Computer Science and Information Systems 22(3):1121–1138 https://doi.org/10.2298/CSIS241119036C

The Integration of Artificial Intelligence and Ethnic Music Cultural Inheritance under Deep Learning

Wenbo Chang

Performing Arts and Culture, The Catholic University of Korea, 43 Jibong-ro, Wonmi-gu, Bucheon-si, Gyeonggi-do 14662, South Korea
changwenbo2023@163.com

Abstract. The traditional music education system faces numerous challenges in inheriting ethnic music culture. Especially in the modern educational environment, the protection and dissemination of ethnic music encounter many difficulties. This work aims to utilize advanced technologies such as deep learning (DL) to explore methods for optimizing the inheritance of ethnic music culture. By summarizing the current situation of ethnic music cultural inheritance, and analyzing its background and content, this work proposes innovative solutions that integrate artificial intelligence (AI) technology. Leveraging a newly constructed model, it performs multi-level and comprehensive analyses of ethnic music elements, uncovering the internal emotional expression mechanism of ethnic music. The experimental results of timbre emotion recognition are presented and compared. The findings reveal that the unsupervised training method improves the feature accuracy by 1.96% compared to the Mel-Frequency Cepstral Coefficients (MFCC), while the supervised training method achieves a 3.46% improvement. In addition, the timbre recognition rate is compared between the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM) and the Deep Neural Network-Hidden Markov Model (DNN-HMM). The result shows that the DNN-HMM is better. These findings highlight the significant advantages of applying DL methods in preserving and transmitting ethnic music cultural inheritance. This work can effectively enhance the accuracy of music emotion recognition, thus providing new technical support for the protection and inheritance of ethnic music.

Keywords: Deep Learning, Artificial Intelligence, Ethnic Music, Cultural Inheritance, Performance Evaluation, GMM-HMM, DNN-HMM.

1. Introduction

1.1. Research Background and Motivations

With the rapid progress of artificial intelligence (AI) technology, particularly significant breakthroughs in Deep Learning (DL), machines are now better equipped to understand and process complex, unstructured data, such as music audio signals. Thanks to their outstanding feature learning and pattern recognition abilities, DL models offer new possibilities for the exploration, analysis, and preservation of ethnic music culture data [1]. Traditional music education systems face several challenges in transmitting ethnic music culture, including uneven distribution of educational resources, difficulties in implementing personalized teaching, and a shortage of qualified teaching staff [2]. Hence, optimizing music teaching methods using advanced technologies like DL, especially in the context of ethnic music inheritance, becomes crucial.

In the era of globalization, local and minority ethnic music cultures are at risk of marginalization or even loss. AI technologies, especially DL, offer the potential to digitize, classify, intelligently retrieve, and reproduce ethnic music resources, thereby providing robust support for protecting and disseminating ethnic music cultural inheritance [3-5]. The education sector universally seeks innovation and keeps pace with the times, aiming to apply advanced scientific technologies to educational practices to enhance teaching quality and efficiency. The fusion of DL and ethnic music cultural inheritance research reflects this trend, aiming to modernize ethnic music education processes through intelligent means [6].

The motivation behind this work stems from two main aspects. One is to recognize the powerful capabilities of DL in handling complex music information, uncovering underlying patterns, and implementing personalized teaching. Thus, new opportunities can be provided to optimize music education methods and enhance teaching effectiveness. The other is to deeply care about the issues of protecting and preserving ethnic music culture both in China and globally. Faced with the limitations of traditional educational approaches in promoting and popularizing ethnic music culture, such as unequal distribution of resources and monotonous teaching methods, AI technology, especially DL, is expected to offer innovative solutions. Therefore, this work explores how advanced AI technologies like DL can be applied to the education of ethnic music inheritance. By leveraging intelligent analysis, personalized recommendations, and interactive learning methods, this work seeks to enhance students' understanding and appreciation of ethnic music culture and promote the modernization of music aesthetic education. At the same time, it effectively fosters the inheritance and development of ethnic music cultures.

1.2. Research Objectives

First, this work explores how DL algorithms can be applied to accurately extract features from ethnic music and recognize patterns, including but not limited to melody, rhythm, harmony, and emotional expression across multiple dimensions. The goal is to achieve intelligent management and utilization of ethnic music cultural resources. Second, personalized music aesthetic education strategies are designed and implemented by integrating AI technology. These strategies can cater to students of different ages and backgrounds, providing learning content and pathways tailored to their characteristics. This approach aims to enhance the effectiveness of ethnic music education and foster greater student engagement. Third, by deeply integrating DL with ethnic music education, the work seeks to overcome geographical and temporal limitations, expanding the dissemination scope of ethnic music culture. This approach is intended to protect and inherit China's rich and diverse ethnic music inheritance, contribute to the modernization of music education, and provide new technological support for international exchanges in ethnic music culture.

2. Literature Review

In the rapidly evolving information technology era, the development of AI and DL has brought unprecedented innovative opportunities for cultural inheritance [7,8]. Particularly within the realm of ethnic music cultural inheritance, the challenge of effectively combining cutting-edge technology with traditional culture to achieve mutual enhancement has become a widely discussed focus in the academic community [9]. This section provides an in-depth analysis of existing literature, highlighting the importance and urgency of integrating advanced technology with ethnic music cultural inheritance. It also delineates the achievements and challenges encountered by scholars in this field, offering a solid theoretical foundation and practical guidance for future research and practice.

2.1. Digital Preservation of Ethnic Music

In exploring the methods and technologies for the digital collection, organization, and storage of ethnic music resources, this work integrates audio signal processing technology with DL algorithms for feature extraction and classification. This provides a solid foundation for music education, academic research, and cultural dissemination. In addition, the technical and practical challenges faced during the digital preservation process, such as sound quality restoration, and long-term data storage stability, are discussed. Huang et al. were dedicated to digitizing, organizing, and storing traditional ethnic music resources, and constructing an ethnic music database. They employed audio signal processing techniques and DL algorithms for feature extraction and classification of ethnic music, laying the groundwork for subsequent music education, academic research, and cultural dissemination [10]. Lei et al. (2024) argued that digital technologies, particularly digital audio and video technologies, could accurately record and restore traditional music's sound characteristics and cultural background. Thus, the limitations of traditional recording and manual preservation methods could be overcome. Furthermore, digital preservation enabled the online sharing of ethnomusicological resources, breaking down geographical and cultural barriers, and allowing future generations to access and understand these traditional arts [11]. Zhang et al. (2023) pointed out that digital preservation was not only a technical issue but also involved respecting and protecting indigenous music's cultural context and intellectual property. They suggested that collaboration with indigenous communities should be considered during the digitization process, ensuring that the preservation methods aligned with local cultural and social practices, to avoid cultural appropriation or misuse [12]. Zhao (2024) highlighted that traditional ethnic music often featured complex melodies and rhythms that were difficult to accurately record using standardized audio formats. Therefore, customized recording equipment and analytical tools were necessary to capture the details of these notes and rhythms [13].

2.2. AI-Driven Personalized Ethnic Music Education

The design and implementation of an AI-based ethnic music teaching platform provide students with a customized learning experience. The system tailors content to students' learning habits, ability levels, and interest preferences, delivering relevant ethnic music materials. Simultaneously, it enhances students' understanding and appreciation of ethnic music through real-time feedback and interaction. Zhang and Romainoor developed such a platform, enabling intelligent recommendation, personalized teaching, and online learning of ethnic music. The system tailored suitable ethnic music content based on students' learning habits, proficiency levels, and interests, thus fostering a deeper understanding and appreciation of ethnic music through dynamic feedback and engagement [14]. Jing et al. (2024) emphasized that AI could personalize teaching content and facilitate the dynamic dissemination of ethnic music, attracting more young people to engage with and appreciate traditional culture [15]. Li et al. (2024) stressed that AI-driven emotional teaching could more comprehensively achieve the educational goals of ethnic music. This enabled students to master performance techniques and to deeply understand the cultural essence behind ethnic music [16]. Qin et al. (2022) highlighted that this personalized teaching approach could significantly enhance students' learning efficiency and engagement, while also aiding in the protection and dissemination of diverse ethnic music cultures [17].

2.3. DL and the Emotional Inheritance of Ethnic Music

This work explores using DL models to analyze and identify the emotional information embedded in ethnic music. It reveals the emotional expressions and aesthetic values in different ethnic music, enhancing students' perception and understanding of the deeper connotations of ethnic music through emotional education. In addition, the work discusses the importance of emotional inheritance in transmissing ethnic music culture and its role in fostering cultural identity. Reshma et al. utilized DL models to analyze and recognize emotional information in ethnic music, exploring the inherent emotional expressions and aesthetic values across different music genres. Their approach aimed to enhance students' perception and understanding of the profound connotations of ethnic music, thus facilitating emotional inheritance in ethnic music culture [18]. Catalina et al. (2023) highlighted the advantages of the Gaussian Mixture Model-Hidden Markov Model (GMM-HMM), particularly in the speech recognition field. GMM could describe the probabilistic distribution of speech signals, while HMM was employed to capture the temporal characteristics of speech signals [19]. Hai et al. (2022) mentioned that HMM provided a natural framework for modeling state transitions in sequential data. In contrast, GMM helped improve the observation probability model of HMM, enabling the system to better fit continuous-valued observation data, particularly in the processing of audio signals or other time-series data [20].

2.4. Comparison between DL and Traditional Methods

The application of traditional ethnic music inheritance methods is compared with that of modern DL technology in the inheritance process. For endangered ethnic music heritage, Ning investigated the use of AI technology for recording and restoring such music, developing scientifically reasonable conservation strategies. How to pass on these music inheritances to the next generation through modern technological means was studied, ensuring the enduring vitality of ethnic music culture [21]. Zhou et al. (2024) argued that DL technologies could spread ethnic music globally by constructing large-scale ethnic music databases and recommendation systems. Through music generation techniques based on Variational Autoencoders (VAE), new variations of ethnic music could be created and pushed to a wider audience via online platforms [22]. Wang (2024) mentioned that modern DL technologies, such as emotion recognition and generation models, had the potential in conveying the emotional aspects of ethnic music. Emotion analysis tools based on DL could extract emotional features from audio data and generate emotionally consistent music segments through models. Although technological methods could not fully replace the emotional depth of live teaching, they could record and reproduce emotional details on a large scale, offering new possibilities for the preservation of ethnic music culture [23]. Yuan (2023) believed that DL technologies could record and analyze the audio features of ethnic music (such as melody, rhythm, and timbre) on a large scale, thus preserving traditional music as a digital resource. By using Generative Adversarial Networks and Convolutional Neural Networks (CNNs), music samples close to real performances could be synthesized, helping more learners access this music. Meanwhile, it could overcome the dependence of traditional inheritance on teachers and scenes [24].

In summary, the integration of AI with ethnic music cultural inheritance from the perspective of DL demonstrates broad application prospects and profound social value. Scholars have tackled numerous challenges in traditional music education by exploring the digital construction of ethnic music, developing intelligent music education systems, emotional recognition and education, heritage protection and inheritance strategies, and global dissemination. These efforts have paved the way for innovative approaches to preserving and promoting ethnic music culture. However, further in-depth research and continuous exploration are required to address emerging challenges in integrating cutting-edge technology and cultural heritage. These challenges may include enhancing DL models' understanding and expression of ethnic music's uniqueness and ensuring that technological means do not weaken cultural essence during the inheritance process. This work continues to evolve and optimize AI technology to propose more mature and comprehensive solutions, effectively advancing the preservation and inheritance of ethnic music culture to new heights.

3. Research Methodology

3.1. The Use of AI in Inheriting Ethnic Music Culture

Chinese excellent traditional culture boasts a long history and profound richness. It is the unshirkable responsibility of cultural workers in the new era to inherit and innovate this cultural heritage, ensuring its enduring vitality and preserving its essential role in society [25-27]. Applying AI to this endeavor can open up vast space for the inheritance and innovation of Chinese culture. The specific functions are reflected in four aspects. First, leveraging computer vision technology can enhance learning efficiency and the acquisition of traditional techniques, thereby contributing to the widespread inheritance of traditional culture. Second, integrating augmented reality (AR) technology with cultural scenic spots can promote cultural tourism and facilitate the digital development of traditional culture. Third, employing DL technology can foster secondary creation, enrich cultural forms, and endow the innovation of traditional culture with more personalized features. Fourth, big data technology can achieve precise promotion, gain insights into user needs, and ensure the accurate inheritance of traditional culture. Figure 1 illustrates these roles in inheriting and innovating traditional culture [28].

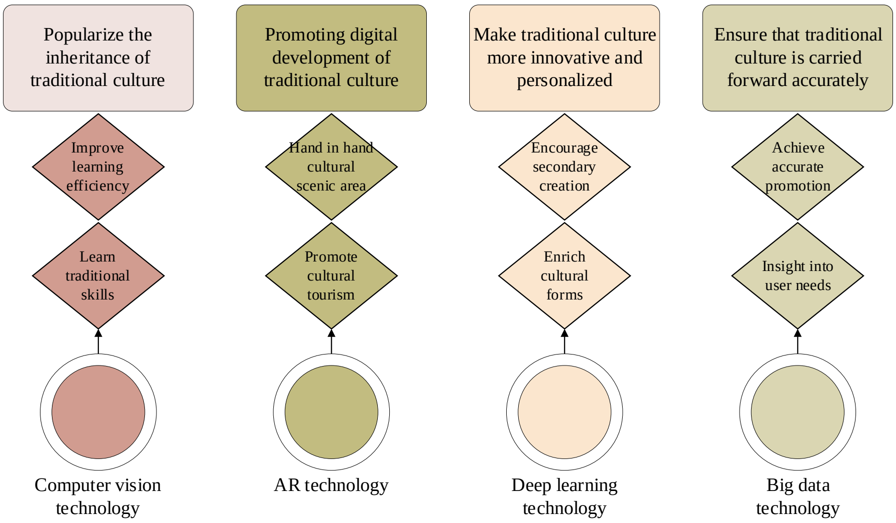

Fig.1. The Role of AI in Inheriting and Innovating Traditional Culture

Based on this, AI technology also plays a crucial role in inheriting ethnic music culture. It utilizes advanced techniques such as DL, big data analysis, and natural language processing (NLP) to digitize and intelligently analyze ethnic music [29]. On the one hand, AI facilitates the efficient collection, organization, classification, and storage of various ethnic music resources, establishing rich databases for easy retrieval and dissemination. On the other hand, intelligent algorithms based on DL can accurately identify and extract features such as melody, rhythm, and mode of ethnic music. These can help people deeply understand and appreciate the unique music styles and emotional expressions of different ethnicities. Additionally, AI can be applied in intelligent music education to design personalized learning paths and provide engaging interactive teaching experiences that engage students more effectively. This fosters greater interest in and engagement with ethnic music among younger generations, ensuring the continued inheritance and promotion of the musical treasures of various ethnic cultures. Moreover, with the assistance of AI technology, ethnic music culture can transcend geographical restrictions, disseminating in a more modernized and internationalized manner to achieve global sharing and exchange [30-32]. Table 1 demonstrates its specific manifestations [33].

Table 1. Elements of AI Used in Inheriting Ethnic Music Culture

Number	Element	Implementation Means
1	Digitization Collection	Utilizing professional equipment and technical means to preserve ethnic music culture in digital form
2	Data Analysis and Processing	Employing big data analysis techniques to organize, annotate, and store collected audio data
3	Feature Extraction and Recognition	Using DL algorithms to extract features from music data and identify unique elements and styles of ethnic music
4	Intelligent Education Platform	Constructing an online music education platform, providing ethnic music-related courses and educational resources
5	Personalized Recommendation	Offering customized learning content based on users' learning records and preferences through intelligent recommendation algorithms
6	Dissemination	Promoting ethnic music culture globally through various channels such as the internet, social media, and mobile applications, facilitating cultural exchange and inheritance

3.2. The Use of DL in Acoustic Models

DL is a machine learning (ML) technology whose fundamental concept is rooted in research on artificial neural networks, especially the design of multi-layer nonlinear network structures. DL models consist of multiple interconnected layers, where each layer progressively extracts more abstract and complex feature representations from input data [34,35]. During the training process, DL algorithms adjust the network's weight parameters via backpropagation, enabling the model to automatically learn from raw input data and extract useful features. This process allows DL to perform complex tasks such as image recognition, speech recognition, NLP, and computer vision [36]. Compared to traditional ML methods, DL excels particularly in solving problems with high-dimensional, unstructured data. With the support of large labeled datasets and substantial computational resources, DL can achieve performance beyond the human level, making revolutionary advancements in various AI fields. Figure 2 compares the structure of acoustic models based on the deep neural network (DNN) [37]. The right side of Figure 2 shows the structure of DNNs for an acoustic model. A DNN consists of an input layer, multiple hidden layers, and an output layer. The input layer receives acoustic features, such as Mel-frequency cepstral coefficients (MFCCs). Neurons in the hidden layers are connected by weights and introduce non-linearity through activation functions, enabling the network to learn complex patterns. The output of each hidden layer serves as the input for the subsequent layer, forming a feed-forward network structure. Model states (e.g., h1, h2, h3, h4, h5) represent the feature representations at different levels in the network. The depth of the DNN is determined by the number of hidden layers, which affects its learning ability and training difficulty. This structure is widely used in acoustic models for tasks such as speech recognition, speech synthesis, and other audio processing applications, as it is highly effective at capturing the intricate features and patterns of input data.

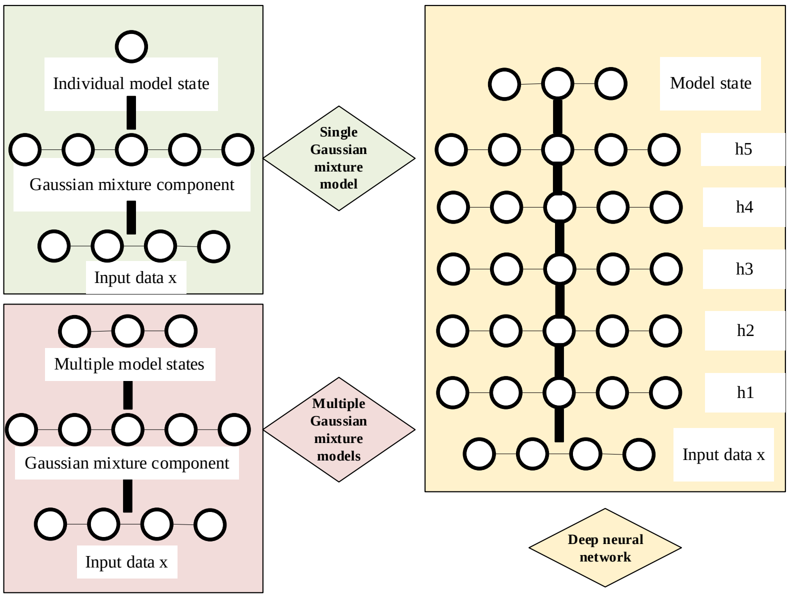

Fig. 2. Model Structure Comparison

In comparison, the structure of the GMM is similar to that of a DNN but features only a single hidden layer. Each node in this layer represents a Gaussian mixture component of the model. The output layer and nodes constitute the model's state vector, derived from the hidden layer nodes, thereby obtaining the posterior probability of the feature state vector from the input layer [38].

Next, the basic process of acoustic modeling is introduced. It is assumed that the feature observation vector at time t is y_{ut} and the activation probability of the output layer state vector is P. Then, the state probability of the DNN-based acoustic model can be computed using the Softmax function as follows [39]:

$$y_{ut}(s) = P(s|O_{vt}) = \frac{exp\{a_{ut}(s)\}}{\sum_{s'} exp\{a_{ut}(s')\}} \quad (1)$$

$a_{ut}(s)$ represents the activation probability of the output layer node state s, that is, the effective output value, and its expression reads:

$$\log p(O_{vt}|s) = \log y_{ut}(s) - \log P(s) \quad (2)$$

The standard mean squared error backpropagation algorithm trains the DNN with the specified optimization objective function. Cross-entropy is chosen as the objective control function for the DNN system, while the optimization algorithm adopts the stochastic gradient descent algorithm [40-42]. Given that acoustic speech recognition is a multi-state classification application, the DNN selects the logarithmic function as the objective function, which can be written as:

$$F_{CE} = -\sum_{u=1}^{U} \sum_{t=1}^{T_u} \log y_{ut}(s_{ut}) \quad (3)$$

s_{ut} refers to the state of the system at the time t; F_{CE} represents the cross-entropy between the reference state vector and the predicted state vector. The gradient calculation equation between the system's objective function and the output layer's node state vector [43] is expressed as:

$$\frac{\partial F_{CE}}{\partial a_{ut}(s)} = -\frac{\partial \log y_{ut}(s_{ut})}{\partial a_{ut}(s)} = y_{ut}(s) - \partial s, s_{ut} \quad (4)$$

$\partial s, s_{ut}$ represents the Kronecker delta function.

3.3. Music Emotion Classification Based on Transfer Learning from a DL Perspective

Ethnic music culture embodies the essence of music art accumulated over the history of various ethnic groups, carrying profound ethnic emotions and cultural memories. Music emotion classification involves analyzing elements such as melody, rhythm, harmony, and timbre to identify and understand the emotional states and emotional connotations conveyed by the music. The inheritance of ethnic music culture plays a critical role in the research and application of music emotion classification. Each ethnic music genre has developed a diverse range of emotional categories, with its unique expressions of emotions and aesthetic concepts [44,45]. Employing modern technologies like DL to classify and interpret the emotional content of ethnic music can help people better understand the emotional characteristics of different ethnic music genres. Moreover, it can effectively preserve and inherit ethnic music culture, allowing listeners to experience and resonate with it at a deeper level, thereby promoting the continuation and development of ethnic music culture. Furthermore, accurate emotion classification can provide valuable insights for music education, assisting teachers and students in better grasping and imparting the intrinsic emotional appeal of ethnic music [46].

Using the trained CNN model for music emotion as a feature extractor, a new target-domain classifier is established to classify the music feature vectors. Figure 3 illustrates the basic process of model transfer [47]. It demonstrates the basic process of transfer learning. Here, Ds represents the source-domain classifier, which is used to pre-train the source-domain classifier. Then, the learned knowledge is applied to the target-domain classifier Dt through model transfer. Finally, feature extraction and fine-tuning are performed on the target-domain data to improve the performance of the target-domain classifier.

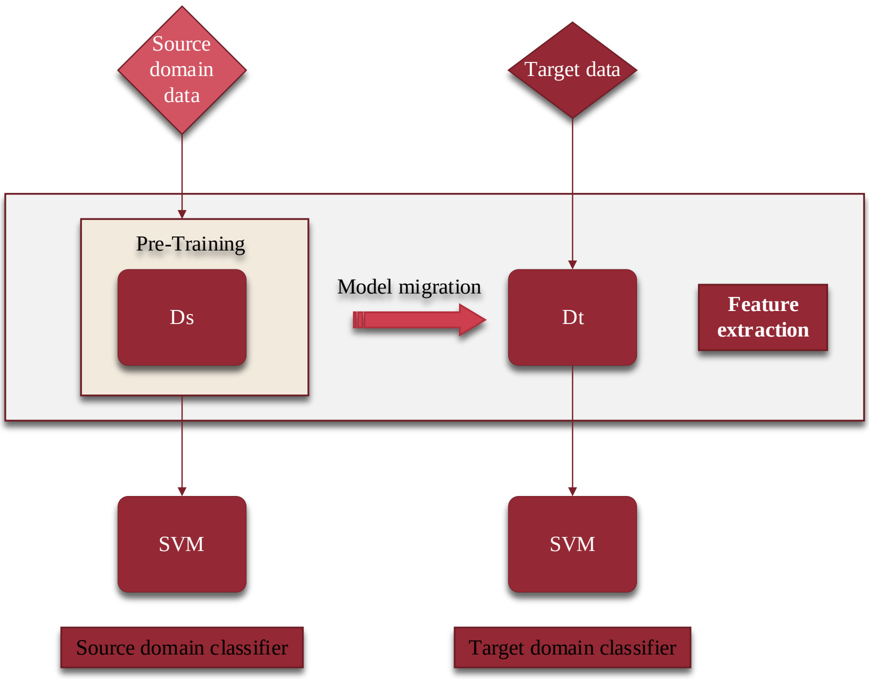

Fig. 3. Model Transfer Process

From a neural network perspective, similar spectral information, such as tones and rhythms, is mainly reflected in the structure, with this structural information concentrated in the layers preceding the convolutional layer. If the low-level information is directly transferred to the high-level, its expressive capacity may be diminished. Therefore, this project intends to merge the middle-level features with the high-level information, thus enriching low-level features with higher-level data. It aims to enhance the fusion ability of features and tasks, reduce the dependency of simple CNNs on high-level information, and improve the network's generalization ability [48,49].

Through music emotion classification, people can analyze the emotional characteristics such as joy, sadness, calmness, and excitement, embedded in different ethnic music works. These characteristics are often closely related to their unique cultural and historical backgrounds. By leveraging AI technologies like DL, emotional features can be extracted from various dimensions of ethnic music such as melody, rhythm, harmony, and timbre. The music can then be classified according to its emotional content [50]. This process helps to reveal and preserve the deep cultural and emotional meanings embedded in ethnic music. Moreover, it provides precise teaching resources and scientific teaching methods for music education. Enabling learners to understand and perceive the emotions of different ethnic music is essential for them to inherit and promote ethnic music culture. Furthermore, it can cultivate students' aesthetic literacy and emotional engagement with ethnic music, thus facilitating the inheritance of ethnic music culture [51].

4. Experimental Design and Performance Evaluation

4.1. Datasets Collection

The dataset used for the music emotion classification experiments is based on the MIREX-like dataset, consisting of 709 songs. