International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-16

https://doi.org/10.70917/ijcisim-2025-0247
Article

Research on the Innovation Mode and Development Path of Ethnic Music Education in the Digital Environment of the New Era

Dan Shen *

Harbin University, Harbin, Heilongjiang, 150086, China; S123456ddd123@163.com

Abstract: In the digital environment of the new era, there are problems that the education model of ethnic music cannot adapt to the needs of students and the teaching content is single. The article first takes the flipped classroom as the basis, and takes the students of major class 1 and class 2 of a music academy as the research object, and designs the ethnomusicology flipped classroom teaching experiment. In order to promote the diversification of the teaching content of ethnic music, this article proposes an intelligent generation model of ethnic music based on MCT-GAN, which utilizes MFCC to extract the audio features of ethnic music, and combines the temporal structure model with the multi-track correlation model, so as to realize the intelligent generation of ethnic music. It is found that the ethnic music generated by the model is more consistent with the chromaticity diagram of the original song, and the concordance, fluency and structure are acceptable to the audience. The application of flipped classroom in the teaching of folk music can significantly improve students' music performance, independent learning ability and music core literacy ($P<0.01$). The active introduction of intelligent technologies and teaching platforms in the digital environment can help promote the innovation of the teaching mode of ethnic music and realize the high-quality development of ethnic music education.

Keywords: MCT-GAN; MFCC; temporal structural modeling; flipped classroom; ethnic music

1. Introduction

Chinese folk music education has been developed to a certain extent in recent years, but due to the influence of historical, social and cultural factors, there are still big gaps in the popularization degree, discipline construction, faculty and teaching methods of folk music education, which is difficult to satisfy the current growth of the demand for folk music talents in the society [1-4]. Rapid changes in the modern social environment and the rapid development of information dissemination have also brought great impact on the folk music inheritance mechanism and education mechanism [5-6]. With the continuous emergence of new musical elements and art forms, more flexible and open innovative thinking and teaching concepts are needed to promote the development of folk music education [7-8]. At present, with the continuous development of information technology, digital education has become a hot spot and trend in the field of music teaching, and the digital application and practice of ethnic music teaching is gradually popularizing and achieving remarkable results [9-11].

Introducing digital technology into folk music teaching can not only provide students with more convenient and diversified learning means, so that students can learn the construction of melody, the use of harmony, and the grasp of rhythm more intuitively [12-14], but also guide students to learn from traditional folk music, providing students with a richer, more three-dimensional, and more personalized music education experience [15-16]. This move helps to maintain the characteristics and artistic charm of traditional Chinese folk music and realize the formation of national music styles in the new era. The development of digital education and the promotion of digital transformation of education is a general

trend [17-18]. The digital application of ethnic music teaching can also realize the organic integration of modern music education and information technology, and promote the inheritance and development of ethnic music culture [19-20]. In the future, the digital application and practice of ethnic music teaching will continue to deepen and expand, providing more opportunities to promote the development and innovation of ethnic music education, and at the same time providing essential cultural sustenance for ethnic music education [21-24].

In this regard, literature [25] elaborates on the application of digital music platforms in ethnomusicology, which is of great significance for the dissemination and preservation of ethnomusicology. Literature [26] examined the teaching of ethnic music under the concept of digital health education, and improved the teaching of ethnic musicology in colleges and universities by analyzing the development of ethnic musicology and enhancing the status of the music discipline, aiming to enhance the national self-confidence. Literature [27] points out that traditional music education does not go in the inheritance of national music culture, and examines the role of deep learning and other technologies in the inheritance of national music culture, revealing the superiority of deep learning in the protection and inheritance of national music culture. Literature [28] emphasizes the importance of preserving and transmitting ethnic music, and cites deep learning technology to verify its impact on the preservation and transmission of ethnic music, and the results of the study show the accuracy and effectiveness of deep learning. Literature [29] synthesized the trends, technologies, and global perspectives shaping ethnomusicology education, and assisted thematic analysis through the use of ATLAS, and the results pointed out that, although ethnomusicology education has made effective progress in terms of technology and global perspectives, there are still challenges in balancing traditions and innovations. Literature [30] introduced intangible cultural heritage (ICH) music and analyzed its educational function and practical value in music teaching in higher education, and the results emphasized that ICH music plays an important role in cultivating students' artistic literacy, creative thinking, and cultural identity. Literature [31] examined the digital application of ethnomusicology in higher education based on literature analysis, interviews, and other methods, and revealed, through literature review and other methods, that digital technology has a significant contribution to the ethnomusicology performance mode in terms of cultural inheritance, education and teaching. Literature [32] examined the impact of the application of digital tools in music education, and based on the literature review showed that most of the studies reported the positive impact of digital tools on the development of musical skills.

In this study, in order to enhance the diversity of ethnic music in the digitized environment, an intelligent generation model of ethnic music based on generative adversarial network, the MCT-GAN model, is proposed. The model takes the audio features of ethnic music extracted by the MFCC algorithm as input, and completes the intelligent generation of ethnic music based on the consideration of multi-track correlation and music generation time structure. The generated ethnic music is then combined with the flipped classroom and applied to the ethnic music flipped classroom teaching model, which aims to further promote the improvement of students' music performance, independent learning ability and music core literacy. In this way, we can realize the innovative development of the teaching mode of folk music, and also provide a new paradigm for the revitalization and inheritance of folk music.

2. Design of innovative models of folk music education

National culture is the basic guarantee for the survival of the nation and the country. In the social environment of economic globalization and market diversification, the mingling and borrowing of different cultures have played a good role in promoting the prosperous development of the cause of national music education. Students in colleges and universities are the future force for building the motherland, and whether they have a good sense of national culture and music literacy determines the future development of individuals and the whole nation. Therefore, under the digitalized environment of the new era, actively exploring the innovative mode of ethnic music education can help to enhance students' understanding of ethnic music and cultural literacy.

2.1. Flipped Classroom Teaching of Folk Music

2.1.1. Concept of Flipped Classroom

Flipped classroom, in fact, is to reform and adjust the traditional teaching process. In the traditional classroom education model, the teacher teaches first and students review after class. In the flipped classroom, students study first and then the teacher explains. That is to say, “teach first and then learn” becomes “learn first and then teach”. Teachers prepare introductory video materials before the new class starts and distribute them to students, who study the relevant content by collecting online materials, record their own puzzles and confusions encountered in the process, and finally discuss the recorded

content with the teacher and classmates in the classroom [33].

The flipped classroom has caused two significant changes to teaching and learning, namely, emphasizing students' independent learning and reorienting the teacher-student relationship. The pre-study process of the flipped classroom allows students to learn effectively according to their own characteristics, to identify problems and actively seek solutions to them, and to enhance their ability to learn independently. Teachers will no longer be the dominant player in teaching, but will become the guide of teaching, and the teacher-student relationship will be transformed into a friend relationship with equal communication.

2.1.2. Modes of teaching music

Under the digital environment of the new era, ethnic music education mode can realize a greater degree of change and innovation by relying on the flipped classroom. The design of the teaching link of the ethnic music flipped classroom is directly related to whether the teaching program is operated effectively and properly. This paper takes the previous research on the flipped classroom teaching model as a reference, combines the characteristics of the discipline, and designs the teaching model of the ethnic music flipped classroom as shown in Fig. 1. Ethnic music flipped classroom teaching model mainly includes a pre-course independent learning stage, a cooperative inquiry stage and a post-course summary and reflection stage, each stage is fully student-oriented and teacher-assisted, in order to maximize the effect of ethnic music education.

