlation with the sweetness of music and the emotion of music. In the field of music, through the analysis of the energy distribution characteristics of the frequency band, the pleasantness and emotional characteristics of the audio

signal can be obtained. Suppose there is a music segment of length M, which contains the voice characteristics of various instruments and human voices. Now we want to find one of the subbands in the frequency domain of the music segment, from a to $a + N$, which contains energy of. First, according to the Fourier transform, the original time-domain music signal $f(t)$ is converted to the frequency-domain signal $F(t)$.

$$F(t) = \int f(t)e^{-jwt}dt. \tag{6}$$

The band energy E is equal to:

$$E = \frac{1}{N}\sum_{a}^{a+N} |F(t)|^2. \tag{7}$$

3.3. Classification Model. The range of sounds is sequential in the time dimension, and the timing information inside the music is ignored by the simple use of the convolution structure. A one-dimensional configuration takes place within the dimension of time and also ignores the sequence relationship between the sound spectrum properties of different time frames while capturing the local sound spectrum characteristics. The music sequence relationship cannot be effectively modelled only by one-dimensional convolution. Thus, we combined the proposed convolution structure with one DNA and a two-way recurrent neural net and proposed a classification module based on a recurring one DNA network and used the mechanism of attention to move the neural network at different times. The output has different attention weights, so that the characteristics of the music style are better represented. The recurrent neural double-way network, in particular, summarizes time domain data so that the model can learn about music's time sequence. Because the musical characteristics of a piece of music can have different effects on a musical category at different times, the attention mechanism is used to assign different weights of attention to the cyclic neural network output at different times and to combine sequence characteristics.

3.3.1. Bi-RNN. RNN can capture the internal structure hidden in the sequence over time. The audio signal itself can be regarded as a time sequence. Using RNN to process music can capture the spatial dependence of the audio signal in the time dimension. The sound spectrum is also expanded in the time dimension. The feature map after one-dimensional convolution can be regarded as a time feature sequence, so the use of RNN to process the sound spectrum features can also play the same role. In order to better capture the multidirectional dependence in the time dimension in the music feature sequence, and be close to the brain's perception of music, this paper uses Bi-RNN to model [29] the music sequence.

Bi-RNN not only considers the previous input but also the latter input may also be helpful for data modeling. Figure 1 shows the structure of Bi-RNN. In the forward calculation, $\overrightarrow{H}^i$ is related to $\overrightarrow{H}^{i-1}$, and in the reverse calculation,

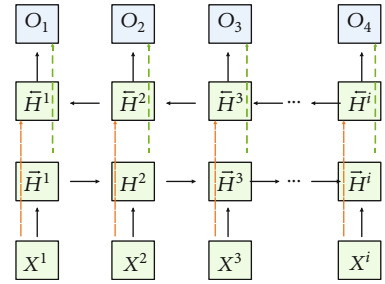

Figure 1: Schematic diagram of Bi-RNN.

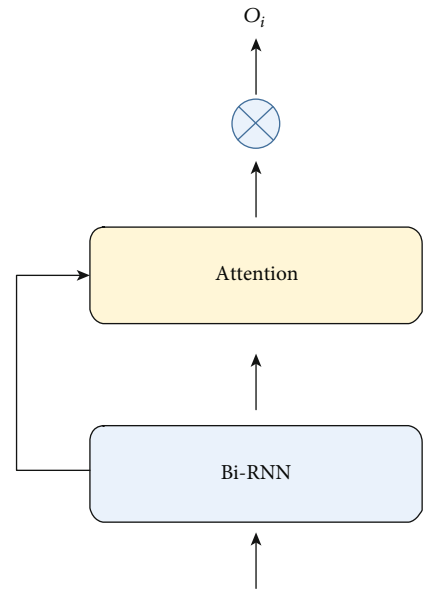

Figure 2: Schematic diagram of attention mechanism of serial structure.

$\overleftarrow{H}^i$ is related to $\overleftarrow{H}^{i+1}$, and $\overleftarrow{H}^i$ represents the state of the hidden layer. The calculation equation of H6 is as follows:

$$\overleftarrow{H}^i = f\left(W'X^i + V'\overleftarrow{H}^{i+1}\right). \tag{8}$$

Next, add the forward and the back of each network step to achieve the final network output:

$$O_i = U\overrightarrow{H}^i + U'\overleftarrow{H}^i. \tag{9}$$

3.3.2. Attention Mechanism. For the music category corresponding to the feature, the specific sound spectrum feature that appeared during different music times may differ. The focus mechanism [15] can each time compute the weight of the characteristics sequence and weightedly sum up the characteristics by weight each time. The overall feature of the music is the fully connected layer. The summarized function is represented.

This paper proposes an attention model with a serial structure, as shown in Figure 2. Since the output O_i of Bi-RNN represents the feature representation learned in the classification model. The attention model uses linear

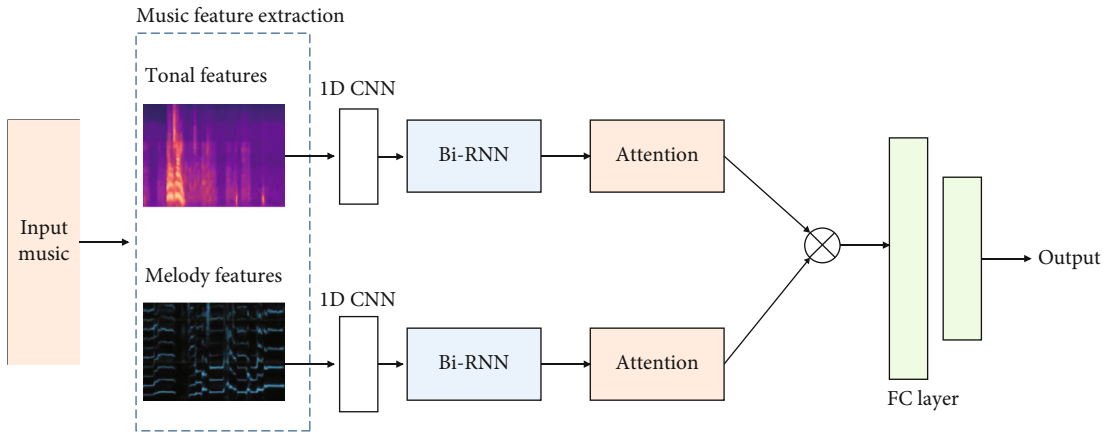

Figure 3: Schematic diagram of the overall structure.

transformation to calculate the attention score. The calculation formula is as follows:

$$e_i = w_i^T O_i, \tag{10}$$

where e represents the attention score assigned to the i-th feature vector, and $E = [e_1, e_2, \cdots e_T]$. O_i is the i-th feature vector. Then, we normalize the obtained attention score to generate the attention probability distribution on the feature representation:

$$a_i = \text{soft max}\,(E) = \frac{\exp\,(e_i)}{\sum_{j=1}^{T} \exp\left(e_j\right)}, \tag{11}$$

where a_i represents the attention probability assigned by the attention model to the i-th feature vector in the feature representation.

3.3.3. Overall Structure. Once the convolutionary layer is learned from various sound spectrums, feature maps with abstract features of high levels can now be obtained. The function maps can be extended in time to achieve sequences of converting features, and the convolution sequences for the music sequence modeling are entered in Bi-RNN. Then using the weight of the focus, the network has learned to weight the feature sequence performed by the Bi-RNN in summary, integrate the output of the Bi-RNN at several times, and translate it to the fully connected layer of the music. The overall network structure is shown in Figure 3 to further learn how to obtain classification results.

4. Experiments

4.1. Lab Environment. This article's hardware configuration is as follows: 16 GB memory, Intel Core i7-7700 processor, and NVIDIA 1050ti graphics card. This article's software environment includes the Windows 10 operating system, the Python programming language, the pycharm development environment, and the librosa voice extraction toolkit.

4.2. Dataset. As a standard data set in the field of music genre classification, the GTZAN data set is widely used to measure the accuracy of the classification method. The data set contains 10 music genres, such as classical, country, and jazz. It contains 1,000 excerpts of songs, and these 1,000 excerpts are evenly distributed among 10 music genres. The duration of each excerpt is approximately for 30 s. In order to ensure the sound quality of various recordings, the excerpted music clips are taken from wireless radios, CDs, and MP3 compressed audio files. Each audio file is stored in a 22050 Hz, 16-bit mono format.

4.3. Experimental Results. We compared the proposed algorithm with some well-known music style classification methods on the GTZAN dataset. The experimental results are shown in Table 1.

It can be seen from Table 1 that the RDNN network without convolutional structure shows the relatively worst classification effect on the GTZAN dataset, while the rest of the networks all adopt convolutional structure for abstract feature extraction, indicating that the convolutional structure can improve the feature extraction ability of the network model on the acoustic spectrum. Although KCNN adopts convolutional structure, it just carries out simple stacking of convolutional units, and its classification performance is inferior to that of NNET2 and NET1 with residual structure. However, the proposed algorithm in this paper adopts a combination of one-dimensional convolutional cyclic neural network and attention mechanism and carries out multifeature extraction. The network can extract the audio features that are more relevant to the music category and obtain the best classification performance.

4.4. Ablation Experiments. In order to further verify the influence of tonal and melody features on the performance of the proposed method, an ablation experiment is set up in this section. Only-tonal means that only tonal features are used, and only-melody means that only melody features are used. The experimental results are shown in Table 2.

It can be seen from Table 2 that only the tonal feature is used to obtain the performance second only to the proposed

Table 1: Comparative experiment results.

Method	Acc
Net1	0.9071
Nnet2	0.8715
KCNN	0.8368
RDNN	0.9301
Ours	0.9199

Table 2: Ablation experiment results.

Method	Acc
Only-tonal	0.9171
Only-melody	0.8915
Tonal and melody	0.9199

algorithm, while the melody feature alone has a greater impact on the performance. Therefore, this shows that the melody feature is more effective.

5. Conclusion

We propose a new algorithm for music classification based on music extraction and a deep neural network in this article. As the classification parameters for music, the algorithm first extracts two types of characteristics, the timbre characteristic and the melody feature. Because the classification method based on a convolutional neural network ignores the audio's timing. Thus, we proposed a music classification module based on a one-dimensional convolution of recurring neural networks by combining the proposed convolution structure with a single-dimensional convolution and a two-way recurrent neural network. To better represent the characteristics of music style, different weights of attention are applied to the output. On the GTZAN data set, we also ran comparison and removal experiments. The experimental results show that the proposed method outperforms several other well-known methods, with an accuracy of 91.99 percent.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author does not have any possible conflicts of interest.

References

[1] M. Reybrouck and T. Eerola, "Music and its inductive power: a psychobiological and evolutionary approach to musical emotions," *Frontiers in Psychology*, vol. 8, 2017.

[2] B. Burger, S. Saarikallio, G. Luck, M. R. Thompson, and P. Toiviainen, "Relationships between perceived emotions in music and music-induced movement," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 5, pp. 517–533, 2012.

[3] F. Nagel, R. Kopiez, O. Grewe, and E. Altenmüller, "EMuJoy: software for continuous measurement of perceived emotions in music," *Behavior Research Methods*, vol. 39, no. 2, pp. 283–290, 2007.

[4] P. Klimek, R. Kreuzbauer, and S. Thurner, "Fashion and art cycles are driven by counter-dominance signals of elite competition: quantitative evidence from music styles," *Journal of the Royal Society Interface*, vol. 16, no. 151, p. 20180731, 2019.

[5] P. Guimaraes, J. Froes, D. Costa, and L. A. de Freitas, "A comparison of identification methods of Brazilian music styles by lyrics," in *Proceedings of the Fourth Widening Natural Language Processing Workshop*, pp. 61–63, Seattle, USA, 2020, July.

[6] B. Schneider, "Community and language in transnational music styles: symbolic meanings of Spanish in salsa and reggaetón.," in *Contested Communities*, pp. 237–260, Brill, 2017.

[7] D. Nobile, "Double-tonic complexes in rock music," *Music Theory Spectrum*, vol. 42, no. 2, pp. 207–226, 2020.

[8] C. Weiß, M. Mauch, S. Dixon, and M. Müller, "Investigating style evolution of Western classical music: a computational approach," *Musicae Scientiae*, vol. 23, no. 4, pp. 486–507, 2019.

[9] A. El Saddik, "Digital twins: the convergence of multimedia technologies," *IEEE Multimedia*, vol. 25, no. 2, pp. 87–92, 2018.

[10] Z. Kotevski and I. Tasevska, "Evaluating the potentials of educational systems to advance implementing multimedia technologies," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 9, no. 1, pp. 26–35, 2017.

[11] Z. L. Kozina, I. N. Sobko, D. V. Safronov, D. O. Goptarev, and V. S. Palamarchuk, "Multimedia technologies as a means of training athletes in student basketball," *Health, Sport, Rehabilitation*, vol. 4, no. 4, pp. 50–61, 2018.

[12] R. Fleischer, "If the song has no price, is it still a commodity?: rethinking the commodification of digital music," *Culture Unbound*, vol. 9, no. 2, pp. 146–162, 2017.

[13] K. Riemer and R. B. Johnston, "Disruption as worldview change: a Kuhnian analysis of the digital music revolution," *Journal of Information Technology*, vol. 34, no. 4, pp. 350–370, 2019.

[14] J. Zhang, "Music feature extraction and classification algorithm based on deep learning," *Scientific Programming*, vol. 2021, Article ID 1651560, 9 pages, 2021.

[15] J. Gan, "Music feature classification based on recurrent neural networks with channel attention mechanism," *Mobile Information Systems*, vol. 2021, Article ID 7629994, 10 pages, 2021.

[16] J. Lee, J. Park, K. L. Kim, and J. Nam, "Samplecnn: end-to-end deep convolutional neural networks using very small filters for music classification," *Applied Sciences*, vol. 8, no. 1, p. 150, 2018.

[17] J. Zhang, W. Wang, C. Lu, J. Wang, and A. K. Sangaiah, "Lightweight deep network for traffic sign classification," *Annals of Telecommunications*, vol. 75, no. 7-8, pp. 369–379, 2020.

[18] H. Bahuleyan, "Music genre classification using machine learning techniques," 2018, https://arxiv.org/abs/1804.01149.

[19] D. S. Lau and R. Ajoodha, "Music genre classification: a comparative study between deep-learning and traditional machine learning approaches," in *Sixth International Congress on*

Information and Communication Technology (6th ICICT), pp. 1–8, London, 2021.

[20] Y. Gu, A. Chen, X. Zhang, C. Fan, K. Li, and J. Shen, "Deep learning based cell classification in imaging flow cytometer," *ASP Transactions on Pattern Recognition and Intelligent Systems*, vol. 1, no. 2, pp. 18–27, 2021.

[21] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-k chip package design," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2021.

[22] W. Sun, P. Zhang, Z. Wang, and D. Li, "Prediction of cardiovascular diseases based on machine learning," *ASP Transactions on Internet of Things*, vol. 1, no. 1, pp. 30–35, 2021.

[23] Y. Jiang, X. Gu, D. Wu, W. Hang, J. Xue, and S. Qiu, "A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18, no. 1, pp. 40–52, 2020.

[24] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–302, 2002.

[25] C. Xu, N. C. Maddage, X. Shao, F. Cao, and Q. Tian, "Musical genre classification using support vector machines," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)*, vol. 5, pp. V–429, Hong Kong, China, 2003, April.

[26] T. Li, M. Ogihara, and Q. Li, "A comparative study on content-based music genre classification," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pp. 282–289, Toronto, Canada, 2003, July.

[27] I. Panagakis, E. Benetos, and C. Kotropoulos, "Music genre classification: a multilinear approach," in *International Symposium Music Information Retrieval*, pp. 583–588, Philadelphia, USA, 2008.

[28] M. A. A. Albadr, S. Tiun, M. Ayob, M. Mohammed, and F. T. AL-Dhief, "Mel-frequency cepstral coefficient features based on standard deviation and principal component analysis for language identification systems," *Cognitive Computation*, pp. 1–18, 2021.

[29] S. I. Kang and S. M. Lee, "Improvement of speech/music classification based on RNN in EVS codec for hearing aids," *Journal of Rehabilitation Welfare Engineering & Assistive Technology*, vol. 11, no. 2, pp. 143–146, 2017.