International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-22

https://doi.org/10.70917/ijcisim-2025-0311
Article

Artificial Intelligence-Driven Innovative Research and Educational Practice of Chinese Minority Music Inheritance Model

Ying Wang *

The Conservatory of Music, Xinjiang Normal University, Urumqi, Xinjiang, 830054, China;
wy13579813120@163.com

Abstract: This paper deeply researches the integration of music education and artificial intelligence, focusing on the innovation of Chinese minority music inheritance model and educational practice. Firstly, it focuses on the application potential of artificial intelligence in the field of music education, and proposes an innovative model of Chinese minority music inheritance. Second, the music classification model MGTN and Markov chain-based intelligent composition model are proposed to realize the educational practice of music inheritance model through the intelligent recognition and automatic generation of ethnic music genres. The experimental results show that the correct rate of folk song classification of MGTN model is 85.11%, which is better than other comparative models. Meanwhile, the intelligent composition model proposed in this paper improves the overall subjective comprehensive evaluation index by 14.91%~34.65%, and improves the overall objective comprehensive evaluation index by 0.89~1.47 times, and verifies that the model generates the result as a melody with global structure through the melodic line drawn by using lifted sampling coding, which provides example reference for the inheritance and innovation of the music of Chinese ethnic minorities.

Keywords: music classification; intelligent composition; MGTN model; Markov model; ethnic music inheritance and innovation

1. Introduction

At present, China's inheritance and promotion of excellent traditional culture has been elevated to a new and unprecedented level, and Chinese minority music, which has been precipitated for thousands of years, is a unique interpretation of Chinese cultural thought and national spirit [1]. Under the surge of this social tide of inheriting and reviving the excellent traditional Chinese culture and exchanging the world's colorful civilization, it has inspired the relevant art practitioners to explore the discovery of Chinese minority music materials and innovative forms. Artificial intelligence, a globally popular technology, is injecting new vitality into the creation, operation and dissemination of Chinese traditional music, especially minority music [2-3]. Such a form can not only meet the modern audience's appetite, but also obtain more commercial added value in the market, and at the same time promote the dissemination and development of Chinese minority music in the footsteps of the times [4]. Similarly, the cross-border integration of artificial intelligence and Chinese minority music is not water without a source and wood without a root. In-depth combing of artificial intelligence and Chinese minority music, thinking about the characteristics of the fusion of the two, the fundamental concept of collaborative expression of the two, and exploring the significance and value of the collision of the two, is the future of Chinese minority music cross-border innovation and inheritance of the basis, but also the key to the dissemination of Chinese minority music [5-7].

With digital technology gradually shifting from “technology empowerment” to “technology empowerment”, the inheritance and innovation of traditional ethnic culture is gradually transformed from pan-entertainment to digitization and intelligence, and it is proposed that traditional ethnic culture should be promoted to the international level through cross-border cooperation, while the inheritance and

innovation of Chinese ethnic minority music should be enhanced. The inheritance and innovation of ethnic minority music should pay more attention to content quality and innovation, and digital technology should serve the diversity and innovation of culture [8-10]. Especially in cultural creativity, digital technology comprehensively empowers cultural product design and creative innovation through data-driven, artificial intelligence and blockchain. For example, literature [11] presents the theoretical basis of CHROMATA's immersive experience platform, which utilizes AI technology to analyze the movements, emotions, and sounds of traditional Greek dance culture in order to provide Internet users and cultural institutions with immersive content retrieval and a dance movement learning experience. Literature [12] proposes the application of digital technologies, such as motion capture and motion algorithms, in the development of virtual games for traditional folk dances, aiming to reduce the risk of loss of folk culture and to broaden the dissemination channels of traditional culture. With the rapid development of the times, people's awareness of the inheritance and development of Chinese minority music has gradually faded. For this reason, literature [13] proposes to use digital technology to present and inherit traditional music in the Internet era, activate and expand the new functions of traditional music, and reach a double balance between social and economic benefits. Literature [14] proposes to optimize the cultural inheritance model of folk music by integrating advanced technologies such as artificial intelligence and deep learning (DL), aiming to analyze the elements of folk music at multiple levels and comprehensively through the newly constructed model, which reveals its intrinsic emotional expression mechanism, thus providing new technical support for the protection and inheritance of traditional folk music.

Literature [15] suggests that digital recording, virtual reality and augmented reality in digital technology play a bridging role in combining traditional folk music with modern inheritance, and analyzes the situation of Chinese folk music to promote cultural inheritance and innovation through technological empowerment. Literature [16] analyzes the application of digital technology in folk instrumental music culture in the article, emphasizes the specific paths of intelligent question and answer, personalized recommendation and intelligent search in artificial intelligence to promote the dissemination of folk instrumental music culture, and highlights that with the help of digital empowerment, the instrumental music culture of the Tujia ethnic group can be integrated into modern creative products, so as to make the Tujia ethnic group instrumental music culture more unique and charming. Literature [17] suggests that motion capture technology opens up a new dimension for the inheritance and protection of ethnic minority music and dance, and that the development and reuse of ethnic dance culture through motion capture technology improves the accuracy and innovativeness of the inheritance, and culture and science and technology promote each other and develop together. Through the above systematic combing of the theoretical foundations of digital technology empowering the inheritance of Chinese folk music, we analyze the current situation and trend of the application of artificial intelligence technology in Chinese minority music, so as to provide theoretical support for the research and build a solid theoretical framework.

This paper explores the application potential of AI in intelligent music analysis, personalized teaching and intelligent music creation, and proposes an innovative model of AI-driven Chinese minority music inheritance. In terms of educational practice, this paper explores the intelligent recognition and automatic generation of music, using AI to assist students in music learning and creation. It mainly includes two aspects: one is the music genre classification model MGTN based on the convolutional attention mechanism, and the other is the folk song melody automatic generation model based on the improved Markov model, and the application effects of the two are explored experimentally respectively.

2. Music inheritance mode innovation driven by artificial intelligence

With the rapid development of science and technology, AI has penetrated into the field of music education, injecting new momentum into the traditional education model. This chapter will deeply explore the application potential of AI in music education, and explore the deep integration of music education innovation and AI from the unique perspective of Chinese minority music.

2.1. The Potential of AI in Music Education

The application of AI technology in music education is deepening, and the coverage area is gradually expanding, bringing far-reaching impact to the whole education system. In this section, we will delve into the application of the three aspects of intelligent music analysis, personalized teaching solutions and intelligent music creation.

2.1.1. Deep Dive into Smart Music Analytics

Initial smart music analysis focused on simple recognition of notes and rhythms. With the continuous

upgrading of technology, modern intelligent music analysis has been equipped with deeper functions. Expansion in areas such as emotion analysis and melodic structure analysis has enabled AI systems to more fully understand students' perception and expression of music, and this deep dive has brought more possibilities for special music education. The system can accurately capture students' emotional changes in the expression of musical works, so as to better adjust teaching strategies and meet different learning needs. At the same time, the in-depth analysis of melodic structure enables the system to better assist students in understanding and analyzing the artistic structure of the work, and improve their appreciation of the musical work.

2.1.2. Provision of Personalized Instructional Programs

AI technology is able to customize a learning plan that meets the needs of each student. Individual differences are very significant in the process of music teaching, and personalized teaching programs can design different kinds of interesting learning activities to stimulate students' interest and give them a profound learning experience, meeting the needs of different students. At the same time, accurate analysis of subject levels allows the system to provide appropriate learning content for each student, ensuring that teaching progress is neither too fast nor too slow, and improving the overall quality of music education.

2.1.3. The push for intelligent music creation

With the continuous development of AI technology, intelligent music composition has become an area of great interest. By training machine learning models, AI can generate relatively artistic musical works, providing students with more opportunities to participate in music creation. In traditional music creation, students may be limited by their own knowledge of music theory and creative experience, while the introduction of AI technology allows students to gain more creative inspiration and possibilities by interacting with the system. The system can analyze students' creative styles and provide creative suggestions, and even synthesize part of the music, prompting students to participate more deeply in the process of music creation, which provides a brand-new way to cultivate students' creative thinking and artistic expression.

Overall, the application of AI in the field of music education has great potential. By digging deeper into intelligent music analysis, personalized teaching solutions and intelligent music creation, we can better promote the innovation and development of China's ethnic minority music inheritance model, and provide students with a richer, personalized music learning experience.

2.2. Reflections on deep integration of traditional music culture

Chinese minority music culture plays an irreplaceable role in music education, and its long and profound historical heritage has become a precious resource for cultivating students' musical emotion and aesthetic ability. In order to better integrate it into the innovation of music education, we urgently need to have a deeper understanding of the uniqueness of Chinese minority music culture.

2.2.1. Uniqueness of China's Minority Music Culture

The music culture of China's ethnic minorities is not only a treasure trove of historical inheritance, but also a breeding ground for unique musical styles and aesthetic concepts. This uniqueness is not only apparent in the form of music, but is also integrated into many aspects of society and culture, which together construct a one-of-a-kind music and culture system. A deep understanding of these qualities is to better guide students to understand and experience the profound connotation of traditional music. Through the in-depth knowledge of music culture, we are able to explore the charm of music more comprehensively and provide richer teaching resources for music education. When exploring the uniqueness of Chinese ethnic minority music culture, we can draw on the socialization perspective to analyze in depth the unique elements generated from various aspects such as local folk culture, labor production, and religious concepts. Through in-depth study of the cultural genes of different traditional music genres, we can explore their development and evolution in the social context. This dimension of reflection will enrich our understanding of minority music culture and give music education a more profound cultural connotation.

2.2.2. Innovation of teaching mode integrating minority music culture

The integration of minority music in music teaching is crucial, not only to help cultural heritage, but also to stimulate students' interest in traditional music. The use of innovative teaching methods, such as interactive teaching and multimedia presentations, allows students to experience the essence of minority

music culture. Such innovations not only help to pass on minority music, but also can deepen students' experience of music, inject new impetus into music education, and make it better adapted to the needs of modern music education. In the in-depth discussion, we can excavate the local music from the perspective of socialization and analyze the cultural genes of the music of different ethnic minorities, including ethnic folk culture, labor production, religious concepts and so on. By exploring the history and social origins of traditional music of ethnic minorities, the meanings and functions of music in different ethnic groups and cultures, as well as the diversity and changes in music culture, we can systematically analyze its cultural connotations, aesthetic expressions, and contemporary values. Such a comprehensive analysis will provide music education with richer teaching resources and theoretical support, so that it can better adapt to the development of the A1 era.

3. Intelligent Recognition of Chinese Minority Music Based on MGTN

In this paper, we extract the features of minority music to realize the intelligent recognition of music, and then realize the automatic generation of music, so as to assist students to use AI technology to carry out the inheritance and innovation of minority music culture. In this chapter, the classification and recognition of minority music is studied, and a bimodal minority music genre classification model MGTN based on a convolutional attention mechanism is proposed.

3.1. Preprocessing of music signals

Before the music signal can be processed at a deeper level, the music information must be pre-filtered, sampled and quantized, pre-emphasized, and windowed and framed.

3.1.1. Pre-filtering

The main purpose of pre-filtering the humming signal of music:

(1) It is generally believed that components or noise in the signal higher than 1/2 sampling frequency will affect the signal, for example, it will cause overlapping of the spectrum, which will result in distortion of the high frequency components of the signal. The pre-filtering process will limit the bandwidth of the signal to a certain range, thus avoiding signal distortion.

(2) Generally, 50 HZ industrial frequency interference will cause the generation of power supply interference, and the pre-filtering process can suppress the 50 HZ power supply industrial frequency interference.

3.1.2. Sampling and quantization

Once the pre-filtering is complete, the music signal needs to be digitized, which is the sampling and quantization of the signal. The main purpose of sampling and pre-quantization is to convert continuous, analog music signals into discrete, digital music signals. Only then can computers process these music signals.

Sampling processing is the conversion of continuous points into points that are discrete in time according to the sampling theorem. The processing method generally used is to take a sample point of the signal at equal intervals. Generally this equal time is not less than half of the signal period, otherwise the signal will be distorted after sampling. The time interval between two consecutively taken samples is the sampling period, and the sampling rate is its reciprocal.

After sampling, the signal is discrete in time, but still continuous in amplitude. This requires quantization to discretize the amplitude of the music signal. The main task of quantization is to split the amplitude of the signal into a finite number of intervals, where the sample points within the same interval are represented by a single amplitude value.

3.1.3. Pre-emphasis

In the music signal, the high-frequency part of the proportion is relatively small, generally around 800HZ, 800HZ above the 6dB/octave rate of landing. But high music in the high pitch part of the signal is generally embodied by the high-frequency part of the signal, due to the high-frequency part of the signal is relatively sparse, so the need for the high-frequency part of the music signal processing, enhance the high-frequency part of the signal to become flat and easy to deal with, and this is the pre-emphasis needs to be accomplished task.

Pre-emphasis processing is generally realized by using a first-order digital filter. Equation (1) represents a first-order digital filter:

$$H(Z) = 1 - uz^{-1} \tag{1}$$

where u is close to 1, usually taken as $u = 0.94$.

3.1.4. Short-time window processing

After sampling and quantizing the music signal, it is actually a very unstable signal in time, in order to be able to process and analyze it in a better way, it is generally assumed that the signal is stable within 10-20ms, then this short-time smooth signal can be analyzed and processed.

After the music signal is processed by adding windows, it is divided into one frame after another, thus forming a stable short-time signal. In order to maintain the continuity of the music signal, the general use of overlapping segmentation method to split the frame, that is, the frame and the frame can overlap each other, the overlap part becomes the "frame shift", the frame shift is generally half of the window length. Rectangular window and Hamming window are two commonly used window functions. The window function of rectangular window is:

$$w(n) = \begin{cases} 1, 0 \le n \le N-1 \\ 0, \text{ Other} \end{cases} \tag{2}$$

The window function of a Hamming window is generally:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left[\dfrac{2\pi n}{N-1}\right], & 0 \le n \le N-1 \\ 0, & \text{Other} \end{cases} \tag{3}$$

where N is the length of the window. In the time-domain analysis of music signals, the advantage of rectangular window is that it has good smoothing, and the disadvantage is that it is easy to lose the details of the waveform, which leads to the generation of leakage phenomenon. The Hamming window can generally effectively prevent the occurrence of leakage phenomenon, so the Hamming window has a wide range of applications. But in different application occasions, it is necessary to choose the two reasonably.

3.2. MGTN model structure

3.2.1. Overall structure

The overall structure of MGTN is shown in Figure 1. In order to study the correlation between the spectrogram and the 2 modes of audio and the hidden correlation between different time steps or feature channels of audio, the MGTN model is improved on the basis of the traditional Transformer [18].

The model inputs are a spectrogram $\mathbf{Image}$ with a vector of audio feature time series $\mathbf{A} \in \mathbf{R}^{m \times n \times d}$, and in this paper we refer to the module that uses the Mayer spectrograms as inputs to the processing as the Image Coding Layer Encoder, which uses an image coding layer containing the Encoder_1 computes the convolutional attention score of the vector $\mathbf{Image} \in \mathbf{R}^{m \times n \times d}$. A positional coding layer is used in the Encoder_1 structure, which helps the network understand the relative or absolute positional information in each sequence. The module that processes time series data constructed using time/frequency domain features as input is called the time series (TS) encoder, which uses both the En-coder_1 and Encoder_2 structures to compute, respectively, the correlation between the vector $\mathbf{A} \in \mathbf{R}^{m \times n \times d}$ time step and the correlation between the feature channels, since the time series encoder deals with time series data, it has no image coding layer structure. The experiment uses Encoder_1 with a convolutional attention mechanism to model the time-step dimension of the music sequence, and Encoder-2 with a multi-head self-attention mechanism to model the feature dimension of the music time-series data, and then the structure of the 2 encoder outputs is connected by a splicing operation. Before the attention score is output, it passes through a random deactivation layer (Dropout) to mitigate overfitting. Subsequent use of the ReLU activation function and passing through a layer normalization layer allows the model to stabilize the training process with some translational invariance. The input to the linear output layer is the output of the previous encoder layer, which projects a high-dimensional representation of the music sequence information into a vector whose dimensions are the number of categories.

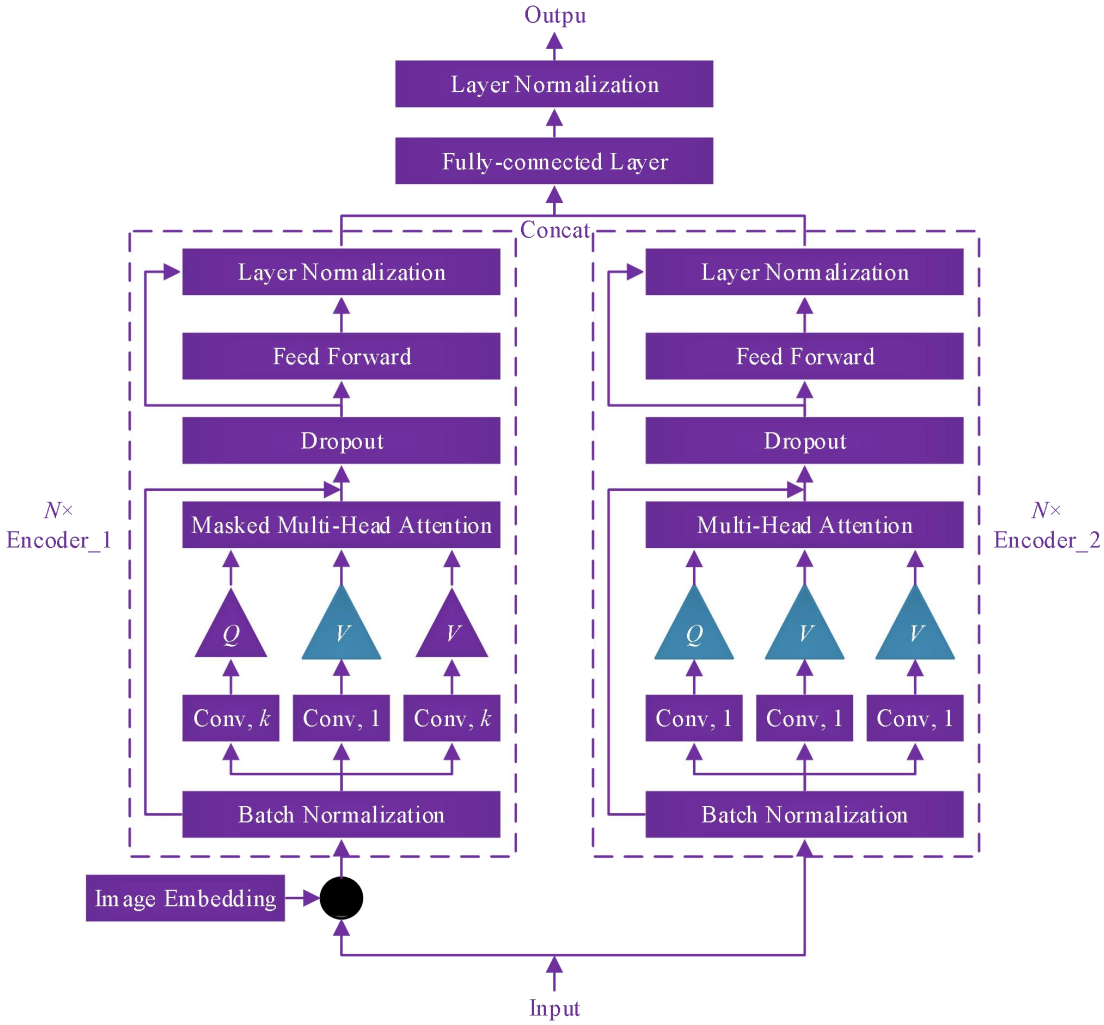


Figure 1. Structure of MGTN.

3.2.2. Image Embedding Layer

The process of converting the spectrogram into the Token categorization information required by the Transformer is shown in Figure 2. Experimentally, the spectrogram with an input size of 128×128 is divided into 64 non-overlapping blocks of a fixed size of 16×16, and each block is subsequently compressed into a one-dimensional vector $\mathbf{w}_1$, with each block containing the original image information as well as the position embedding containing the positional information.The length of the input sequence to the Transformer network α is the number of blocks, and the one-dimensional vector $\mathbf{w}_1$ represented by each block is equivalent to the word vector encoding length β. Due to the large dimensionality of the one-dimensional vector $\mathbf{w}_1$, the experiment compresses the dimensionality of the one-dimensional vectors $\mathbf{w}_1$ of each block after stretching through a linear layer, which also enables the feature transformation process.

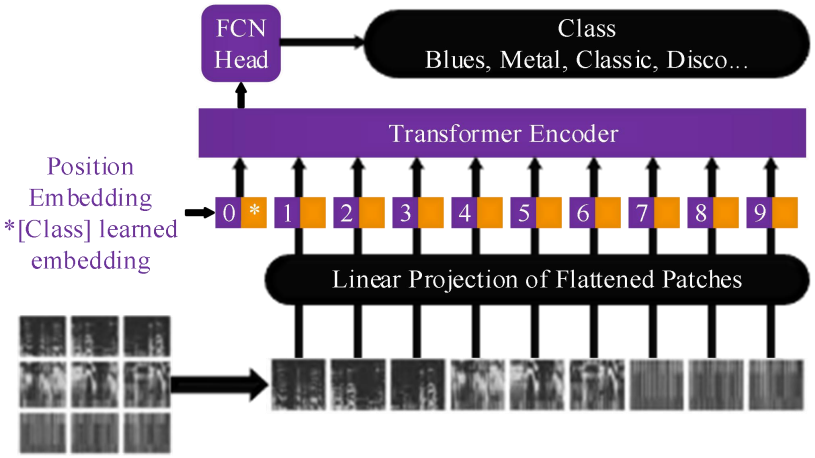


Figure 2. Sketch map of image embedding layer.

3.2.3. Convolutional Multiple Self-Attention Layers

The experiments in this paper apply the attention mechanism to spectrogram and audio feature data, and the computational principles of the self-attention mechanism and the convolutional attention mechanism are shown in Figures 3 and 4, respectively.

$\mathbf{InputMatrix} \in \mathbf{R}^{m\times n\times d}$ is used as input, where the time step length is m. The time feature dimension is n and contains n-dimensional audio features such as center distance, zero-crossing rate, etc. The d is the model hyperparameter d_{model}. The input matrix $\mathbf{InputMatrix} \in \mathbf{R}^{m\times n\times d}$ is first mapped to dense vectors by linear layer mapping. A convolution kernel of size $k\left(k \in \{1, a\}\right)$ is used in the experiments to perform convolution operations in computing the query and keyword matrices, which enhances the model's ability to model the local contextual information in the music time-series data. In the attention layer, a multi-head attention sublayer simultaneously transforms the input matrix $\mathbf{InputMatrix} \in \mathbf{R}^{m\times n\times d}$ into H distinct query matrices $\mathbf{Q}_h = Y\mathbf{W}_h^Q$, the keyword matrix $\mathbf{K}_h = Y\mathbf{W}_h^K$, and the value matrix $\mathbf{V}_h = Y\mathbf{W}_h^V$, where the number of attention polytopes $h = 1, \cdots, H$. Finally, the score matrix $\mathbf{ScoreMatrix} \in \mathbf{R}^{m\times n\times d}$ is then obtained by multiplying the query matrix with the value matrix as the output.

For spectrograms, each output element represents a link between every 2 convolutional blocks. For audio timing data, each output element represents the correlation score between every 2 time steps or every 2 temporal features. In the experiments of this paper, the attention layer of Encoder_1 is realized by the convolutional multi-head attention mechanism and uses an upper triangular mask matrix with element values of $-\infty$ so that the query matrix does not pay attention to the keyword matrix in order to prevent the ensuing leakage of the temporal sequence information. And the self-attention layer of Encoder_2 is the traditional Multi-Head Self-Attention mechanism Multi-Head Self-Attention.

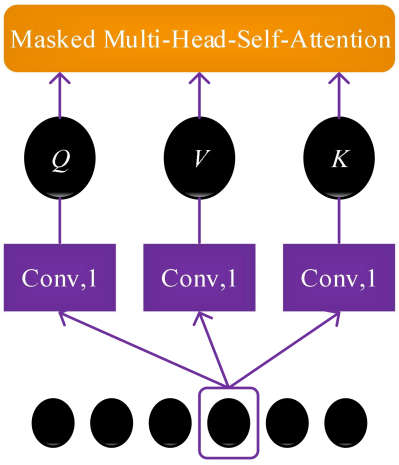


Figure 3. Calculation Principle of Convolutional Attention Mechanism.

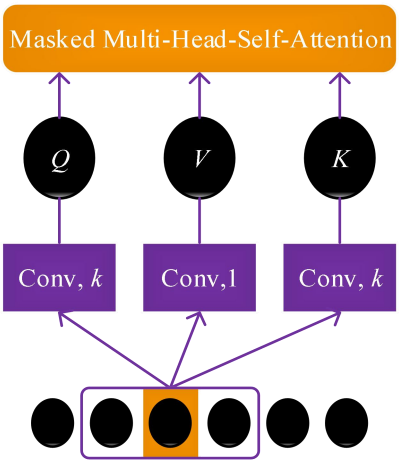


Figure 4. Calculation principle of convolutional attention mechanism.

3.3. Data pre-processing methods

This paper provides details of the data preprocessing steps of the experiments for the 2 modal data for the classification of Chinese minority music genres, and describes in detail the 2 data processing methods for solving this classification problem.

The experiments performed data augmentation on the dataset. In order to ensure that the amount of data was sufficient and that the musical feature information of the source audio was not lost, each music file with a duration of 30s was divided into multiple small segments of 3s in length, and the classification label of each small segment was kept the same as before division. Subsequently, a signal transformation is performed for each fragment: a Gaussian noise $f(x)$ with random amplitude ranging from 0.005 to 0.020 is added to the audio signal:

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{4}$$

The audio signal is then multiplied by a random amplitude factor in the range of -12dB to 12dB to keep the volume of the audio signal in a smooth range. Next, the audio signal is normalized for loudness, and a gain with a constant value of K is applied to achieve a suitable volume range. Finally, the pitch of the audio is transformed between 8 and -8 semitones without changing the tempo. The audio signal is stretched in rows of 0.5 to 1.5 on-the-fly without changing the pitch. Each transformed audio segment maintains its original classification label.

3.3.1. Generating a Mel Spectrogram

A spectrogram is a two-dimensional representation of a signal with the x -axis representing time and the y -axis representing frequency. A color map is used to quantify the magnitude of a given frequency within a given time window. The Mel spectrogram is a spectrogram of frequencies converted to the Mel scale. In this paper, each music file is converted to a Mel spectrogram, which is obtained by setting the window with a duration of 46.44ms and a size of (22050Hz, 1024points).The size of the Mel-bins is set to 128, so the input size of the Mel spectrogram for each small clip is 128×128.

3.3.2. Extracting audio features

Unlike Mayer spectrograms as input, extracting audio features requires the selection of an accurate feature set to ensure classification accuracy. Extracting features in audio feature engineering usually produces a n -dimensional vector, where the value of n depends on the length of the audio clip. If the value of n is very large, it is very difficult to deal with high-dimensional feature vectors, so it is necessary to downscale high-dimensional feature vectors. Therefore, for the input audio feature vector $\mathbf{V}=(v_1,v_2,v_3,\cdots,v_n)$, this paper explores the following feature representations, and the librosa library of Python is used for feature extraction:

Mean: the average value of the feature $\mathbf{V}$, calculated as shown in equation (5):

$$\mu(\mathbf{V}) = \frac{1}{n}\sum_{i=1}^{n} v_i \tag{5}$$

Standard deviation SD: a measure of the distribution of the feature $\mathbf{V}$, calculated as shown in equation (6):

$$\sigma(\mathbf{V}) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(v_i - \mu(\mathbf{V})\right)^2} \tag{6}$$

MFCCs are very useful features for tasks such as speech recognition. First, the short-time Fourier transform STFT of the signal is obtained using $window\_size = 2048$, $hop\_size = 512$ and a Hann window. In this paper, after the experimental computation of the power spectrum, a triangular Mel filter bank is applied to simulate human perception of sound, and finally, a discrete cosine transform is applied to the logarithms of the energies of all the filter banks to obtain the MFCCs. MFCCs.In this paper, the parameter $nmels$ corresponding to the number of filter banks is set to 20.

3.4. Simulation experiments

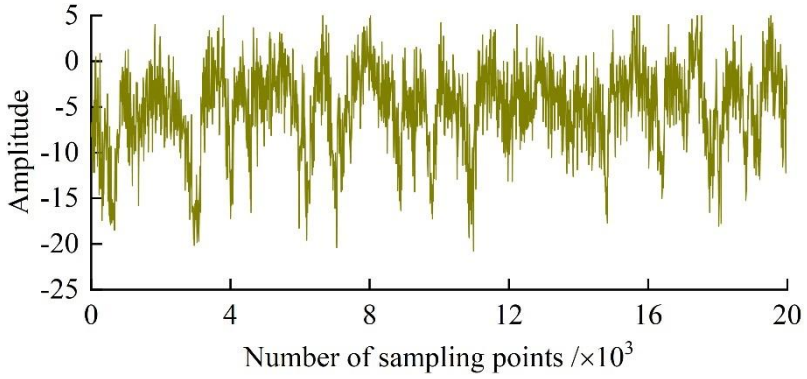
In order to validate the effectiveness of the MGTN minority music classification model, this section investigates and compares different speech features and classification methods in order to determine the best classification method applicable to minority music in Chinese regions. The experimental hardware platform is an Intel CoreTM i7 processor with a main frequency of 3.4GHz and 32GB of RAM, running MatlabR2023a on a Windows operating system.

The experiment firstly extracts the 40-dimensional MFCC, 12-dimensional LPCC, 1-dimensional short-time average energy and 1-dimensional short-time average over-zero rate parameters of the ethnic minority folk songs from 10 geographical regions, namely Guangdong, Guangxi, Guizhou, Hunan, Jiangxi, Inner Mongolia, Qinghai, Shaanxi, Sichuan and Tibet, with a total of 54-dimensional features as feature vectors. The whole folk song dataset has 3000 samples, each folk song contains 300 samples, each sample data is 55 dimensions, the first 54 dimensions are audio features, the last 1 dimension is the classification labels of the MGTN model, and 1~10 represent the types of folk songs of 10 different geographical regions, respectively, and the feature vectors of the folk songs are randomly divided into the training set and the test set according to 7:3.

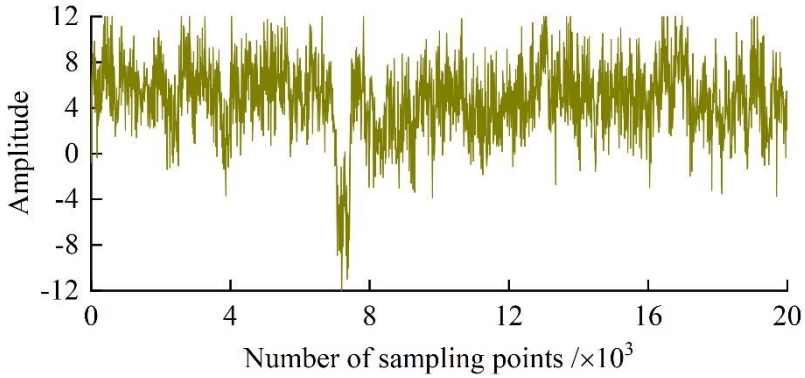
3.4.1. Comparison of audio features

The values of MFCC, LPCC, short-time average energy and short-time average over-zero rate characteristic parameters of three regional folk songs, namely, Guangdong, Shaanxi and Sichuan, are shown in Figs. 5-8, in which (a)~(b) all denote Guangdong, Shaanxi and Sichuan, respectively.

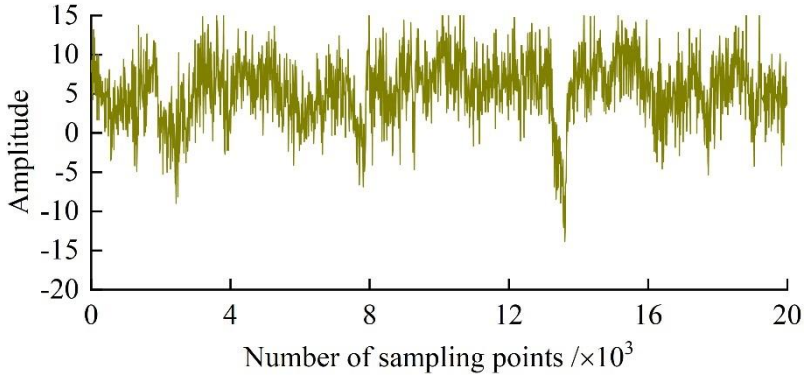
The MFCC parameters of the five regional folk songs are obtained after a series of processing operations, such as pre-emphasis, frame-adding windowing, and fast Fourier transform, etc. for the continuous speech signals. As can be seen from Fig. 5, the amplitude of MFCC in the same regional folk songs is different in each signal frame. As a whole, the amplitude changes of three different regional folk songs are even more different.



(a) Guangdong folk songs



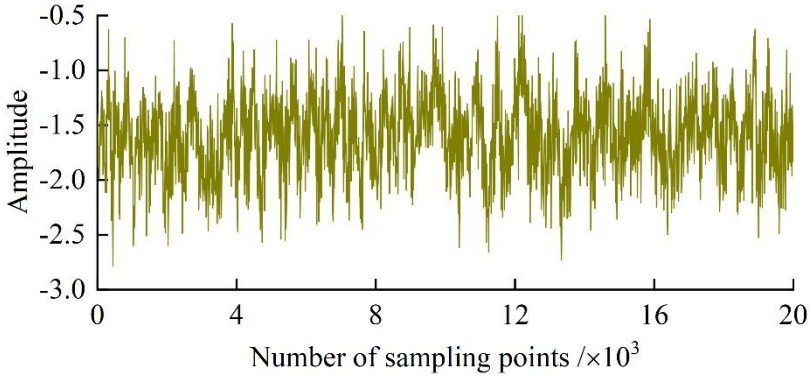
(b) Folk songs from northern Shaanxi



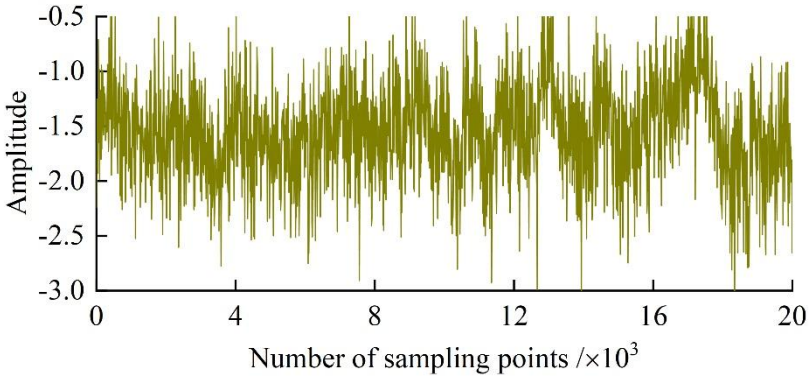
(c) Sichuan folk songs

Figure 5. *MFCC* values of three regional folk songs.

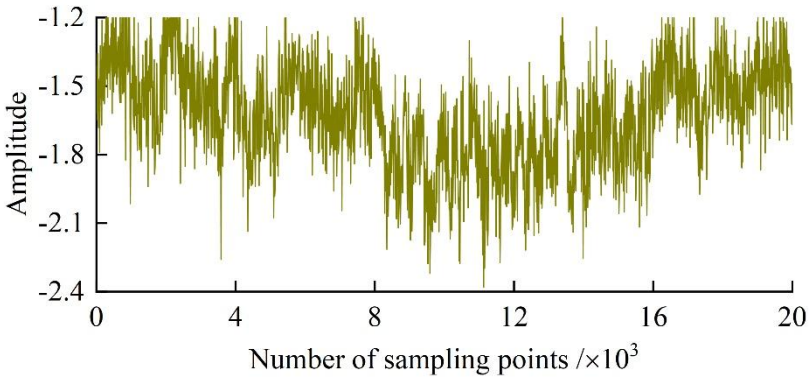
The speech characterization coefficient LPCC can well describe the principle of human vocalization and has good robustness in the problem of minority folk song classification. As can be seen from Fig. 6, the LPCC parameters of Guangdong folk songs are more centralized, while those of Sichuan folk songs behave more discrete.



(a) Guangdong folk songs



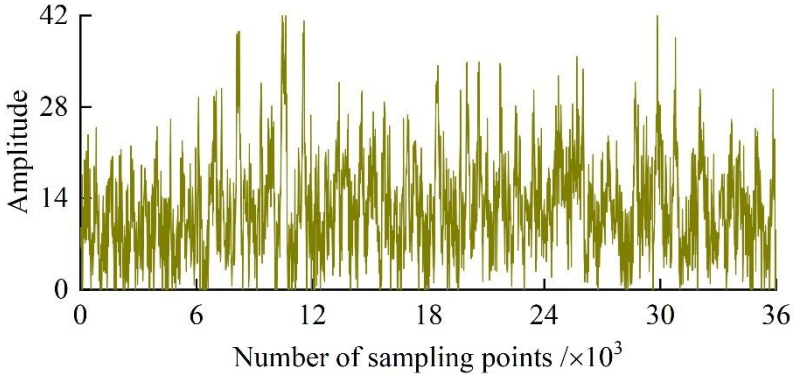
(b) Folk songs from northern Shaanxi



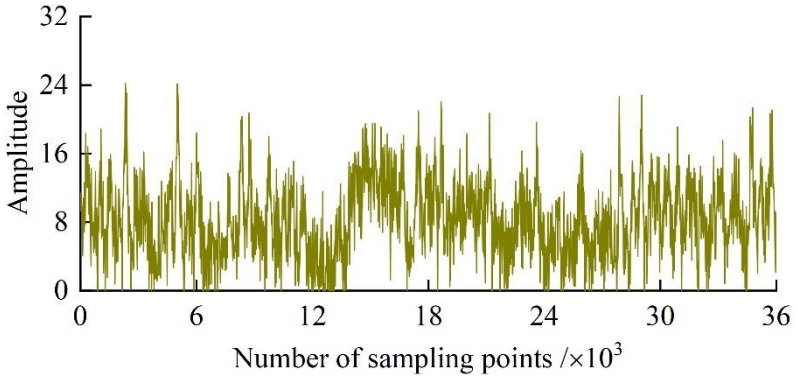
(c) Sichuan folk songs

Figure 6. *LPCC* values of three regional folk songs.

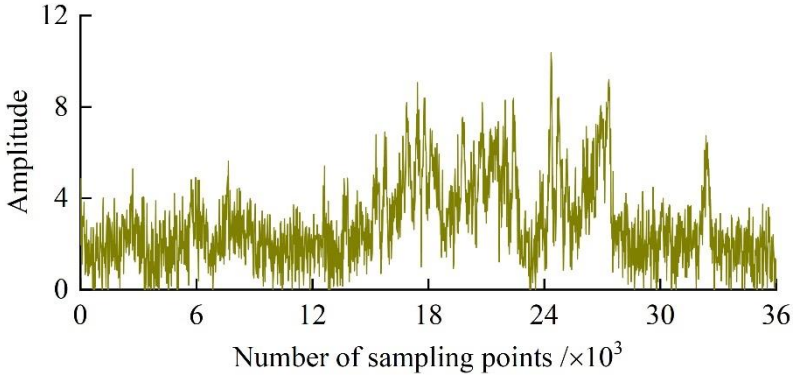
The folk songs from three different regions were sub-framed, and the short-time average energy of each frame was calculated, and the time of each frame after sub-framing was calculated to plot the short-time average energy of the folk songs. As can be seen from Fig. 7, the short-time average energy variation of Guangdong folk songs is more prominent, while the short-time average energy variation of Shaanxi folk songs is less obvious.



(a) Guangdong folk songs



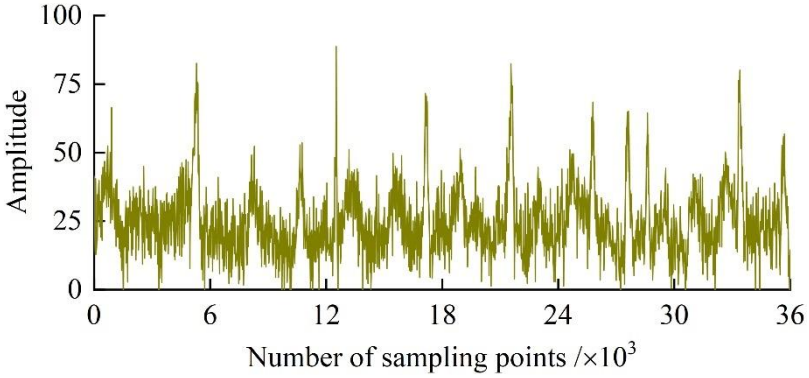
(b) Folk songs from northern Shaanxi



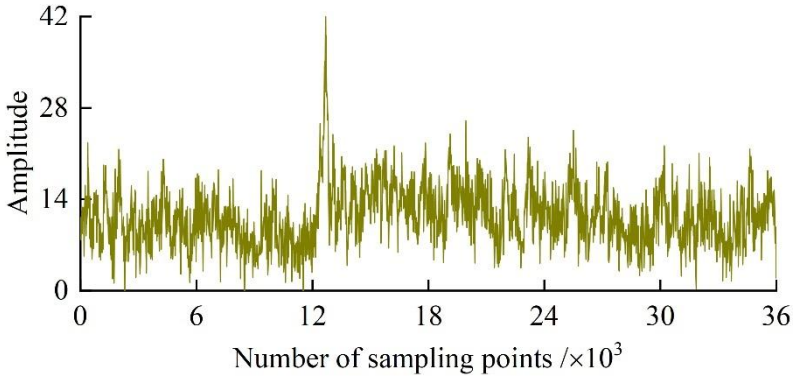
(c) Sichuan folk songs

Figure 7. Short-term average energy values of three regional folk songs.

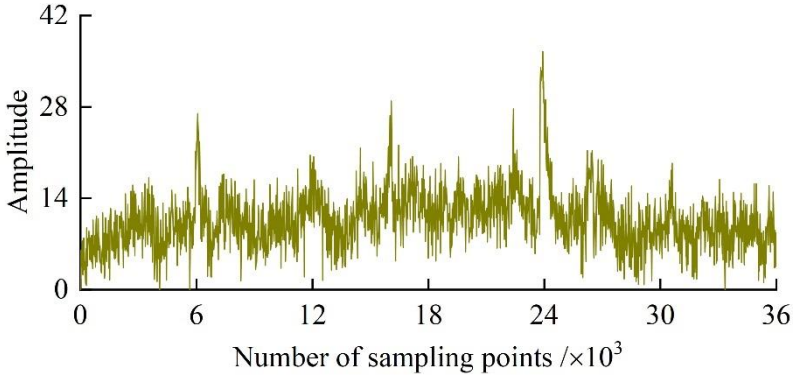
Over-zero in continuous speech signal means that the waveform of the signal in the time domain passes through the time axis, because high frequency means higher short-time average over-zero rate and low frequency means lower short-time average over-zero rate, and it can be considered that the clear and turbid tones are related to the high and low short-time average over-zero rate. Therefore, by analyzing the short-time average zero crossing rate of folk songs from different regions in Fig. 8, it is possible to distinguish the clear and turbid tones in folk songs from different regions.



(a) Guangdong folk songs



(b) Folk songs from northern Shaanxi



(c) Sichuan folk songs

Figure 8. Short-term zero-crossing rates of three regional folk songs.

The four kinds of features, *MFCC*, *LPCC*, short-time average energy and short-time average zero crossing rate, are combined separately, and Transformer is chosen as the classification model to carry out the classification experiments of 10 regional folk songs, and the correct rate of folk song classification under each combination of features is obtained through the simulation results as shown in Table 1.

It can be seen that the classification correct rate of Chinese minority folk songs is the highest when combining the four feature coefficients of *MFCC*, *LPCC*, short-time average energy and short-time average zero crossing rate as feature vectors, reaching 81.56%, which is better than that of other feature combinations, and better reflects the differences and characteristics of the folk songs of different regions.

Table 1. Classification results of folk songs of algorithms using different feature combinations.

Feature combination	Classification accuracy rate /%
MFCC	69.73
LPCC	44.94
MFCC+*LPCC*	78.25
MFCC+ Short-term average energy + Short-term average zero-crossing rate	74.18
MFCC+*LPCC*+ Short-term average energy +Short-term average zero-crossing rate	79.56

3.4.2. Comparison of classification models

(1) Simulation results of MGTN classification model

In this section, the effectiveness of the folk song classification model MGTN is verified on the problem of ethnic minority folk song classification. Classification experiments are carried out on 10 regional styles of folk songs, and the results such as classification prediction results and confusion matrices obtained for the test set of folk songs are shown in Figs. 9~10. The axes 1~10 in the figure represent the folk song samples from Guangdong, Guangxi, Guizhou, Hunan, Jiangxi, Inner Mongolia, Qinghai, Shaanxi, Sichuan and Tibet, respectively.

The number of samples in the test set in Fig. 9 is 900, and the string of data points with the largest consecutive length in each folk song sample indicates the result that the predicted classification matches the actual classification. By comparison, it can be concluded that the classification correctness of the

final test set reaches 85.11%, which is improved by 5.55% compared with the 79.56% of the standard Transformer model. The prediction bias of the experiment is small, and the classification results achieve good results. Figure 10 Each matrix element $Y_{a,b}$ of the confusion matrix represents the number of folk songs with actual category a predicted category b. Obviously, the diagonal elements of the confusion matrix then represent the number of folk songs that are predicted accurately. From the confusion matrix, it can be seen that category 1 and category 2 have the highest rate of categorization correctness, i.e., Guangdong folk songs and Guangxi folk songs have the best rate of categorization correctness, both at 95.56%.

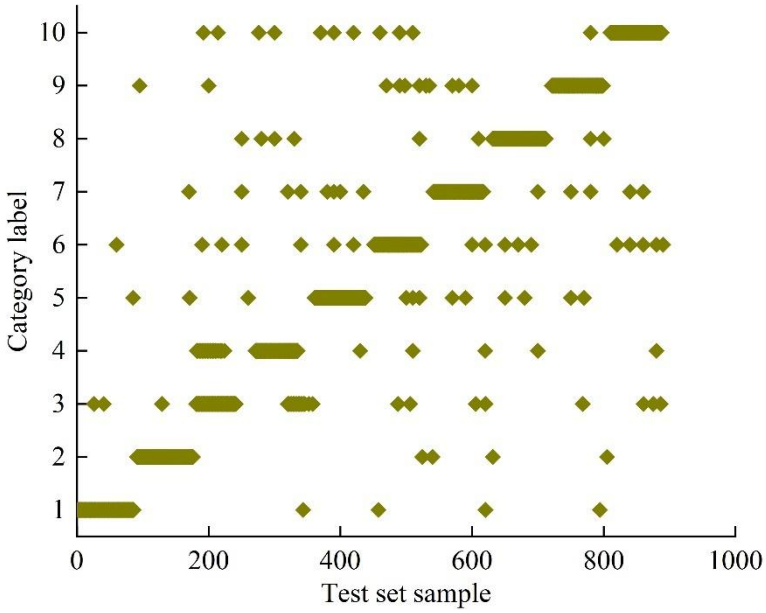


Figure 9. Prediction bias plot of the MGTN model on test set.

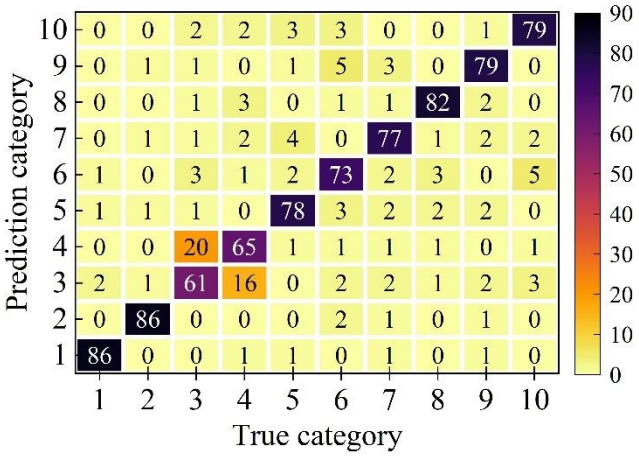


Figure 10. Confusion matrix of the MGTN classification model.

(2) Comparison of classification performance of different models

Keeping the dataset unchanged, comparing the classification accuracy of different models, the obtained experimental results of different models are shown in Table 2.

It can be seen that the MGTN model in this paper has a higher classification accuracy rate compared to other models, and has better results in the task of Chinese minority music classification. Meanwhile, compared with the models that utilize a single Image Encoder and TS Encoder, the classification accuracy of this paper's model is improved by 4.86% and 2.55% respectively, which verifies the effectiveness of this paper's model improvement method.

Table 2. Comparison of operation results of different models.

Model	Training data volume	Test data volume	Classification accuracy rate /%
VGG-16 CNN	2100	900	65.24
Random Forest	2100	900	74.51

Logistic Regression	2100	900	77.83
Transformer	2100	900	79.56
SVM	2100	900	79.72
NNet2	2100	900	79.83
MGTN Image Encoder (ours)	2100	900	80.25
LSTM+SVM	2100	900	81.12
BRNN+PCNNA	2100	900	81.43
MGTN TS Encoder (ours)	2100	900	82.56
MFMCNN	2100	900	84.72
MGTN (ours)	2100	900	85.11

4. Automatic Generation of Minority Folk Songs Based on Markov Chains

In this chapter, a model for automatic generation of folk song melodies of Chinese ethnic minorities is designed based on Markov chain algorithm, and experiments on automatic generation of folk song melodies are conducted based on this model.

4.1. Composition process

Markov chains [19] are relatively common mathematical statistical probabilistic methods used for sequence prediction. Without considering other complex conditions, music can be regarded as a sequence consisting of pitch and duration, so this chapter uses first-order Markov chains for the study of Chinese minority music generation.

Assuming that the sequence of states at moment x_t is the sequence of states at moment j, x_{t-1} is the sequence of states at moment i, and from the definition of the Markov model, the state j at moment x_t is only related to the state i at moment x_{t-1}, the transfer probability is mathematically described as follows:

$$p_{ij} = p(i \to j) = p\left(x_t = j \mid x_{t-1} = i\right) \tag{7}$$

Multiple transfer probabilities can form a matrix of transfer probabilities, and the state transfer matrix p is represented as follows:

$$P = \left[p_{ij} \right]_{k \times k} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1k} \\ p_{21} & p_{22} & \cdots & p_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1} & p_{k2} & \cdots & p_{kk} \end{bmatrix} \tag{8}$$

Among the above formulas, k denotes the total number of states, and $p_{ij}(i, j = 1, 2, 3, \cdots, k)$ denotes the transfer from the current state i to the next state. Where the probability of j should satisfy the following conditions:

$$0 < p_{ij} < 1 \tag{9}$$

$$\sum_{k}^{j} p_{ij} = 1 \tag{10}$$

According to the logic of Markov chain and the conditions necessary for generating Chinese minority music melodies in this chapter, the Markov composition process is shown in Figure 11. Firstly, the sample song set is constructed by collecting and processing classical ethnic minority folk songs, and then the songs of the same tonality are obtained through the pre-processing of the song melody, which are analyzed to derive the probability matrix of pitch as well as time value in the melody, and the probability matrix is used to conduct the experiments on the generation of the sequence of pitch and time value, which is ultimately transformed to obtain the melodic segments of the music.

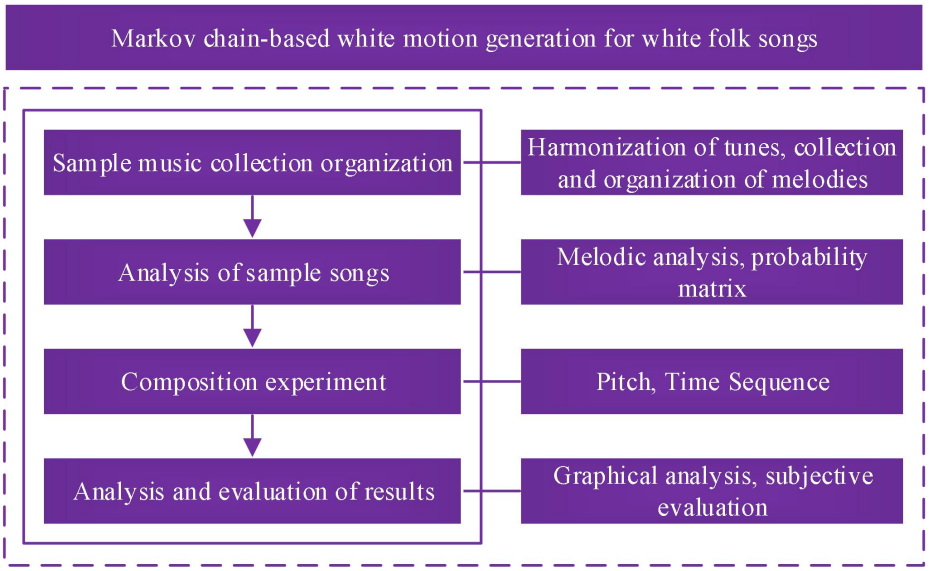


Figure 11. Automatic generation process of folk songs based on Markov Chain.

4.2. Analysis of Sample Songs

4.2.1. Obtaining Pitch Sequences and Time Value Sequences

In the processed sample melodies, the pitches of all the notes in each song are extracted separately to form a pitch sequence. Now take the sample song "Yasasai" as an example, the obtained monophonic melody is composed into a pitch sequence as follows:

$$S_{p_1} = \left\{ c^2, a^1, g^1, \cdots, c^1, a \right\} \tag{11}$$

where S_{p_1} denotes the set of all note pitches in the song "Yasasai". The series in it denotes the order in which the pitches appear as well as the number of them.

Using the same method, the sequence of note tensors of the sample song "Yasasai" can be obtained as shown below:

$$S_{t_1} = \left\{ \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \cdots, \frac{1}{8}, \frac{1}{4} \right\} \tag{12}$$

where S_{t_1} denotes the set of all note tensors in the song "Yasasai", where the series denotes the order in which the note tensors appear as well as the number of note tensors, and the note tensors correspond to the pitch order.

4.2.2. Getting the note distribution

In this subsection, the note distribution consists of the pitch and the time value of the note, and the Markov model requires the acquisition of the initial distribution of the pitch and the time value, respectively.

The acquisition of the initial pitch distribution satisfies the following conditions:

$$D\left(p_m\right) = \frac{pn}{PN} \tag{13}$$

where the initial pitch distribution is the ratio of the number of a certain pitch in a certain pitch sequence to the number of all pitches in the sequence. $D\left(p_m\right)$ is the distribution of a particular pitch in a particular pitch sequence, pn denotes the number of times a particular pitch occurs in that pitch sequence, and PN is the total number of pitches in the pitch sequence.

The initial time-valued distribution is obtained satisfying the following conditions:

$$D\left(t_m\right)=\frac{tn}{TN} \tag{14}$$

where the initial time-value distribution is the ratio of the number of a certain time value in a certain time-value sequence to the number of all time values in the sequence. $D\left(t_m\right)$ is the distribution of a certain time value in a certain time value sequence, tn denotes the number of times a certain time value occurs in that time value sequence, and TN is the total number of time values in the time value sequence.

After subjecting all sample songs to the above process, the initial distribution of overall notes and time values will be obtained, and the total distribution will satisfy the following conditions:

$$Dpk=\frac{\sum pn}{\sum PN} \tag{15}$$

where Dp_k is the distribution of a given pitch across all sample song pitch sequences. $\sum pn$ is the number of a given pitch in the sequence of all sample song pitches, and $\sum PN$ is the total number of all pitches in the sequence of sample song pitches.

The set SDp is the initial distribution of all note pitches in the entire set of sample songs:

$$SDp=\left\{Dp_1,Dp_2,Dp_3,\cdots,Dp_k\right\} \tag{16}$$

where Dt_k is the distribution of a particular time value in the sequence of time values of all sample songs. $\sum tn$ is the number of a particular time value in the sequence of time values of all sample songs, and $\sum TN$ is the total number of all time values in the sequence of time values of the sample songs. Then there are:

$$Dt_k=\frac{\sum tn}{\sum TN} \tag{17}$$

The set SDt is the distribution of all note tensors over the entire set of sample songs:

$$SDt=\left\{Dt_1,Dt_2,Dt_3,\cdots,Dt_k\right\} \tag{18}$$

4.2.3. Obtaining the note transfer probability matrix

The note transfer probability distribution is obtained in a similar way as the initial distribution of pitch.

From Eq. (7), the transfer probability between two states is P_{ij}, therefore, the transfer probability pitch transfer probability distribution of the note is obtained in the following way:

$$P_{ij_n}=\frac{(i,j)}{(i,x)} \tag{19}$$

Where P_{ij} denotes the shift probability distribution of the note $i\to j$ in a particular sample pitch sequence, and (i,j) denotes the number of times the event occurs in the sequence starting with the pitch i and the next tone being j. (i,x) denotes the number of times an event occurs in a sequence that starts with pitch i and the next tone is any pitch x, and the ratio of the two is the transfer probability of pitch $i\to j$.

By counting each sample song and summing up all P_{ij}, the transfer probability Pp of the note in all sample songs can be obtained as shown in Equation (20), at this time, the pitch transfer probability matrix of all sample songs is shown in (21):

$$Pp=\sum Pij \tag{20}$$

$$Pp = \left[p_{ij} \right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} \\ p_{21} & p_{22} & \cdots & p_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} \end{bmatrix} \tag{21}$$

Using the same method one can obtain the matrix of transfer probabilities of note time-values in the sequence.

Where P_{ij} denotes the transfer probability distribution of the note $i \to j$ in a particular sample time-valued sequence, and (i, j) denotes the number of times the event that starts with the time-value i and the next time-value is j occurs in the sequence. (i, x) denotes the number of times an event occurs in a sequence that starts with the time value i and the next time value is any time value x, and the ratio of the two is the transfer probability of the time value $i \to j$.

By counting each sample song and summing up all P_{ij}, the transfer probability P_t of the note's time value in all sample songs can be obtained as shown in Equation (22), and at this time, the matrix of time-value transfer probability is shown in (23) in all sample songs:

$$Pt = \sum Pij \tag{22}$$

$$Pt = \left[p_{ij} \right] = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1i} \\ p_{21} & p_{22} & \cdots & p_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} \end{bmatrix} \tag{23}$$

4.3. Construction of a new melody

The construction of the new melody is based on the transfer probability matrix of note pitch and time value, which constitutes a pitch sequence and a time value sequence according to the relationship between note pitch and time value, respectively, and transforms the two into a musical score.

From the Markov model, if there is any note i, the next note j needs to be acquired. According to the transfer probability matrix, starting with the note i, the next note is inside the matrix, and the note with the largest transfer probability among the events starting with the note i constitutes the state $i \to j$. Repeat this step to obtain the pitch sequence $Psong$ and the time value sequence $Tsong$ of the note:

$$Psong = \{ p_1, p_2, \cdots, p_n \} \tag{24}$$

$$Tsong = \{ t_1, t_2, \cdots, t_n \} \tag{25}$$

Because the sample song is a Chinese minority folk song, note repetition occurs in the construction of the note sequence, a situation that does not conform to the rules of composition. In this regard, this chapter adds a process of randomly selecting the next note when constructing the note sequence, so that the case with the highest probability is not always selected when constructing the note sequence.

For example, when calculating the sequence of time values, the initial probability distribution of the time values and the transfer probability distribution of the sample songs are collected and combined with the current time values to predict the next time values. The time values are represented as whole notes, half notes, quarter notes, eighth notes, and sixteenth notes by (1, 1/2, 1/4, 1/8, and 1/16), respectively. After obtaining the sequence of pitches and time values, the two correspond to each other to obtain a new sequence of notes. This is expressed in the following form:

$$Song = \left\{ \left(a, \frac{1}{2} \right), \left(c^1, \frac{1}{8} \right), \cdots, \left(n, \frac{n}{x} \right) \right\} \tag{26}$$

The pitch sequence is computed in a similar way, with the final step yielding a combination of the pitch sequence and the time value sequence of the song. The sequences are eventually converted into a musical score.

4.4. Experimental results and analysis

4.4.1. Experimental evaluation methods

In order to verify the feasibility of the experimental model in this paper in generating melodies of Chinese ethnic minority folk songs, it is necessary to make a comparative evaluation of the generated results, and here we use the standard Markov model, the attention_mn model in the Magenta project, and the improved Markov model in this paper to do a comparative analysis of the melodies. In order to ensure the fairness of the evaluation, 12 melodies were generated by each of the three methods, all using the collected 90 Chinese ethnic minority folk songs dataset, and choosing the same parameters to control the key, tempo, beat number and other information of the generated melodies, and evaluating the 12 randomly generated melodies.

In this paper, a combination of subjective and objective evaluation is used to assess the generated works. Firstly, subjective evaluation, the aesthetics of a musical work is a process from shallow to deep, from intuitive experience to emotional expression and then to resonance, for these 3 levels, 5 indicators often used to evaluate music are used, which are the ethnicity of the melody, structure, pleasantness to the ear, association, and resonance. Since aesthetic differences exist objectively, 12 students each from music majors and non-music majors were invited to give ratings to the generated melodies in each of these 5 areas.

Objective evaluations are usually designed to compensate for the aesthetic differences and subjectivity of subjective evaluations. Here, based on the rules of music theory and related concepts of music aesthetics combined with peer research, the generated melodies were objectively assessed using three indicators: the relational relevance of neighboring phrases, R, the degree of intervals conformity, I, and the degree of rhythmic sparsity intervals, D. The results of the objective evaluation were presented in the form of the following three indicators.

4.4.2. Experimental results

Twelve melodies were generated using the traditional Markov model and the attention_rnn model, of which two MIDI files were in Sibelius pentatonic form, and the melodies were in 2/4 time. The Improved Markov Model (IM-Markov) of this paper was used to generate 12 melodies, of which 4 MIDI files were in Sibelius pentatonic form, and the melodies were all in 2/4 time.

The score of subjective evaluation is set to 1-10 points, and the subjective evaluation results of the composition examples are shown in Table 3, where the values are the average values. It can be seen that the evaluation results of each subjective index of the model in this paper are better than the other two models.

Table 3. Comparison of subjective evaluation results.

Subjective evaluation index	Markov		attention_rnn		IM-Markov	
	Music major	Non-music major	Music major	Non-music major	Music major	Non-music major
National character	7.34	7.62	6.71	7.95	8.72	9.24
Structural	6.47	6.95	8.44	8.62	9.61	9.73
Pleasantness	5.84	6.46	7.35	7.35	8.71	9.08
Degree of association	5.23	6.35	6.69	7.30	7.26	7.35
Degree of resonance	5.21	5.64	6.41	6.96	7.32	7.38

Based on the formula: comprehensive evaluation value = average value of music major evaluation * 0.6 + average value of non-music major evaluation * 0.4, the subjective comprehensive evaluation results are obtained as shown in Table 4.

Analyzing the subjective comprehensive evaluation data, it can be seen that this paper's model improves about 19.87%~45.48% and about 5.34%~23.86% in five subjective evaluation indexes compared with the traditional Markov model and attentions_nn, respectively. Overall, this paper's model improves about 14.91%~34.65% on the overall subjective composite evaluation metrics compared to attention_mnn model and Markov model.

Table 4. Comparison of subjective comprehensive evaluation results.

Subjective evaluation index	Model		
	Markov	attention_rnn	IM-Markov
National character	7.45	7.21	8.93
Structural	6.66	8.51	9.66
Pleasantness	6.09	7.35	8.86

Degree of association	5.68	6.93	7.30
Degree of resonance	5.38	6.63	7.34

The objective comprehensive evaluation data are shown in Table 5, and all the data in the table are average values. Analyzing the evaluation data, it can be seen that the model in this paper has obvious improvement in R, I and D indexes compared with the traditional Markov model and attention_rnn model, and the overall objective comprehensive evaluation indexes have been improved by a factor of 0.89~1.47.

Table 5. Comparison of objective comprehensive evaluation results.

Objective evaluation index	Model		
	Markov	attention_rnn	IM-Markov
R	2.24	6.27	8.31
I	0.85	0.66	0.95
D	3.16	1.25	6.17

In conclusion, the improved Markov model proposed in this paper has certain advantages in the generation of melodies for Chinese ethnic minority folk songs, which are summarized as follows:

(1) Clear modulation: the modulation of the melody generated by the IM-Markov model is in the Gong, Levistic, and Symbolic modulation, which is in line with the modulation characteristics of Chinese ethnic minority folk songs, while the modulation of the melody generated by the traditional Markov model and the attentional_rnn model is not obvious.

(2) Accurate interval relationship: the melody generated by attention_rnn model is smaller than the index solution strategy of I. Although the melodies generated by IM-Markov model and Markov model both satisfy the index solution strategy of I, by analyzing the interval relationship between adjacent notes, it is found that the interval relationship between adjacent notes of the melody generated by Markov model can be an octave relationship, which can reduce the quality of the musical work. Comparatively speaking, the model in this paper is more accurate than the Markov model and the attention_rnn model in grasping the intervals of the Chinese minority folk songs, and the melodies generated are more characteristic of the minority music melody.

(3) Rhythmic regularity: The rhythm of the melodic phrases generated by the IM-Markov model is regular, and most of the phrases in each melody end with a short rhythm into a long rhythm. On the other hand, the Markov model and attention_mnn model generate melodies with more scattered rhythms, which makes it difficult to have a clear grasp of the phrase length, and also results in the melodies not having a relatively stable sense of suspension.

(4) Clear melodic structure: the melody generated by the IM-Markov model consists of two sections, and the melodic line trend of the first section of one of the melodies generated by this model is shown in Fig. 12, and the melodic line trend of the second section is shown in Fig. 13. By comparing the melodic lines of the corresponding phrases of the two sections, the second section is a varied repetition of the first, with a high degree of correlation between the phrases within each section; the second phrase is a repetition or modal progression of the first; the third phrase shows the function of transitions, which is the culmination of the whole section; and the fourth phrase serves as a summary, which is typical of the rising and falling structure. In contrast, the Markov model generates melodies anchored on microscopic notes, so the correlation between phrases is lower and the global structure is less relational. The melodies generated by the attention_rnn model depend on the quantity and quality of the data, and due to the limited number of existing ethnic minority folk songs in China and the lack of a publicly available database of ethnic minority folk songs, the generated melodies show some meaningless repetitions and a relatively less obvious section structure.

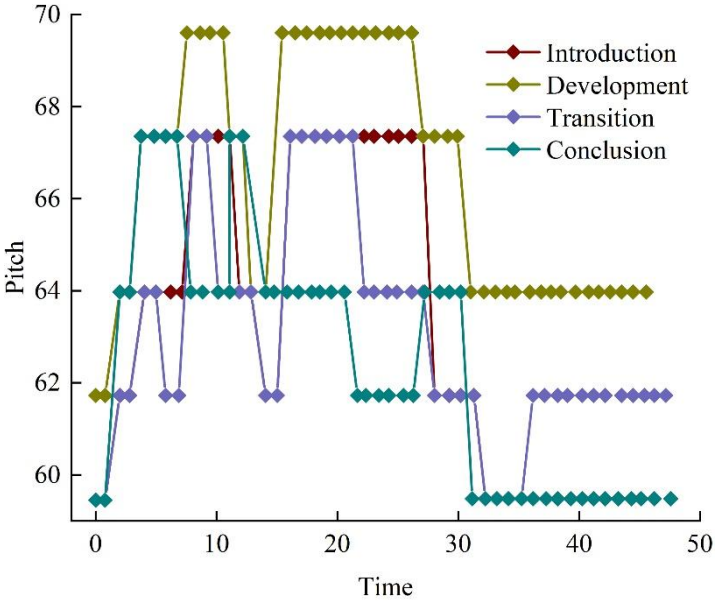


Figure 12. The melodic line trend of the first section.

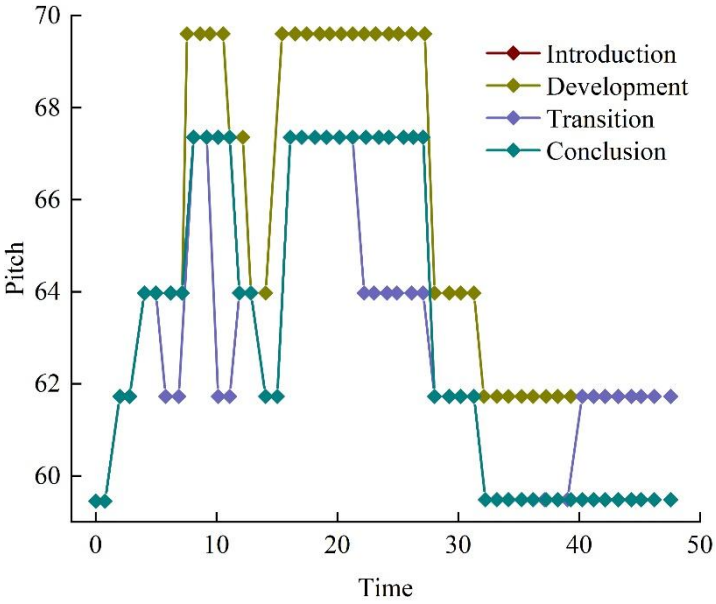


Figure 13. The melodic line trend of the second section.

5. Conclusion

In the field of music education, AI has the potential to be applied in intelligent music analysis, personalized teaching and intelligent music composition, etc. Incorporating ethnic music culture into music education and AI technology for inheritance is an important breakthrough for the innovative development of Chinese ethnic minority music. This paper focuses on the two aspects of intelligent music recognition and automatic generation to explore the innovative mode of AI-driven Chinese minority music inheritance, and proposes the music genre classification model MGTN and the intelligent composition model based on IM-Markov.

The MGTN model finally achieves 85.11% correct classification rate for Chinese minority folk songs, which is 5.55% higher than the 79.56% of the standard Transformer model. Meanwhile, compared with the MGTN model utilizing a single Image Encoder and TS Encoder, the classification accuracy of the MGTN model in this paper is improved by 4.86% and 2.55% respectively, which verifies the effectiveness of the model design in this paper.

The IM-Markov model proposed in this paper for generating folk music achieves 14.91%~34.65% and 0.89~1.47 times improvement in the overall subjective comprehensive evaluation index and

objective comprehensive evaluation index, respectively, and it has the advantages of tonal clarity, accurate intervallic relationship, rhythmic regularity, and clear melodic structure in generating the melody of Chinese ethnic minority music.

The development of intelligent categorization and automatic generation technology for Chinese minority folk songs requires more in-depth research and exploration, and the future work of this research is:

(1) Data feature improvement. In this paper, when constructing the model, the main consideration is the interval relationship and rhythmic pattern data composed of the pitch and time value of the notes, respectively, and in the future, more note features, such as tone intensity and timbre, can be considered to make the generated music more stylistic features.

(2) Emotional and rhythmic expression. Chinese ethnic minority folk songs are often emotionally rich and rhythmically beautiful forms of music that express people's emotions and life experiences. However, how to make AI algorithms understand and express these emotions and rhythms is a further issue that needs to be studied.

References

1. Wong, C. F. (2020). Hearing the minorities in modern Chinese music. International Communication of Chinese Culture, 7(2), 117-131.
2. Yin, L., & Guo, R. (2025). An Artificial Intelligence-Based Interactive Learning Environment for Music Education in China: Traditional Chinese Music and Its Contemporary Development as a Way to Increase Cultural Capital. European Journal of Education, 60(1), e12858.
3. Zhang, C. (2025). The analysis of Chinese National ballad composition education based on artificial intelligence and deep learning. Scientific Reports, 15(1), 9215.
4. Huang, R. S., Sturm, B. L., & Holzapfel, A. (2021, November). De-centering the West: East Asian Philosophies and the Ethics of Applying Artificial Intelligence to Music. In ISMIR (pp. 301-309).
5. Xu, Y. (2023). Exploration of the Influence of Music Communication Methods on the Inheritance of Minority Music. Art and Performance Letters, 4(8), 44-49.
6. Guo, D., Buarabha, H., & Wannapipat, W. (2024). Cross-border integration and cultural inheritance: the historical evolution and artistic innovation of Chinese Peking Opera films. Library of Progress-Library Science, Information Technology & Computer, 44(3).
7. Zhang, L. X., Yang, Y., & Leung, B. W. (2025). Innovative music classroom teaching in China. In Innovative Teaching and Classroom Processes (pp. 130-145). Routledge.
8. Hill, R., Betts, L., & Gardner, S. (2015). Empowerment and enablement through digital technology in the generation of the digital age. Computers in Human Behavior, 48, 1-23.
9. Xiao, Y. (2025). Research on the Innovation Model of Music Education Empowered by Information Technology. Journal of Sociology and Education, 1(8).
10. Gamaliia, K., Turchak-Lazurenko, L., Lavrenyuk, O., Penchuk, O., & Lytvynenko, N. (2023). Synergy of design, culture, and innovation in pedagogy: New horizons for education. Research Journal in Advanced Humanities, 4(4), 175-190.
11. Pistola, T., Diplaris, S., Stentoumis, C., Stathopoulos, E. A., Loupas, G., Mandilaras, T., ... & Kompatsiaris, I. (2021, May). Creating immersive experiences based on intangible cultural heritage. In 2021 IEEE International Conference on Intelligent Reality (ICIR) (pp. 17-24). IEEE.
12. Stavrakis, E., Aristidou, A., Savva, M., Himona, S. L., & Chrysanthou, Y. (2012, October). Digitization of cypriot folk dances. In Euro-Mediterranean Conference (pp. 404-413). Berlin, Heidelberg: Springer Berlin Heidelberg.
13. Shi, H. (2021, August). Research on the digital presentation and inheritance of traditional music in the internet era. In Journal of Physics: Conference Series (Vol. 1992, No. 4, p. 042041). IOP Publishing.
14. Chang, W. (2025). The integration of artificial intelligence and ethnic music cultural inheritance under deep learning. Computer Science and Information Systems, (00), 36-36.
15. Peining, M., Ghani, D., & Siman, Y. (2025). Digital Inheritance of Chinese Folk Music: Opportunities and challenges of multimedia technology. Environment- Behaviour Proceedings Journal, 10(SI32), 93-97.
16. Qingtang, L., Xinghan, Y., Linjing, W., Fengjiao, T., & Liang, C. (2023). Construction of Intelligent Application Service System ofEthnic Instrumental Music Culture from the Perspective ofDigital Protection and Inheritance. Libraly Journal, 42(386), 113.
17. Shi, Y. (2019, April). Research on the Inheritance Method of Minority Music and Dance Art based on Motion Capture Technology. In 1st International Symposium on Education, Culture and Social Sciences (ECSS 2019) (pp. 218-222). Atlantis Press.
18. Jiyang Chen, Xiaohong Ma, Shikuan Li,Sile Ma, Zhizheng Zhang & Xiaojing Ma. (2024). A Hybrid Parallel Computing Architecture Based on CNN and Transformer for Music Genre Classification. Electronics,13(16),3313-3313. https://doi.org/10.3390/ELECTRONICS13163313.
19. Adhika Sigit Ramanto & Nur Ulfa Maulidevi. (2017). Markov Chain Based Procedural Music Generator with User Chosen Mood Compatibility. International Journal of Asia Digital Art and Design Association,21(1),19-24. https://doi.org/10.20668/adada.21.1_19.