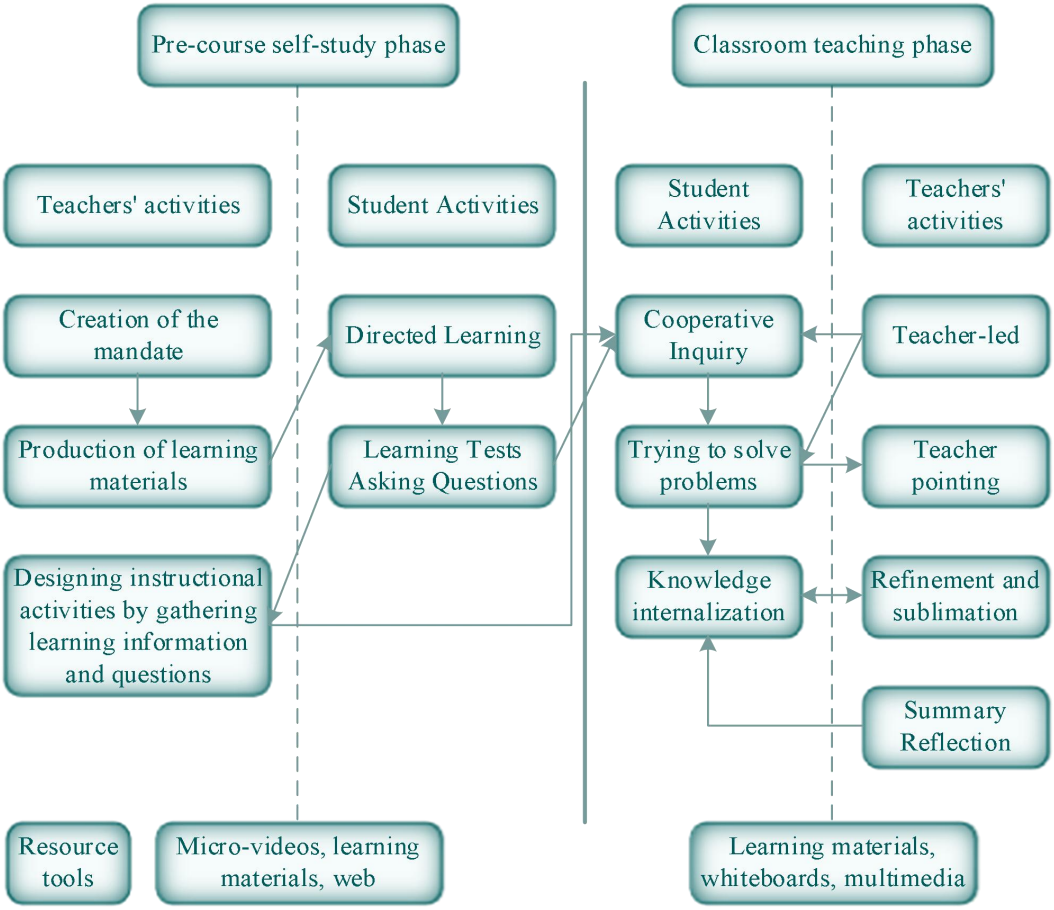


Figure 1. National music flipped classroom teaching model.

(1) The stage of independent study before class. Teachers should, according to the requirements of the teaching objectives, combined with the students' learning conditions, make learning materials for students' independent study. The design and production of learning materials should be centered on the teaching of the important and difficult points, so that the theme of the materials is clear and the objectives are prominent. The length of the micro-video should be controlled in 5-10 minutes, the selection of micro-video can be from the network platform, or by the teacher according to the actual situation of the

students in the class, recording their own micro-lesson video.

(2) Cooperative inquiry stage in the class. Based on the feedback of students' pre-course learning and teaching objectives, teachers create activities that are in line with the teaching theme, encourage group cooperative learning, and try to solve pre-course problems in cooperation. If they encounter problems that are difficult to solve, teachers give appropriate guidance and help. Throughout the process, teachers should pay attention to individual guidance for difficult students.

(3) Summarizing and reflecting stage after the lesson. After the end of the teaching session, teachers should consciously summarize and reflect on the effect of learning materials, students' classroom performance, the achievement of the class standards, whether the activities are in line with the theme and other aspects of the situation, and constantly improve and enhance the level of teaching.

2.2. Experimental design for teaching folk music

2.2.1. Design of teaching experiments

(1) Experimental subjects. In this study, a total of 80 students in two classes, majoring in Class 1 and Class 2 of the School of Music of a university in S Province, with 40 students in each of the two classes, were used as experimental subjects. It was found by the pre-test that there was no significant difference between the ethnic music performance and the core literacy level of the students in these two classes, which satisfied the basic conditions for conducting the experimental study.

(2) Experimental time. After communicating with the teachers, it was determined that the appropriate teaching content would be selected to conduct the ethnomusicology flipped classroom teaching experiment during the second half of the semester in 2024.

(3) Experimental tools. Ethnic music achievement test paper, which mainly contains the quantification of four dimensions, such as rhythm, music notation, beat, and music literacy, and each dimension is scored out of 10 points, and it was distributed to students in two classes before and after the beginning of the teaching, and its purpose is to understand the change of the students' ethnic music achievement.

(4) Experimental variables. The professional 1 class of the experimental group was taught by the flipped classroom teaching method, and the professional 2 class of the control group was taught by the traditional teaching method. During the experimental period, both classes were taught by the same teacher with the same teaching content and used the same class time for teaching, thus better controlling the control variables and making the results obtained from the study more accurate and effective.

2.2.2. Teaching process design

Applying the flipped classroom teaching mode to the teaching of folk music should be implemented in the following aspects when conducting teaching experiments:

(1) Define teaching objectives and plan teaching scientifically

According to the ethnomusicology curriculum standard, the three-dimensional objectives of cultivating students are knowledge and skills, process and method, and emotional attitude and values. The teaching objectives and appreciation contents are all aimed at guiding students to establish good aesthetic views and values. In order to improve students' own aesthetic ability, attention should be paid to cultivating students' three-dimensional goals during the teaching process.

(2) Understanding the actual situation of students and rationalizing teaching

In the understanding of the usual teaching, students will use many network software to discuss and exchange in the time after class. Students are not very satisfied with the existing music teaching mode, and they hope that the music class is interesting and creative, and that many music works can be shown in video. Through practical understanding, it is known that the teaching mode of flipped classroom can be used in the teaching of ethnic music.

(3) Design micro-lesson videos to guide independent learning

Although the micro-lesson video is only a few minutes' content, it is actually a very important part in the flipped classroom. If you want to achieve the expected effect, so that students can find the relevant classroom content in the video as well as the teaching key points of the lesson, then you need teachers to pay attention to the reasonable arrangement of the video content when making the video, and you need to be able to let students communicate and interact better in the classroom.

(4) Discussion and exchange summary, internalization of classroom knowledge

After watching the micro-teaching video, students need to go home to complete the homework assigned by the teacher, check the relevant information of the course content, review the problems again in class, exchange and discuss, and group activities. The assignments should be centered on the relevant knowledge points, and the questions should be designed in a hierarchical manner so that students' familiarity with the course content can be easily checked and students can be guided to think and study on

their own.

2.2.3. Questionnaire construction

In order to further clarify the influence of the ethnomusicology flipped classroom teaching model on students' music core literacy and independent learning ability, this questionnaire was set up from two aspects respectively. For the effect of students' independent learning ability cultivation, it was mainly studied from four dimensions: learning motivation, learning planning, learning environment, and learning reflection. For students' music core literacy, it was mainly quantified from the degree of music enjoyment, music basic quality, music class learning, feeling of music class, and comprehension of music (HX1~HX5).

The questionnaire was presented in the form of a five-point Likert scale (1-5 points), and a total of 40 questionnaires were distributed to the students in the experimental classes before and after the teaching experiment began, and 40 questionnaires were retrieved, with the effective questionnaire recovery rate reaching 100%. Moreover, when the questionnaire was formally distributed, the reliability of the questionnaire was also verified, and the results showed that the reliability of the questionnaire amounted to 0.874, the KMO value was 0.917, and the significance of the Bartlett's sphericity test was 0.000. The above results show that the questionnaire of the present paper has a good degree of reliability and validity, and the data obtained can provide accurate conclusions for the study.

3. Construction of an intelligent generation model for folk music

The rapid development of digital society provides new opportunities for the reform of folk music teaching. In ethnic music flipped classroom teaching, relying on the online teaching platform can provide teachers and students with diversified teaching resources of ethnic music, on the basis of which teachers and students can use the digital intelligent model to generate ethnic music that meets their own expectations after class, so as to better promote the diversification of ethnic music resources, to provide a new paradigm for the inheritance and innovation of ethnic music, and to help improve the quality of ethnic music education.

3.1. Characterization dataset for folk music

3.1.1. Audio Characterization

For a piece of music, the descriptive information (music name, synopsis, lyrics) as well as the audio itself can represent its unique properties. Among them, the audio features are the most unique and effective information representation of a piece of music, through which this paper can make the maximum distinction between different pieces of music. Extracting audio features of music is a key step for the research goal of this paper. Audio features contain both time domain and frequency domain categories, in order to fully extract audio features, this paper will look for feature representation methods that can express both time domain and frequency domain information of audio.

The original music audio files are stored as one-dimensional time-domain signals in formats such as MP3. In order to get a frequency domain signal representation of the music, the Fourier transform can be used. However, it is not possible to show the relationship between the audio frequency over time. To address this problem, many literatures have proposed time-frequency domain transform methods, including short-time Fourier transform, Laplace transform and so on. Transforming the sound signal into an image representation can better extract features using convolutional neural networks. Among them, short-time Fourier is the most commonly used transform method, which converts the audio signal into an acoustic spectrogram representation by time-frequency transform to obtain the frequency-energy distribution of audio on the time axis.

3.1.2. Audio Feature Extraction

Meier inverted spectral coefficient method is to analyze the sound signal from the perspective of spectral transformation, the spectral transformation detailed process for the sound signal $x(n)$ first through the Fourier transform, will be convolved with the processing, to obtain the multiplicative signal, that is:

$$x(n) = H(n) \times (n) \tag{1}$$

Taking logarithms on both sides of equation (1) at the same time, the equality sign holds, then:

$$\begin{aligned} \lg[|x(n)|] &= \lg[|H(n) \times E(n)|] \\ &= \lg[|H(n)|] + \lg[|E(n)|] \end{aligned} \tag{2}$$

The multiplicative signal is converted to an additive signal and then the additive signal Fourier inverse transformed to:

$$x(n) = h(n) + e(n) \tag{3}$$

where $h(n)$ is the Mel inverse spectral coefficient and $e(n)$ is the logarithmic energy of the output of the k rd Mel filter.

The Mel filter bank is a collection of triangular filters divided according to the Mel scale, and the process is to process the spectrum $X(k)$ squared obtained from the discrete Fourier transform to convert the signal's non-uniform frequency into the Mel frequency, which is converted into the Mel domain by the Mel filter bank to calculate its logarithmic energy, and the Mel frequency conversion formula is:

$$f_{\mathrm{m}} = 2595 \cdot \lg\left(1 + \frac{f}{700}\right) \tag{4}$$

Where f_{m} is the Mel frequency and f is the actual frequency of the signal.

Mel inverted spectral coefficient method is to discrete cosine transform the logarithmic energy of each filter output, but in practice, the standard Mel inverted spectral coefficients for the static existence, in order to characterize the dynamic characteristics of the sound signal, the first need to analyze the difference spectrum of the static characteristics of the sound signal, so as to obtain the dynamic differential parameters of the sound signal [34]. The dynamic differential parameters of the sound signal are solved by the formula:

$$d_t = \begin{cases} C_{t+1} - C_t t, & t < K \\ \dfrac{\sum_{K=1}^{K} k\left(C_{t+k} - C_{t-k}\right)}{\sqrt{2\sum_{k=1}^{k} k^2}}, & \text{Other} \\ C_t - C_{t-1}, & t \mp Q - L \end{cases} \tag{5}$$

where d_t is the first-order difference parameter of the t nd MFCC, C_t is the t th MFCC parameter, L is the order of the MFCC, Q is the MFCC prescribed order, and K is the first-order derivative time difference, which is taken as 1 or 2.

3.1.3. Music characterization data

The purpose of generative modeling in application is to find the similarity between the music so that the generated music is more in line with the user's needs, then theoretically similar music should have a similar music feature set. However, music features can be considered as a kind of temporal features, so the similarity of this music feature set can be considered as a kind of relative similarity, which does not mean that there is similarity in a certain period of time, but there is similarity in general. This requires that more attention be paid to the totality of features in the process of constructing the music feature set in order to serve the music intelligent generation model well.

In the actual application process, the box dimension approximation of the feature set is usually used to find the fractal dimension of the feature set, and the calculation process is as follows:

Assuming that datasets $D = \{A, B, f\}$, A represent a feature set with n attributes $\{A_1, A_1, \ldots, A_n\}$, E represent an object set with k tuples, and f represent categorical attributes, if we map B into a n -dimensional space and ρ equate $(\rho = 2, 4, 8\ldots)$ each dimension, then we get $h \times \rho$ cells, and then number each h -dimensional cell in turn according to the order of the feature set, such that the feature record $R_i = \{R_{i1}, R_{i2}, R_{i3}, \ldots, R_{in}\}$ corresponding to the cell in the n -dimensional space order can be obtained from $\{R_{i1} / R_1, R_{i2} / R_2, R_{i3} / R_3, \ldots, R_{in} / R_n\}$ where R_h is the value taken by the attribute A_h of the feature set after being ρ equipartitioned, and the number of points into which the

j th cell falls is noted as $N(\rho) = \sum_j (c, j)^2$, then the fractal dimension of the feature set is denoted as:

$$D(s) = -\frac{\partial \log(s(\rho))}{\partial \log \rho} \tag{6}$$

After feature extraction, assuming that there is m attribute in the feature set, the box dimension of the feature set is calculated to be h , then the effective feature dimension of the feature set can be considered to be h . Here the strategy used in this paper is not to discard the $(m-h)$ redundant features as is usually done, but to reduce the weight of the redundant features according to the relevance, so as to improve the accuracy of the prediction of the information.

3.2. Intelligent Generation Model for Ethnic Music

3.2.1. Generating Adversarial Networks

Generative Adversarial Network (GAN) is an unsupervised generative model which consists of two parts: generator (G) and discriminator (D) [35]. The generator and the discriminator are two mutually independent networks, which are trained separately and alternately during training. The generator and discriminator can use either artificial neural networks based on multilayer perceptron or deep convolutional neural networks. When the model is trained, random noise z of fixed dimensions is sampled from the latent space obeying a specified distribution is input to the generator network G. The generator is trained to maximize the distribution of the output generated samples G(z) to fit the real data distribution. The discriminator is a true-false dichotomous classifier, which inputs the generated samples G(z) and the true samples x into the discriminator network, and trains the discriminator to correctly discriminate between the true samples and the false samples as much as possible. GAN optimizes the model by letting the generator and the discriminator play with each other, and the two optimizers need to be defined to optimize each other separately in the training, and fixing the parameters of one model to update the other, and their performance improves continuously during training until reaching N The performance of the two is continuously improved during training until Nash equilibrium is reached. The generator learns the distribution of real data well, and the discriminator has difficulty in distinguishing whether the input samples are from real data.

The loss function of GAN can be expressed as:

$$\max_G \min_D V(G,D) = E_{x\sim P_{(data)}(X)}\left[\log D(x) + E_{Z\sim P_z(z)}[\log(1-D(G(z)))]\right. \tag{7}$$

Where $p_{(data)}(x)$ represents the data distribution of the real sample in the data set, x represents the real data, $p(z)$ represents the data distribution in the latent variable space (which usually obeys a Gaussian distribution), z represents the random noise data sampled from the prior distribution, and $G(z)$ represents the generated sample obtained by inputting random noise z into the generator. $D(x)$ and $D(G(z))$ are the discriminative results obtained by inputting the real sample and the generated sample into the discriminator, i.e., the probability that the sample is real or not. The discriminator is trained to maximize this loss function, i.e., maximize $\log D(x)$ so that the probability of discriminating the true sample as true is larger, and maximize $\log(1-D(G(z)))$ so that the probability of discriminating the generated sample as true is smaller. The generator's goal is to minimize this loss function, i.e., minimize log(1-D(G(z))) so that the discriminator has a greater probability of discriminating the generated samples as true, thus optimizing the generator network to generate more realistic samples. Compared with other traditional generative models, the generative adversarial network does not require complex Markov chains, the model is trained based on an unsupervised approach, and has the advantage of generating clearer and more realistic samples.

3.2.2. Time-structured models

Because music has the characteristic of time, music as a whole is also characterized by “succession and continuity”. Music attaches great importance to the coherence of music between different phrases and bars. Using the Pianoroll format as the generator format to generate music, the model generates music according to the music bars, and the music generated is likely to be incoherent between different bars. Therefore, a temporal structure-based music generation model is designed to incorporate a temporal structure generator into a generative adversarial network, which allows the model to generate multiple

consecutive music bars that make up a longer piece of music.

It is designed that when the generator generates a certain bar of music, a vector carrying temporal information will be generated sequentially by the temporal structure generator, which will be used as an additional dimension of the music bar generator in the ethnomusicological generative network model, and will be inputted into the music bar generator, thus generating music bars containing inter-bar information and combining them to form a fixed-length music phrase.

The temporal structure based GAN music generator G consists of two parts, G_{order} represents the temporal structure generator and G_{bar} represents the Scratch music bar generator, while the process of generator modeling the data distribution can be represented as:

$$G(Z)=\left\{G_{bar}\left(G_{order}\left(Z_t\right)\right)\right\}_{t=1}^{T} \tag{8}$$

where Z_t is the random vector on which the temporal structure generator depends, and T is the total number of bars contained in each musical phrase. G_{order} Using Z_t as a noise input, for each ethnic music bar a potential vector carrying timing information is generated. This potential vector will be input to the G_{bar} music bar generator to generate ethnic music bars in a certain order. Eventually a series of ethnomusic bars in Pianoroll format in chronological order will be generated.

3.2.3. MCT-GAN modeling

In order to realize the accurate generation of ethnic music, this paper further investigates a music generation model (MCT-GAN) that integrates the multi-track correlation model and time structure model, and its framework is shown in Fig. 2.

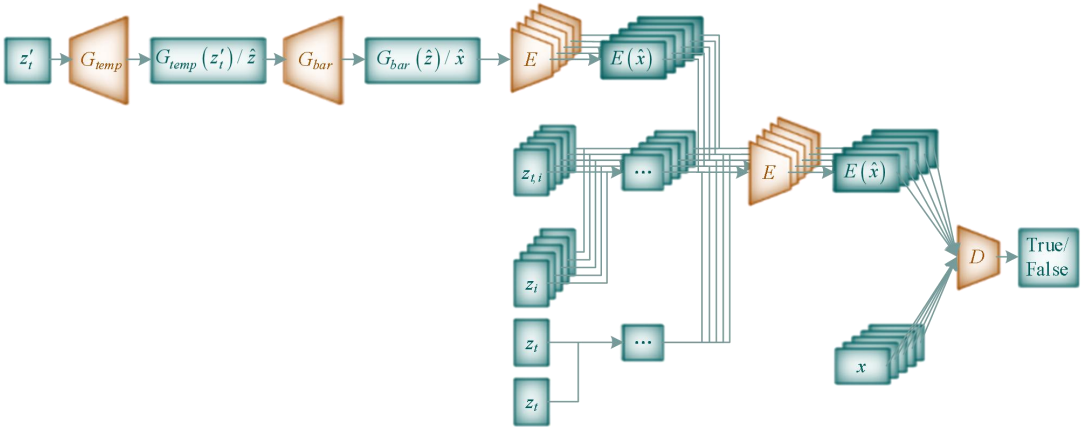


Figure 2. MCT-GAN model structure diagram.

The MCT-GAN network is divided into two main parts, the first part generates a specific track of music, G_{temp} with z_t' as input, G_{bar} with the output $G_{temp}\left(z_t'\right)$ (or $\hat{z}$) of G_{temp} as input, and then outputs a track of music $G_{bar}(\hat{z})$, or $\hat{x}$, viz:

$$\hat{x}^{(t)}=G_{bar}\left(G_{temp}\left(z_t'\right)^{(t)}\right)\forall t=1,2,\ldots,T \tag{9}$$

The second part consists of four inputs: time-independent random vector z between audio tracks, time-independent random vector z_i within audio tracks, time-dependent random vector z_t between audio tracks, and time-dependent random vector $z_{t,i}$ within audio tracks. First, connect the time-independent random vector z between audio tracks and the time-independent random vector z_i within audio tracks respectively to obtain $\hat{z}_i$, and connect the time-dependent random vector z_t between audio tracks and the time-dependent random vector $z_{t,i}$ within audio tracks to obtain $\hat{z}_{t,i}$. After

connection, it is input into the generator G together with $E\left(\hat{x}^{(t)}\right)$ generated in the first part and mapped by decoder E, and other audio tracks of music are generated in sequence. Then:

$$G_i^{(t)}(\overline{z}) = G\left(\hat{z}_i \circ \hat{z}_{i,t} \circ E\left(\hat{x}^{(t)}\right)\right) \forall t = 1, 2, \ldots, T \tag{10}$$

Finally, the discriminator D is trained using the data generated by G along with the real data $\tilde{x}$. Since MCT-GAN uses CT-GAN as its generative adversarial network model, then a consistency penalty term (CT) needs to be added to the objective function of MCT-GAN.

Denote by $d(a,b)$ the l_2 distance between a and b, if for the discriminator $D: x \mapsto y$ if there exists a constant $M \geq 0$ such that any $x, x' \in X$ that satisfies the inequality (11), then that case need not be penalized, and if it is not possible to satisfy that inequality, such cases need to be penalized, then:

$$d\left(D(x), D\left(x'\right)\right) \leq M \cdot d\left(x, x'\right) \tag{11}$$

The consistency penalty rule described above can be realized by adding the following consistency penalty term to the objective function. To wit:

$$CT\big|_{x,x'} = E_{x,x'}\left[\max\left(0, \frac{d\left(D(x), D\left(x'\right)\right)}{d\left(x, x'\right)} - M'\right)\right] \tag{12}$$

Overall, for MCT-GAN, random noise is used as the input to the generator, and the goal of the generator is to transform the random noise into a pitch-time matrix representation of Pianoroll to make it look like a Pianoroll representation of real music, and this transformation is realized by a special convolution operator in the transposed convolutional layer in the CNN. At the same time, the generated data in the form of Pianoroll produced by the generator is used as an input to the discriminator, which predicts whether it is from real or generated Pianoroll, and after that the prediction is fed back to the generator, so that the data generated by the generator after the generator update looks more realistic.

4. Ethnic music generation and validation of pedagogical effectiveness

The study of folk music is the foundation of future Chinese national music creation. Therefore, contemporary music colleges and music departments of comprehensive universities pay special attention to the teaching of folk music. However, many students are not interested in the study of folk music, showing indifference and even contempt. Therefore, the real situation requires us to reflect deeply on the teaching of this course and make exploratory reforms to try to change the basic work of the bad academic style.

