Application Model Construction of Emotional Expression and Propagation Path of Deep Learning in National Vocal Music

Zhangcheng Tang*

Music and Dance Academy, Hunan First Normal University, ChangSha, 410205, China

Abstract—Emotional expression is important in Chinese national vocal music art. The emotional expression in national vocal music is based on the art of national vocal music, with distinct characteristics and requirements. The ultimate goal is to spread the expression of various emotions in the national vocal music art. Promoting the spread of national vocal music singing art using modern media is an urgent requirement for the inheritance and development of national vocal music singing art. With the rapid development of science and technology, integrating deep learning and traditional music has become the general trend. It has been gradually applied to melody recognition, intelligent composition, virtual performance, and other aspects of traditional music and has achieved good results, but also hidden behind a series of ideas and technical and ethical issues. In this paper, the application of deep learning has been discussed and prospected. The recognition rate of emotional expression in national vocal music is 92 %. In terms of communication, combined with the deep learning algorithm, this paper analyzes the characteristics and requirements of emotional expression in the art of national vocal music singing and puts forward a new method of promoting the development of the art of national vocal music singing, hoping to attract more attention and enhance the social awareness of the application field, to promote the steady development of Chinese traditional music in the information age.

Keywords—Deep learning; national vocal music; innovation; emotion; dissemination

I. INTRODUCTION

Emotional expression has always been particularly valued in national vocal music. The emotional expression running through the art of national vocal music is the soul of the art of national vocal music. The art of national vocal music without emotional expression will lose its significance. The emotional expression ability of singers is the focus of national vocal music singing art. The emotional expression of Chinese national vocal music singing art has accumulated rich levels and diversity in the development of thousands of years, with its unique characteristics and requirements. As the pearl in the art treasure house, the inheritance of national vocal music singing art is related to the integrity of traditional art [1], [2]. While analyzing the emotional expression of national vocal music singing art, continuously improving its artistic and emotional expression ability and meeting the current social and cultural needs, the inheritance of national vocal music singing art has also become a social problem. In the face of the new era of openness and diversification, how to take effective communication paths to make the national vocal music singing art occupy a place in the field of communication and inherit it for a long time is a problem that must be thought and solved simultaneously in addition to paying attention to the emotional expression of the national vocal music singing art.

Music education is gradually developing towards intelligence and online in the Internet era. As an emerging discipline, deep learning is mainly used to research, develop, and extend the method theory of human intelligence. Nowadays, many things are related to deep learning, such as fingerprint recognition, intelligent search, face recognition, language translation, and automatic planning. Deep learning and hearing has a great relationship, and deep learning for music education has a unique advantage. The breakthrough of material civilization also accompanied the development of music art. Deep learning will change music education's teaching mode and theory, especially in teaching means and methods, to provide positive practical value for music education in the Internet era [3].

Combining deep learning algorithms and traditional national vocal music has become an important research direction in the future. The singing of traditional national vocal music will usher in great changes under the impetus of deep learning. It makes the emotional expression of traditional folk music singing easier to identify and push to the audience who like to change the type of folk music so that it is not limited to the nation. The technology of retrieving music based on music's emotional attributes has also been developed. Many kinds of music are stored in the network music database. The essence of music information retrieval is music recognition and classification. Most music end users often like a certain type of music, and the diversity of music gives it unique attributes [4], [5]. Therefore, a music recognition and classification system can help people retrieve and manage music more effectively. MIR contains many sub-tasks, such as music emotion recognition, instrument recognition, genre recognition, and author recognition [6].

The traditional extraction of music emotion features often lacks the temporal structure and related semantics of music because it is usually extracted in frames, and the deep semantics of chords, melody, and rhythm in music that change with time is important for music emotion recognition. To analyze the correlation between temporal information and emotional expression in the research process, some scholars have done verification experiments in 2014: First, the music is

transformed into a feature vector of the time structure, and then the Gaussian mixture model, Markov and Hidden Markov model and other generative models are used to form the temporal characteristics of the music. Finally, the model is passed through the Probability Product Kernel for music emotion recognition. Traditional folk music has been rapidly upgraded, making full use of the convenience of the mobile Internet and spreading through more channels. Therefore, the mobile Internet has unprecedented advantages for promoting music works and the comprehensive utilization of traditional Internet, mobile communication networks, mobile phone ringtones, and other wired and wireless platforms for communication [7]–[10]. This mode of creation, production, dissemination, and acceptance of music products mainly relies on digitalization, reflecting technological and cultural progress and transforming and upgrading communication modes. The Internet and mobile terminal technology have achieved unprecedented development speed since the 21st century, especially in the past five years. It has not only formed a pattern of national coverage of online music but also gradually replaced the traditional music platform as the main channel in the rapid development of this mode of communication and platform. The mode of communication has developed into one of the most fashionable ways of music communication. Some scholars have applied deep learning to music content analysis, especially style, and artist recognition, producing meaningful features from the preprocessed spectrum. The deep learning features have higher accuracy than those standard Mel Frequency Cepstral Coefficients features for music style recognition. There is also a convolutional neural network (CNN) that MFCC features vectors and inputs them into three hidden layers. Finally, it can automatically extract image features for classification, indicating that CNN can capture changing image information features [11]–[13].

The role of national vocal music art in the mass media has also changed; specific singing techniques and emotional expression also began to change. Although from the perspective of modern communication, this is the development and progress of national vocal music art, to some extent, it can also be said to be desalination. In the modern media environment, maintaining a proper relationship between the dissemination of national vocal music singing art and the ' native ' cultural roots and maintaining unique characteristics in emotional expression, artistic technology, and other aspects is a problem facing the inheritance of national vocal music singing art. National vocal music technology should effectively use modern media to explore new communication paths. Based on this article, the convolutional neural network in deep learning is combined to realize the recognition and propagation path optimization of emotional expression in national vocal music. Firstly, the emotional feature extraction of vocal music signals is carried out, and then the time and frequency domain features are analyzed. The different emotions of vocal music signals in national vocal music are labeled and classified through feature extraction. The signal feature is the input index, and the emotional label is the output feature. Finally, the effective identification of emotions in national vocal music singing based on a deep learning algorithm is obtained, which provides efficient national vocal music classification based on a machine learning algorithm for following different categories of listeners and realizes targeted dissemination for different users.

II. RESEARCH ON DEEP LEARNING IN NATIONAL VOCAL MUSIC

A. National Vocal Music Multi-Vocal Emotion Parameter Recognition Design

In the expression of national vocal music, it is necessary to present different national vocal music styles and emotions by concretizing the abstract space of national vocal music. Effectively determining the similarity of two music points is an important part of concreting the abstract space of national vocal music. Music emotion recognition has become a popular trend. Under this condition, it is necessary to study how to enhance the accuracy of music emotion recognition and provide support and help for music search [14].

Currently, the mainstream music emotion recognition classification structure is shown in Fig. 1. There are three main steps:

1) Select the emotion model (continuous emotion model and discrete emotion model).

2) Preprocessing music, extracting useful music features and information as input.

3) Input into the recognition model for emotion recognition.

The most important part is extracting music emotion features and constructing the recognition classification model. Previous studies often used single emotional features or classifiers based on traditional machine learning models. A single emotional feature often does not have unity and cannot fully express musical emotions. It needs to be re-extracted when performing different recognition tasks. Although the recognition effect has been improved, the efficiency is too low. The traditional machine learning model only in the unique music feature recognition results; strange music is not ideal, and poor generalization ability shows the main breakthrough in the second and third steps.
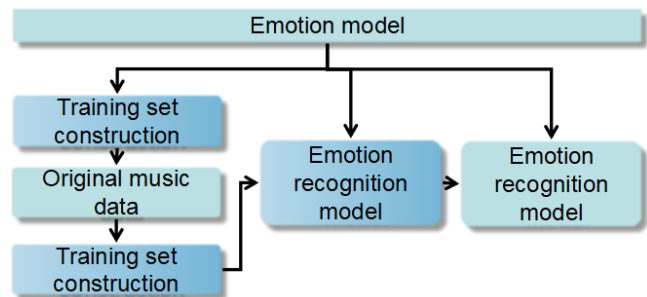

Fig. 1. Deep learning music emotion expression recognition basic structure.

Previously, music expression produced a series of music with the same style, whether relying on artificial technology or automation technology. In the abstract space of music, it isn't easy to have a variety of different styles of national vocal music at the same time. The reason for this phenomenon is that there is a certain probability of large similarity between national vocal music points with similar or different styles in the abstract space, which leads to an uneven transition between

different styles of national vocal music. Therefore, in the expression of national vocal music, the determination of the style of national vocal music creation needs to adopt some attributes as the characteristics of learning, and the style of national vocal music is numerically expressed. The objective function expresses the style value, and the national vocal music style is expressed based on the value of different objective functions.
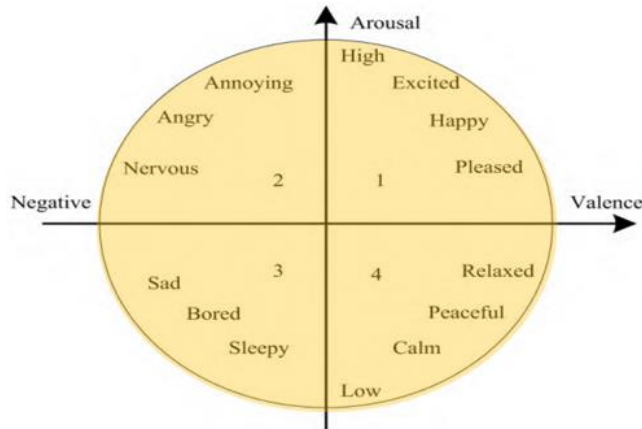

Fig. 2. National vocal music emotional space.

Different emotions are distributed in the two-dimensional space shown in Fig. 2, which is the emotional space. Based on the two-dimensional emotional space, in the expression of national vocal music, it is only necessary to project the emotional characteristics of music into the two-dimensional emotional space, take the coordinate scale value in the emotional space as the input information, and then use the artificial neural network to train the sample information. In this process, first, judge the various national vocal music elements in the training set, give the emotional category to determine its position in the emotional space, one-to-one correspondence with the emotional characteristics in the figure, obtain the emotional scope of the national vocal music, and determine the emotion and style of national vocal music creation.

The data of national vocal music is regarded as a time series, two different national vocal music sequences Y1 and Y2, their characteristic quantities are (a1μ, b1μ) and (a2μ, b2μ), and μ=1, 2,..., n represents the time length of national vocal music Y1 and Y2. At each time point μ, the Euclidean distance D (μ) between (a1μ, b1μ) and (a2μ, b2μ) is calculated, and then the similarity between the national vocal sequence Y1 and Y2 can be defined as μ = 1, 2,..., n.

$$S_{Y_1 Y_2} = \frac{1}{n} \sum_{\mu=1}^{n} D(\mu) \tag{1}$$

However, the above similarity calculation process only applies to the same time length of two national vocal music sequences. For two different time lengths of national vocal music sequence, let the two-time series are X = (x1, x2,..., xi) and Y = (y1, y2,..., Yi), and set the metric function k (x1, y1), can measure the distance between component x1 and y1, namely (a1μ, b1μ) and (a2μ, b2μ) Euclidean distance.

B. A New Model of National Vocal Music Communication Under Deep Learning

1) The traditional 'live' mode of transmission of national vocal music mainly is the early context of national vocal music; people can sing works, watch, listen to activities in indoor and outdoor places, and participate in live communication activities. These modes of transmission, especially the performance of the 'stage' as a platform to sing, appreciate the more common way. At this time, the music performers and the audience of the works are in the same space-time environment for close emotional and artistic information exchange, which is a materialized, information-sharing, face-to-face communication mode. This traditional communication mode, which has lasted for many years, emphasizes that both the dissemination of song information and the acceptance of songs in the activities show the characteristics of face-to-face and the same time and space. This traditional mode of creative song transmission does not even need the assistance of other media. The instant communication and feedback of national vocal music and emotional information can be realized between the information disseminator and the song receiver, which has a strong feature of real-time information sharing.

However, there have been new changes in the current model - with live video support; people can achieve instant information exchange and communication in different times and spaces. Disseminators of songs can achieve 'off-site instant' information exchange through mobile Internet through performances anywhere. In the high development of modern mobile media, the "on-site" communication of this kind of creative song has developed into a new "on-site" communication mode relying on live video broadcast, which has become one of the most important forms of contemporary creative song communication.

2) New deep learning model in disseminating national vocal music. Deep learning in various national vocal music communication modules has been innovative from the basic information dissemination acquisition-generation distribution process. From the perspective of music information collection, a deep neural network (DNN) has many applications in music information retrieval (MIR), including beat estimation, melody mining, music emotion recognition, etc. If music is regarded as a binary data stream, Convolutional Neural Network (CNN) can efficiently mine the feature attributes of musical moments. A recurrent neural network (RNN) can analyze the characteristics of music at different times, mining its more macro features, such as emotion. QQ music, Netease cloud music, and other platforms ' listening song recognition function is the basic application of MIR. Deep learning has made more innovative attempts from the generation of music information. Nowadays, deep learning technology uses generative adversarial networks (GAN) and reinforcement learning (RL) to inject data into music and dynamically evaluate it in real-time. Deep learning can complete the evaluation of music emotion only by relying on the binary

representation of music; it can also judge the good and bad after creating and generating music and modifying it. Deep learning technology to generate music is a process of repeated creation and reflection. From the distribution of music information, the network distribution platform of music has become the world of deep learning. Music platforms rely mainly on recommendation algorithms to accurately distribute music resources to corresponding users. Deep learning represented by recommendation algorithms can process a large amount of data faster and more accurately and make more accurate recommendations for users to get a better experience.

C. Opportunities and Challenges brought by Deep Learning to National Vocal Music

1) Deep learning enables people who do not understand music theory to create music and have ownership. Then, the problem arises. If the creator creates music without knowing how it is created, it goes beyond the scope of ordinary 'music creators'. At the same time, if deep learning music creation tools create new music based on imitating existing music, then defining similarity becomes a problem. A similar piece of music, the ownership of its creative subject, is the owner of the tool, the creator of the original music, or belongs to the tool? These are issues that need careful consideration.

2) Deep learning brings great convenience to music creation, but many creations can create, far from 'can create well'. Like other arts, music creation needs to think, feel, and understand. Only in this way can it be truly appreciated and endowed with value. The music constructed by deep learning technology only proves that deep learning can generate music. This music can only be used as experimental products, which cannot make waves and even reduce the quality of the whole music.

3) Music copyright is a behavior of endowing value for art, a very important issue on the road of music commercialization. Attaching importance to copyright is to protect creators and make them have a healthy creative environment. Deep learning allows online music resources to be easily collected and quickly spread on a large scale, which increases the difficulty of tracing the path of music transmission. With such speed and breadth of transmission, the traceability of music copyright and other audio and video resources, software copyright, etc., can only be out of reach. In addition, the difficulty of positioning the creative subject will lead to copyright issues. If a piece of music cannot clarify its subject attribution, it cannot discuss its proper attribution.

4) Deep learning technology originates from the deep neural network constructed by computer scientists to imitate the ' neurons ' of the human brain, which promotes the rapid development of music communication. Music's collection, generation, and differentiation are more intelligent with the participation of deep learning. Our in-depth understanding of deep learning technology will make related concepts' definitions more mature. After jumping out of the 'learning' mode, deep learning may truly perform music creation with machine rationality.

From the current technical level, artificial intelligence has achieved computational and perceptual intelligence, but cognitive intelligence is still insufficient and needs to continue developing into cognitive intelligence. To achieve breakthroughs in national vocal music, it is necessary to combine new things such as artificial intelligence, deep learning, and machine learning to achieve more efficient emotional expression and classify national music with different emotions to share with audiences of different preferences.

III. ANALYSIS AND IMPLEMENTATION OF NATIONAL MUSIC DEEP LEARNING SYSTEM

As the main component of music, audio plays a vital role in music emotion recognition. As an important part of music information retrieval, music emotion classification based on audio has attracted more and more attention. The second-level emotional expression in national vocal music mainly refers to the singer's realization of the emotional state through deeper excavation, such as through sound, expression, and movement form [15]. The article builds the model shown in Fig. 3.

A. Network Structure

The first few layers of the CNN network structure are feature extractors that automatically get image features through supervised training; the last layer is classified and identified by the SoftMax function. The CNN network structure is shown in Fig. 4, which shares the basic framework with the traditional AlexNet. It contains eight layers; the first five layers are convolutional layers alternating with the pooling layer, and the remaining three layers are fully connected layers for classification. The input images of the CNN network are harmonic spectra and shock spectra separated by HPSS and the spectra of the original music signals. The input image size is normalized to 256 * 256, and then it is input into the first convolution filter. In the deep network structure, the first convolution layer uses 96 kernels with a size of 11*11 and a step size of 4 pixels (the distance between the receptive field centers of adjacent neurons in the same kernel mapping) to filter the input image. Next, the max pooling layer takes the output of the first convolution layer as input and filters with 96 kernels with a size of 3*3, and the response is normalized. Using these five convolutional layers, 256 feature maps of size 6*6 were finally obtained, fed to three fully connected layers containing 4096, 1000, and 10 neurons, respectively. The final identification result is the output of the last fully connected layer, as shown in Fig. 4.
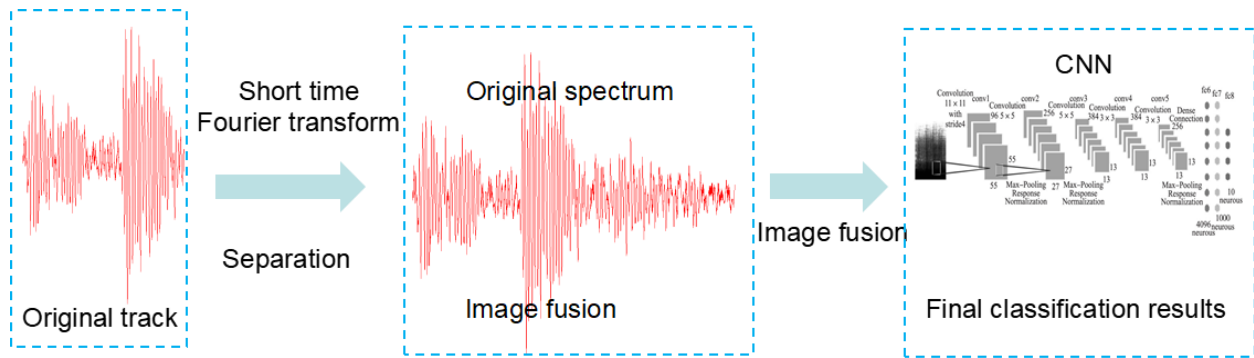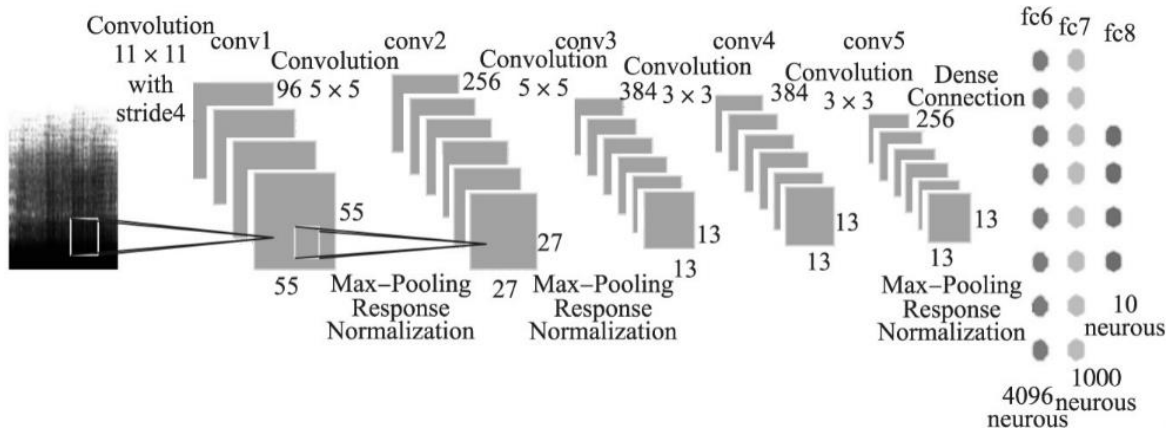
Fig. 3. CNN classification research model.

Fig. 4. CNN model.

B. Methods of Network Training and Learning

The Pooling layer uses Max Pooling, and the convolutional and pooling layers appear alternately in the convolutional neural network. Using the stochastic gradient descent (SGD) training network model, it is found that smaller weight attenuation is very important for model learning. Weight attenuation can decrease the training error in the model. Therefore, in the experiment in this paper, fine-tuning to 0.0005 usually, dropout and momentum can improve the learning effect. Since the convergence for all layers using dropout is time-consuming, the dropout value is set to η=0.5, μ= 0.9, λ = 0.0005 in the fully connected layer in the experiment of this paper. There are three fully connected layers in the network structure. The last fully connected layer, i.e., the eighth layer, is the output layer, and the output of the seventh layer is its input, which contains m neurons corresponding to m types of music styles. The output probability is p = [p1, p2,..., pm] T. The SoftMax regression formula is as follows.

$$p_j = \frac{\exp\left(X_8^j\right)}{\sum\limits_{i=1}^{m}\exp\left(X_8^i\right)} \tag{2}$$

where, Xj 8 is the input of the SoftMax function, j is the current class being computed, j = 1,..., m ; pj represents the true output of class j.

C. Model Results and Analysis

In the experiment, the Caffe framework is utilized to train the CNN model to realize the recognition of music style. Using the recognition rate as the performance index, the music signal on the Chinese folk music collection database is the input index, and the 8 types of emotions are the output indexes. Each style category contains multiple audio recordings. The folk music emotion classification model is in Fig. 5.

Fig. 5. Folk music emotion classification model.

TABLE I.   TRAINING-RELATED HYPERPARAMETERS

Hyperparameters	η	Size	μ	λ	Dropout
Value	0.02	15	0.9	0.0004	0.45

The parameter is to be adjusted when the error rate of the training set becomes small enough and stable. The hyperparameters obtained through this adjustment process are summarized in Table I.

Relation image between iteration times and hyperparameters is shown in Fig. 6 to Fig. 9.
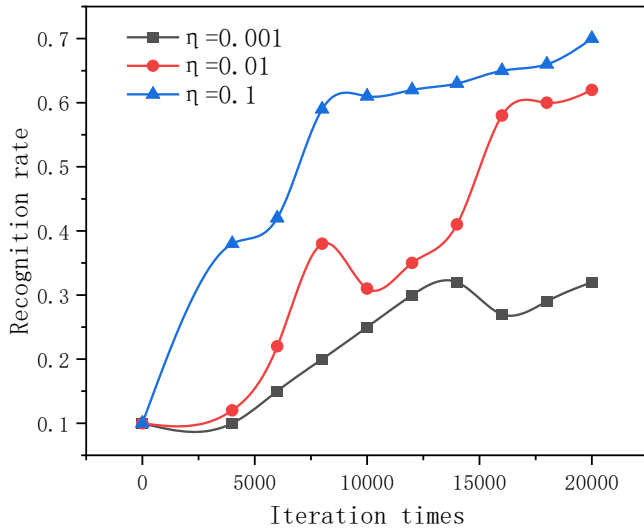

Fig. 6.   Learning rate and number of iterations.

Fig. 6 shows that the learning process will be very slow, and the recognition rate is unstable when the training samples are 20000 iterations, and the learning rate η is relatively small as 0.001. Appropriately increasing η can effectively improve learning efficiency. At the same time, if η is too large as 0.1, the learning process will be unstable, and the classification performance will be decreased. Fig. 8 and Fig. 6 illustrate the influence of momentum μ and weight attenuation λ, respectively. Fig. 7 indicates that using momentum μ can accelerate the learning process well.

As shown in Fig. 9, Dropout is a technique for preventing overfitting during the training of neural networks. In the literature, dropout with a 70 % reduction in output is applied to the last fully connected layer. This experiment adopts this technique and scrambles the training data in each round to reduce overfitting. In such research, the output of hidden layer neurons is usually set to 0 with a certain probability, so such neurons will not play any role in forward and backpropagation.

In this experiment, using the hyperparameters set in Table I, the recognition rate is about 73 % without data expansion. The index to realize the correct classification of national vocal music emotion is on the diagonal of the matrix in Table II.
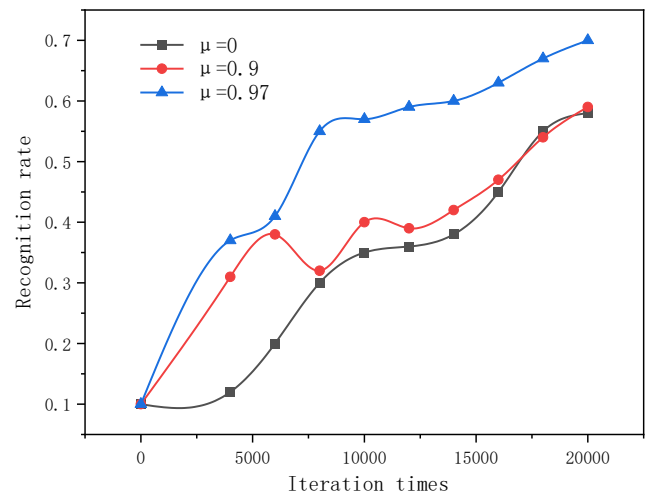

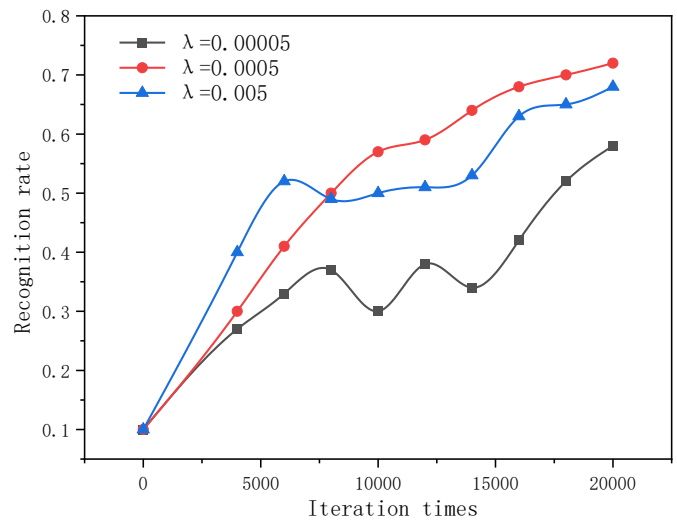Fig. 7.   Momentum coefficient and number of iterations.

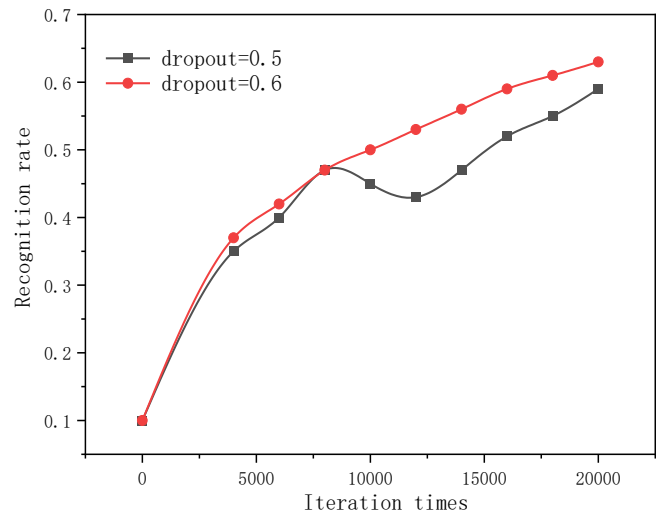Fig. 8.   Attenuation coefficient and iteration times.

Fig. 9.   Dropout value and the number of iterations.

TABLE II.   CONFUSION MATRIX

Emotion categorization	1	2	3	4	5	6	7	8
1	82.5	8.3	0.0	0.0	0.0	7.3	0.0	0.0
2	10.2	74.6	0.0	0.0	0.0	12.5	0.0	0.0
3	0.0	4.5	91.7	0.0	0.0	0.0	0.0	0.0
4	4.0	4.1	4.3	79.1	0.0	0.0	0.0	8.3
5	0.0	0.0	0.0	0.0	75.6	0.0	0.0	0.0
6	3.3	8.2	0.0	0.0	0.0	76.2	0.0	0.0
7	0.0	0.0	0.0	0.0	20.2	0.0	92.0	0.0
8	0.0	0.0	0.0	12.0	0.0	4.0	0.0	82.7

D. Research on National Vocal Music Dissemination Mode of Different Emotion Types

After getting different emotional types of folk music, the emotional types of music can be classified and put on the corresponding label in the current mainstream music software. Take the dissemination of national music in microblogs as an example. After researching and analyzing the elements in the communication mechanism, it is not difficult to find that the music communication mode in microblogs is similar to the epidemic communication theory on complex networks. On this basis, by transforming the classic infectious disease communication model, a model consistent with the music communication information in microblogs is constructed to describe the communication process of music information in microblogs and a series of behavioral characteristics to facilitate the study of music communication mode in microblog in the future. Disseminating music information in Weibo, when the communicator sends out the music information, it can be seen by its friends and fans on Weibo. When the recipient receives the information, it can have one or more behaviors, such as forwarding, commenting, and liking, paying attention to private letters, and collecting the choice of music information feedback. Music knowledge, hobbies, friend intimacy, mood, and other conditions influence this choice. Similarly, the complexity of the information itself also determines its communication effect. Whether different music information has a certain value is the basis of its broadcast. The source of music information also restricts the spread of this music. The music information sent by the disseminator with high popularity and great influence is more likely to be disseminated among the audience. This music information can be more identified as meaning or value. Affected by this, the disseminator has a certain probability to choose to forward the content of the entry containing the music information. It can be regarded as an 'infected person', and the music information transmitted can be compared to the 'virus' of an infectious disease. The audience is not interested in the music information contained in this microblog and is regarded as an 'immune person'. They may hinder the praise of friends, spit in the comments or even directly ignore this information. That is the termination of music transmission. Because the music communication network in microblogs is complex and diverse, the scope of the music audience is wide, and the behavior is random and uncertain. To facilitate the research, the music communication is set in microblogs that can only spread from the music communicator to the music audience as its fans. The communicator is set as a node; lines connect the relationship among them, and the music information can only be transmitted through lines. Combined with the influence of the factors in the music communication mechanism, the classical SIR model is improved, and the music communication mode in the mechanism is modeled and studied in Fig. 10.

When the public in the micro-blog does not receive the music information, its node is S after it is 'infected' by the music information, it will be converted to the I state or IR state and the probability of converting into two states is uncertain. If it is transformed into state I, it continues to propagate along the line through the node it is concerned with and eventually transforms into state R; if it is transformed into IR state, it may need to be transformed into R state, or I state through the influence of some other information. In Weibo, music information often has certain timeliness. After some time, the heat of information will gradually decrease, so the characteristics of relay propagation will decrease. At a certain point, the IR state will also change to the R state until only two nodes, S and R, are on Weibo. In studying the music transmission mode in the music transmission mechanism in microblogs, the SIR model of music transmission is constructed by improving the SIR model of classical infectious diseases. It fully considers the rules of music transmission mode in microblogs as much as possible and discusses the influence of nodes and transmission lines. Based on the SIR model of classical infectious diseases, it can combine the dynamic principle of infectious diseases and complex networks to further study the music transmission mode in the microblog. It provides a reference basis and maybe another new direction for music transmission research.
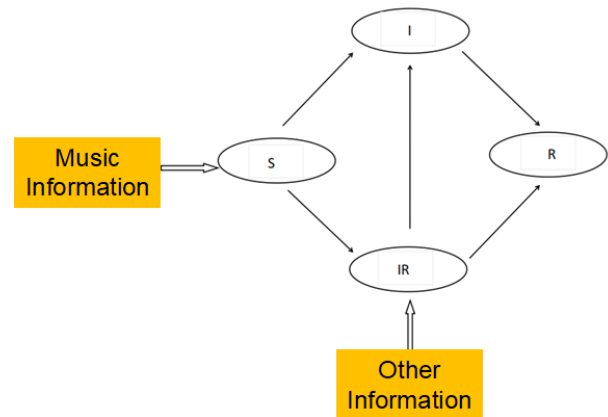

Fig. 10. National vocal music propagation model.

IV. Conclusion

Deep learning accelerates the scope and speed of information dissemination and expands the impact of various disciplines on music dissemination. Based on the innovation of deep learning in music communication, the present work analyzes the challenges brought by deep learning to music communication and looks forward to the future development of the rational application of deep learning in music communication activities. The efficient identification of emotional expression in national vocal music singing through the CNN model has been realized, and the identification accuracy is as high as 92%. With the help of microblog software in China, a new national vocal music promotion model has been designed. The article research is a new attempt at national vocal music research and has certain research value.

Music emotion recognition is an important research direction in music information retrieval. The research of music emotion involves many interdisciplinary such as music and psychology. In this paper, before the model construction, some Chinese folk music was mobile phoned, and a database was established, including eight different types of folk music audio with different emotional characteristics. Firstly, the music features are filtered and extracted, then the deep learning network model is used to process the music, and the features are used as the input of the CNN model to realize the emotional classification of Mongolian music. The final study shows that the accuracy of emotional identification of national vocal music is as high as 92 %, and the new national vocal music communication mode can achieve more efficient national music sharing. Although many researchers have conducted some research on some of these sub-areas and have achieved initial results, music emotion recognition is still in its infancy; there is still a lot of research space. The music emotion recognition work based on Mongolian music faces many problems due to limited conditions and late start.

Availability of Data and Materials

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of Interest

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

Authorship Contribution Statement

Zhangcheng Tang: Writing-Original draft preparation Conceptualization, Supervision, Project administration.

References

[1] W. Hongdan, S. SalmiJamali, C. Zhengping, S. Qiaojuan, and R. Le, "An intelligent music genre analysis using feature extraction and classification using deep learning techniques," Computers and Electrical Engineering, vol. 100, p. 107978, 2022.

[2] D. Xiaofeng, "RETRACTED: Application of deep learning and artificial intelligence algorithm in multimedia music teaching," Journal of Intelligent & Fuzzy Systems, vol. 38, no. 6, pp. 7241–7251, 2020.

[3] C. Kereliuk, B. L. Sturm, and J. Larsen, "Deep learning and music adversaries," IEEE Trans Multimedia, vol. 17, no. 11, pp. 2059–2071, 2015.

[4] C. Kereliuk, B. L. Sturm, and J. Larsen, "Deep learning, audio adversaries, and music content analysis," in 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE, 2015, pp. 1–5.

[5] J. Nam, K. Choi, J. Lee, S.-Y. Chou, and Y.-H. Yang, "Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach," IEEE Signal Process Mag, vol. 36, no. 1, pp. 41–51, 2018.

[6] S. Zhao et al., "Emotion-based end-to-end matching between image and music in valence-arousal space," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2945–2954.

[7] H. Hong and W. Luo, "The method of emotional education in music teaching," The International Journal of Electrical Engineering & Education, p. 0020720920983559, 2021.

[8] N. Thammasan, K. Moriyama, K. Fukui, and M. Numao, "Continuous music-emotion recognition based on electroencephalogram," IEICE Trans Inf Syst, vol. 99, no. 4, pp. 1234–1241, 2016.

[9] M. B. Er and I. B. Aydilek, "Music emotion recognition by using chroma spectrogram and deep visual features," International Journal of Computational Intelligence Systems, vol. 12, no. 2, pp. 1622–1634, 2019.

[10] Y. Hu, "Music Emotion Research Based on Reinforcement Learning and Multimodal Information," Journal of Mathematics, vol. 2022, pp. 1–9, 2022.

[11] N. V. Kimmatkar and B. V. Babu, "Novel approach for emotion detection and stabilizing mental state by using machine learning techniques," Computers, vol. 10, no. 3, p. 37, 2021.

[12] S. Nag, M. Basu, S. Sanyal, A. Banerjee, and D. Ghosh, "On the application of deep learning and multifractal techniques to classify emotions and instruments using Indian Classical Music," Physica A: Statistical Mechanics and its Applications, vol. 597, p. 127261, 2022.

[13] S. Chowdhuri, "PhonoNet: Multi-stage deep learning for raga preservation in hindustani classical music," J Acoust Soc Am, vol. 146, no. 4, p. 2947, 2019.

[14] R. A. Wittmann-Price and M. Godshall, "Strategies to promote deep learning in clinical nursing courses," Nurse Educ, vol. 34, no. 5, pp. 214–216, 2009.

[15] R. T. Dean and J. Forth, "Towards a deep improviser: a prototype deep learning post-tonal free music generator," Neural Comput Appl, vol. 32, pp. 969–979, 2020.