4.1. Verification of Intelligent Generation of Ethnic Music

4.1.1. Feature extraction results

Experiments are conducted according to the structure of the audio feature extraction algorithm described in the previous section, and the data returned from the use of ethnomusicology is stored in Dataset and saved as a local file Dataset.dat. The hardware environment for the experiments is an Intel(R) Core(TM) i3-10500H 3.60 GHz processor and an NVIDIA GeForce TX 4090. The software environment is the deep learning framework PyTorch, Python version 3.8.

The crawled audio files are usually MP3 format files, which are converted to wav.format files with a sampling rate of 44000Hz by Python. The length of the music is generally about 3 to 5 minutes, and the dimension of the directly generated Mel Spectrogram is more than 18000x1000. For the Mel Spectrogram with large input dimension, the input dimension can be reduced by sliding the window or setting the appropriate step size to sample the Mel Spectrogram. The appropriate window size and step size can be selected according to the characteristics of the task and the temporal nature of the input data.

Figure 3 shows the original Mel spectrogram converted from a folk music spectrogram, and its dimensionality reduction processing can get 32 512*512 Mel spectrogram subplots. Figure 4 shows the sub-sequence map after dimensionality reduction processing. The dimensionality reduction process reduces the dimensionality of the input while retaining as much important information as possible, thus improving the training efficiency.

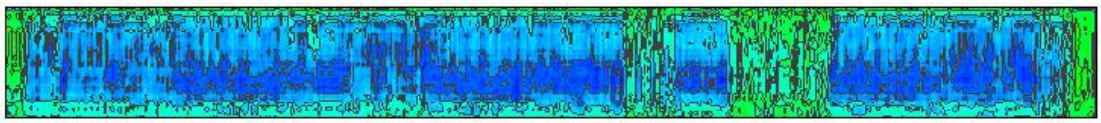


Figure 3. Original Mel Spectrum.

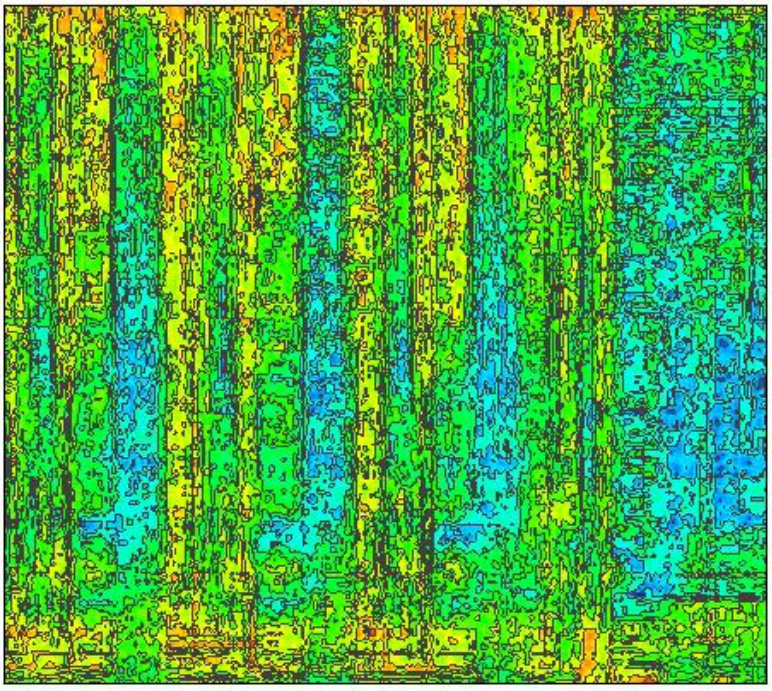


Figure 4. Mel frequency spectrogram after dimensionality reduction.

For the evaluation of ethnic music audio feature extraction algorithms, the three main dimensions are pitch estimation, voiced and unvoiced detection, and total accuracy. Among them, for the evaluation of pitch estimation, RPA and RCA will be used, for the detection of voiced and voiceless, VR and VFA will be used, and for the evaluation of pitch estimation as well as the detection of frames both in total, total accuracy OA will be used.

In order to verify that the audio feature extraction method for ethnic music proposed in this paper can improve the pitch estimation accuracy, the method of this paper is evaluated with RPA and RCA as evaluation indexes, and Auditory Streaming Cues (AudS) and Source/Filter Modeling (SFM) are selected as comparison algorithms, and in addition to the Dataset established in this paper, ORCHSET and MedleyDB are additionally selected as supplementary datasets. The results of different algorithms on different datasets are obtained as shown in Table 1.

As can be seen from the table, compared to the methods for auditory flow cues, the RPA of this paper's method on the three datasets is improved by 3.45%, 6.65%, and 3.09%, respectively, but it is lower by 4.08%, 1.37%, and 4.83%, respectively, when compared to the source/filter modeling methods. Since RCA maps pitch values to an octave range to calculate the pitch estimation accuracy, comparing the difference between RPA and RCA on the three datasets separately, the self-constructed dataset Dataset obtained a difference of 4.06%, 8.59%, and 5.77% from top to bottom, and the dataset ORCHSET obtained a difference of 10.19%, 19.07%, and 16.41% from top to bottom and the dataset MedleyDB gets 3.03%, 4.31% and 10.54%, it can be concluded that the method of this paper further reduces the multiplicative error.

Table 1. Pitch precision comparison (%).

Method	Dataset		ORCHSET		MedleyDB	
	RPA	RCA	RPA	RCA	RPA	RCA
MFCC	66.43	70.49	58.04	68.23	61.36	64.39
AudS	62.98	71.57	51.39	70.46	58.27	62.58
SFM	70.51	76.28	59.41	75.82	66.19	76.73

In order to verify the effectiveness of the algorithm of this paper, the algorithm of this paper is compared with C3, Sal, and Bit on the self-built dataset Dataset. The evaluation indexes selected in the previous paper are used to evaluate the algorithms, and the validation results of different algorithms on the dataset are obtained as shown in Figure 5.

From the figure, it can be seen that the introduction of the MFCC algorithm in this paper can effectively reduce the error rate of the sound frames in the audio feature extraction of ethnic music, but at the cost of the reduction of the recall rate. From the RPA and RCA metrics, it can be seen that this paper's algorithm improves the pitch estimation accuracy and reduces the octave error. Combining the other four evaluation metrics, this paper's algorithm obtains the maximum total precision (55.04%), which is also 4.66 percentage points higher compared to the C3 algorithm, which performs second best, indicating the effectiveness of this paper's algorithm for extracting audio features and melodies of ethnic music.

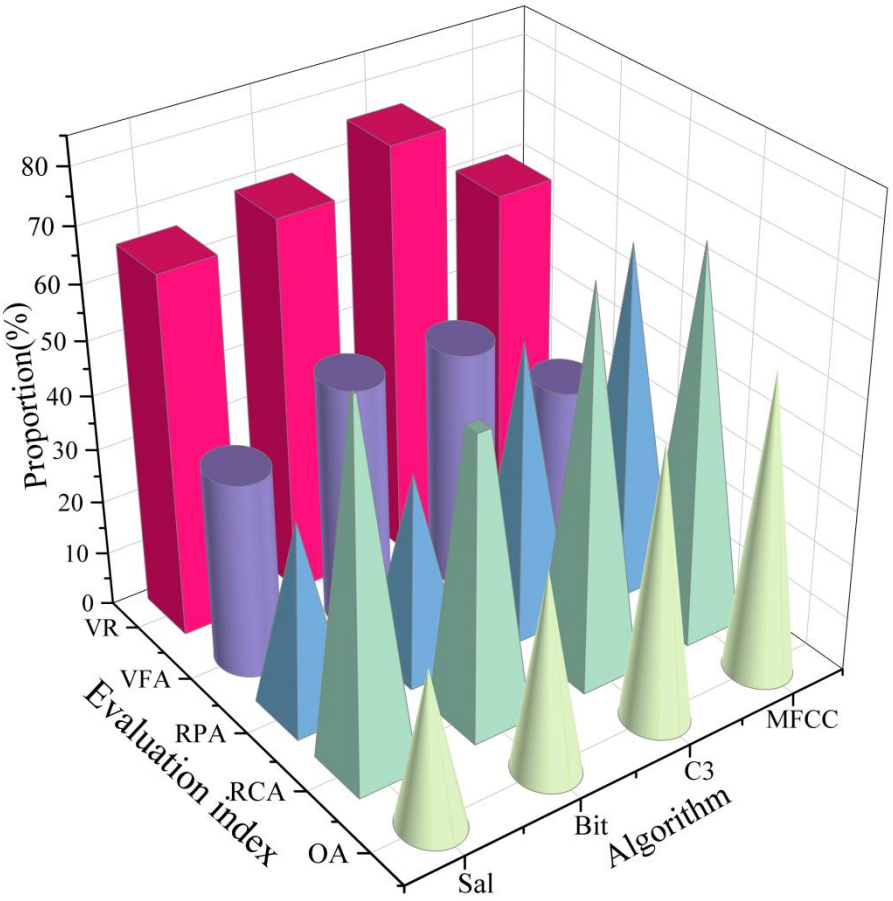


Figure 5. Results in the data set.

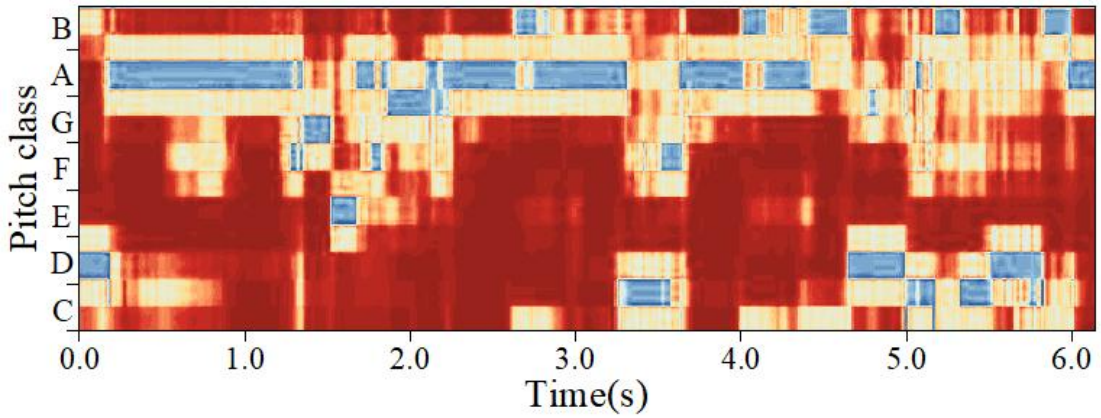
4.1.2. Music generation effects

In order to verify whether the MCT-GAN model is capable of generating ethnic music with the same pitch as the score, in this section, the real music clip corresponding to the score in the test set is used as the original song, and controlled experiments are conducted in terms of music pitch by comparing between the original song and the generated music.

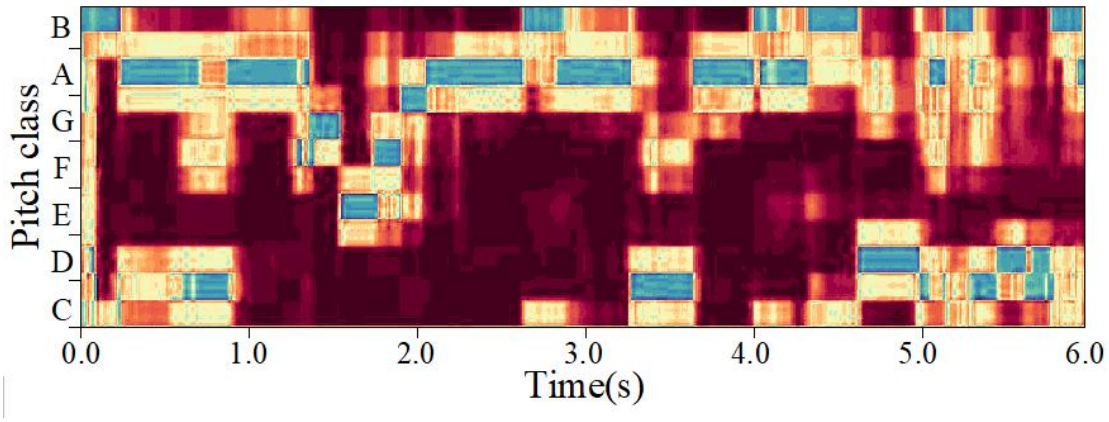
In order to verify that the MCT-GAN model is able to generate ethnic music with the same pitch as the score, this experiment uses the chromaticity map features to represent the ethnic music pitch, and then calculates the histogram similarity between the chromaticity map features of the generated music and the chromaticity map features of the original song in order to assess whether the generated ethnic music has the same pitch as the original song. The calculated numerical results range from 0 to 1, with higher values representing higher pitch similarity between the generated music and the original song. Specifically, in this experiment, 30 music scores are randomly selected from the test dataset, the chromaticity map features of their respective original songs are calculated, 30 audio clips of music with a length of 6 seconds are obtained and their chromaticity map features are calculated through the MCT-GAN model generation, and then the histogram similarity of the two chromaticity map features is calculated to

evaluate the pitch accuracy of the generated ethnic music. Fig. 6 shows a control example of a segment of ethnic music in the test set, where Figs. 6(a)~(b) show the chromaticity maps of the real music and the generated music, respectively.

As can be seen from the figure, the chromaticity diagram of the original song is more similar to that of the generated music, i.e., the pitch levels are matched. Overall, by calculating the histogram similarity between the 30 pieces of generated music and the original music in the test set, the average value of histogram similarity obtained in the dataset is 0.872. Not only that, this paper also conducts the pitch MOS test, and the results also show that the music generated by the model has a higher consistency with the source music in terms of pitch in terms of the human's subjective perception. Taken together, the above experimental results show that the generated folk music has similar pitch level with the source music, which means that the model better understands the pitch information in the input folk music data, and thus generates the corresponding pitch folk music.



(a) Original chromaticity diagram of the song



(b) Generate the chromaticity map of music

Figure 6. A comparison example of national music.

4.1.3. Assessment of music quality

In order to further illustrate the effectiveness of the intelligent generation model of folk music designed in this paper to generate music, 10 volunteers were selected from the students in the major 1 class to generate 10 pieces of folk music using the model, and were asked to rate (on a 10-point scale) the concordance, fluency, and structure of the generated folk music. In addition, 10 additional non-music majors were selected to give the same ratings to the 10 pieces of folk music, and the ratings of the two types of experimenters were combined to obtain the results of the volunteers' assessment of the quality of the generated folk music as shown in Fig. 7.

For such artworks as folk music, the quality of folk music is mainly measured through people's subjective feelings, and different people will get different evaluation results for the same piece of folk music. The scoring data were collected from the 10 participants described above using music evaluation

software, and from the box plots of the auditory scoring results, it can be seen that the participants with a foundation in music were more demanding than ordinary people in terms of the concordance of the generated folk music, and did not give high ratings on the indicator of concordance as the ordinary people did, and in the indicators of fluency and the overall structural nature of the folk music, the two types of participants gave The difference between the ratings is not obvious, and both of them gave high ratings, indicating that the quality of ethnic music generated by this paper based on the MCT-GAN model was recognized by the participants.

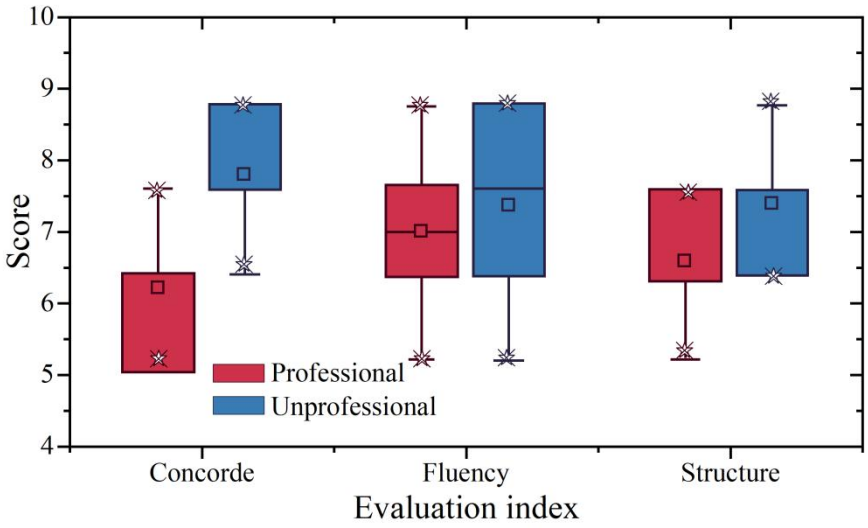


Figure 7. National music quality assessment results.

4.2. Teaching Effectiveness of Flipped Classroom Model

4.2.1. Changes in music scores

In this paper, in addition to carrying out the design of the intelligent generation model of ethnic music, we also designed a flipped classroom teaching model of ethnic music based on an online education platform. In order to verify the effectiveness of the teaching model, a teaching experiment was carried out. At the end of the teaching experiment, the students of specialty 1 and 2 classes were tested for music ability mastery assessment. The data before and after the beginning of teaching were counted, and the independent sample t-test was performed on the two scores, and the results of pre- and post-test data analysis were obtained as shown in Table 2.

As can be seen from the data in the table, the P-value of rhythm, music notation, beat, and music reading ability of professional class 1 and class 2 before and after the experiment are different, and the P-value of professional class 2 is greater than 0.05, which indicates that the variability of musical ability of professional class 2 before and after carrying out the teaching experiment is not obvious, while the P-value of professional class 1 is less than 0.05, which indicates that professional class 1, in the practice of flipped classroom teaching, the students' musical ability appeared significant variability.

In addition, in rhythm, music notation, beat, and music reading ability, there is a significant change in the students of Professional 1 class, and their before and after mean values have a large difference, especially in the three modules of music notation, beat, and music reading, the teaching effect is more obvious. The before-and-after mean differences in these three modules in Major 1 class are 4.99, 4.66 and 3.25 points respectively, and their p-values are all 0. It fully indicates that the application of the flipped classroom teaching mode in the folk music classroom can rapidly improve the students' ability in music notation, beat and reading music scores. For such changes, the study concluded that the teaching principles of student-centeredness, focusing on students' individual differences, and building a dynamic teaching framework in the flipped classroom teaching model are integrated with the artistic expression and auditory art of the ethnic music discipline. The teaching role of flipped classroom teaching in the ethnic music classroom is greatly utilized, consolidating students' knowledge of ethnic music and improving their basic ethnic music abilities.

Table 2. Analysis of the results of the music performance data.

-	Class	Before	After	*T*	*P*
Rhythm	1	5.41±0.51	7.04±0.34	5.327	0.000
	2	5.38±0.42	5.42±0.47	0.518	0.423
Musical notation	1	3.16±0.24	8.15±0.19	7.169	0.000
	2	3.24±0.35	3.37±0.28	0.423	0.208
Beat	1	4.09±0.26	8.75±0.31	5.286	0.000
	2	4.14±0.21	4.28±0.37	0.317	0.263
Genealogy reading	1	5.18±0.27	8.43±0.18	4.456	0.000
	2	5.23±0.35	5.57±0.26	0.428	0.197

4.2.2. Analysis of questionnaires

On the basis of clarifying that the ethnic music flipped classroom teaching mode helps to improve students' music performance, this paper also investigates and analyzes students' independent learning ability and music core literacy. After obtaining the data from the questionnaire, the SPSS paired-sample t-test was used to analyze the level of independent learning ability of the students in the Professional 1 class, and to analyze the difference between the pre and post-tests, in order to grasp the effect of the cultivation strategy of enhancing students' independent learning ability in the flipped classroom teaching as a whole. That is, one by one, from the four dimensions of independent learning ability to determine whether the cultivation strategy to enhance students' independent learning ability in flipped classroom teaching has an effect on students' independent learning ability, and the results are shown in Table 3.

In the pre-practice period, the mean value of independent learning ability of the students in the professional 1 class was 2.308 points, and it was 3.424 points after the practice, and the total value of independent learning ability of the students in the class before and after the practice increased by 1.116 points, which is an extremely significant difference with the total value of the level t=10.467, P=0.000<0.05. Therefore, through this practice of enhancing students' independent learning ability in flipped classroom teaching, the independent learning ability of the students in Major 1 class was significantly improved. Among them, the overall means of students' self-directed learning ability in motivation, learning planning, learning environment and reflection on learning and the overall means of students' self-directed learning ability increased significantly after the practice, and there are significant differences (P less than 0.01). This indicates that through this teaching practice of flipped classroom teaching of ethnic music course and implementation of strategies to enhance independent learning ability, students' independent learning ability significantly increased and was significantly promoted in terms of motivation, learning planning, learning environment and learning reflection.

Table 3. The difference in autonomous learning ability before and after tests.

Index	Test	Means	STD	T	Sig.(2-tailed)
Learning motivation	Before	2.059	0.412	8.792	0.002
	After	3.176	0.587		
Learning plan	Before	2.583	0.573	12.519	0.000
	After	3.684	0.415		
Learning environment	Before	2.415	0.423	15.328	0.001
	After	3.481	0.546		
Learning reflection	Before	2.174	0.528	27.186	0.005
	After	3.356	0.435		
Total	Before	2.308	0.382	10.467	0.000
	After	3.424	0.401		

In addition, to investigate the students' music core literacy, the obtained data were carried out the pre and post-test paired samples t-test, and the results of the data analysis of the music core literacy of the students in the professional 1 class were obtained as shown in Table 4. As can be seen from the table, before and after the beginning of the teaching experiment, the mean of the total score of music core literacy of the students in Major 1 class was 2.151 points and 3.762 points, respectively. After the implementation of the flipped classroom teaching mode, the total score of music core literacy of the students in Major 1 class increased by 74.9%, and the paired samples test result t=18.514, P=0.002<0.05. Therefore, it can be shown that the flipped classroom teaching mode can help to improve the core literacy of the students in the folk music classroom, make the students more willing to learn folk music, and lay the foundation for the innovative development of folk music.

Table 4. Data Analysis of Students' Core Musical Literacy.

Index	Test	Means	STD	T	Sig.(2-tailed)
HX1	Before	1.742	0.345	5.792	0.001
	After	3.493	0.363		
HX2	Before	2.058	0.527	14.556	0.002
	After	3.529	0.565		
HX3	Before	2.164	0.401	9.835	0.000
	After	3.737	0.481		
HX4	Before	2.529	0.428	10.261	0.003
	After	4.065	0.494		
HX5	Before	2.261	0.549	24.678	0.000
	After	3.984	0.473		
Total	Before	2.151	0.345	18.514	0.002
	After	3.762	0.363		

5. Conclusion

The purpose of this study is to improve the effect of ethnic music education and promote the diversification of ethnic music teaching resources. To this end, this paper designs an ethnomusicology intelligent generation model based on the MCT-GAN model and constructs an innovative model of ethnomusicology education in combination with the flipped classroom. It is found that the quality of ethnic music obtained by the ethnic music intelligent generation model meets the audience's needs in terms of concordance, structure and fluency, which can further promote the diversified development of ethnic music forms. Under the flipped classroom teaching model, students' music performance, independent learning ability, and music core literacy were significantly improved (p<0.01), which fully demonstrated the feasibility of the application of the flipped classroom in the teaching of ethnomusicology.

This study has certain limitations while achieving research results. When extracting audio features of ethnic music, only MFCC was used to directly extract audio features without preprocessing the ethnic music signal, which may lead to the appearance of noise and affect the accuracy of audio feature extraction. When carrying out teaching experiments, the overall teaching time is relatively short, which cannot fully guarantee the stability of the data results. In the future research, the audio feature extraction process of folk music should be further improved, and the generated music should be introduced into the classroom practice, which guides the students to utilize the digital tools to create folk music, and further promotes the innovative inheritance and development of folk music.

References

1. Ji, W. (2022). research on the influence of the inheritance and development of ethnic music education in coll eges and universities on alleviating college students’psychological anxiety. Psychiatria Danubina, 34(suppl 2), 452-452.
2. Fan, C. (2023). research on ethnic music education in the concept of multiculturalism from the perspective of music anthropology. agricultural sciences, 39, 13.
3. Wang, L. (2025). research on the current status of the application of ethnic music elements in secondary scho ol music education. journal of education, 12(1), 23-28.
4. Wang, W. (2022). Ethnic minority cultures in Chinese schooling: manifestations, implementation pathways a nd teachers’ practices. Race Ethnicity and Education, 25(1), 110-127.
5. Yu, X. (2019). The Construction Strategy of Ethnic Music Teaching System. In 2019 International Conferen ce on Arts, Management, Education and Innovation (ICAMEI 2019).
6. Li, S., Yang, Y., Peng, Y., & Zhan, Y. (2024). Quality Education and the Reform of Ethnic Instrumental Mu sic Teaching in Basic Music Education. Cultura: International Journal of Philosophy of Culture and Axiolog y, 21(1).
7. Li, Y. (2025). Modeling Chinese Traditional Cultural Resonance in College Ethnic Instrumental Music Class room Teaching using Energy driven Bipolar Interval-Valued Neutrosophic Graph. Neutrosophic Sets and Sy stems, 90(1), 69.
8. Huang, Y. (2024). Cultural harmonies: Exploring compositional techniques and cultural fusion in Guizhou et hnic minority music. Pacific International Journal, 7(1), 216-221.
9. Karkina, S. V., Batyrshina, G. I., & Valeeva, R. A. (2020, October). A sustainable approach to music educati on: towards a cultural ecology in the digital age. In Eighth International Conference on Technological Ecosys tems for Enhancing Multiculturality (pp. 535-541).
10. Tian, Y., Chen, H., & Zhou, K. (2024, October). Research on Enhancing the Awareness of the Chinese Natio nal Community through Ethnic Music Based on Big Data Technology. In 2024 8th International Seminar on Education, Management and Social Sciences (ISEMSS 2024) (pp. 477-483). Atlantis Press.

11. Zhang, S., & Wu, C. (2023). Revitalizing endangered traditions: Innovative approaches to safeguarding Yun nan's ethnic minority music as intangible cultural heritage. Herança, 6(1), 101-128.
12. Albert, D. J. (2015). Social media in music education: Extending learning to where students “live”. Music Ed ucators Journal, 102(2), 31-38.
13. Bannerman, J. K., & O’Leary, E. J. (2021). Digital natives unplugged: Challenging assumptions of preservic e music educators’ technological skills. Journal of Music Teacher Education, 30(2), 10-23.
14. Zhang, R., & Wang, H. P. (2024, July). The impact of Chinese university music teachers’ teaching beliefs on creative teaching behaviors: the mediating role of technological acceptance. In Frontiers in Education (Vol. 9, p. 1404541). Frontiers Media SA.
15. Lin, C. (2025). Research on the Implicit Disciplinary Mechanism of Ethnic Identity in Primary and Secondar y School Music Education: From the Perspective of the Core Competency of “Cultural Understanding” Unde r the New Curriculum Standards. Human Resources, Education and Public Policy, 1(1), 21-40.
16. Ouyang, M. (2023). Employing mobile learning in music education. Education and Information Technologie s, 28(5), 5241-5257.
17. Clements, A. (2018). A postdigital future for music education: Definitions, implications, and questions. Acti on, Criticism & Theory for Music Education, 17(1).
18. Parkita, E., & Trzos, A. P. (2016). Digital environment in music school education. International Journal of M usic and Performing Art, 4(2), 53-64.
19. Çetinkaya, P. R., & Kaya, A. (2023). Tendencies of Prospective Music Teachers to Use Technology and the Status of Technology Use in Teaching Practice Course. Cumhuriyet Uluslararası Eğitim Dergisi, 12(4), 1066 -1080.
20. Bayley, J. G., & Waldron, J. (2020). “It’s never too late”: Adult students and music learning in one online an d offline convergent community music school. International Journal of Music Education, 38(1), 36-51.
21. English, H. J., Lumb, M., & Davidson, J. W. (2021). What are the affordances of the digital music space in al ternative education? A reflection on an exploratory music outreach project in rural Australia. International Jo urnal of Music Education, 39(3), 275-288.
22. Wang, Q. (2024). Examining the Role of World Multicultural Music Education in the Inheritance of Foreign Ethnic Music Culture. Cultura: International Journal of Philosophy of Culture and Axiology, 21(4), 69-85.
23. Wagner, C. (2017). Digital gamification in private music education. Antistasis, 7(1).
24. Lee, L., & Liu, Y. Y. (2025). Integrating digital technology systems into multisensory music education: a tec hnological innovation for early childhood learning. Applied System Innovation, 8(5), 125.
25. Fitriana, Y. N., & Putra, R. D. (2022). “Preserve our culture”: The use of digital music platform in the ethnic music community. Bricolage: Jurnal Magister Ilmu Komunikasi, 8(1), 041-050.
26. Shen, H. (2023). The Integration of Digital Technology in the Preservation and Promotion of Ethnic Music i n Colleges and Universities. Journal of Commercial Biotechnology, 28(3).
27. Chang, W. (2025). The integration of artificial intelligence and ethnic music cultural inheritance under deep l earning. Computer Science and Information Systems, (00), 36-36.
28. Hui, F. (2023). Transforming educational approaches by integrating ethnic music and ecosystems through R NN-based extraction. Soft Computing, 27(24), 19143-19158.
29. Miao, P., Faudzi, M. A., Ji, L., Jiang, X., & Wenhong, H. (2024). AA Thematic Analysis of Folk Music Edu cation: Trends, Technology, and Global Perspectives (2013–2023). Música Hodie, 24.
30. Wang, T. (2024). Research on the Application of Intangible Cultural Heritage Music Culture in Higher Educ ation Music Teaching. Journal of Art, Culture and Philosophical Studies, 1(2).
31. Huan, L., Jirajarupat, P., & Liu, L. (2024). Innovative Research on Ethnic Music Performance Models in Sou thwestern Chinese Universities under Digital Technology. Journal of Roi Kaensarn Academi, 9(12), 2076-20 88.
32. Yihan, L., Cuong, T. V., Kiss, B., Oo, T. Z., Szabó, N., & Józsa, K. (2025). The Use and Effectiveness of Di gital Tools in Elementary Music Education: A Systematic Review. Music & Science, 8, 2059204325136333 8.
33. Geni Wu. (2025). Research on the Student-centered Music Flipped Classroom Model. Region - Educational Research and Reviews,7(2),https://doi.org/10.32629/RERR.V7I2.3490.
34. J. Jayanthi & V. Upendran. (2025). Raga Recognition of Indian Classical Music using Meerkat Optimization Based MFCC and Fine Tuned BILSTM-XGBOOST. Circuits, Systems, and Signal Processing,44(7),1-29.htt ps://doi.org/10.1007/S00034-025-02999-W.
35. Yang Zhang & Shu Yu. (2025). Harmonizing AI: A GAN–Transformer fusion for expressive multimodal mu sic synthesis in IoT systems. Alexandria Engineering Journal,131,368-382.https://doi.org/10.1016/J.AEJ.202 5.07.043.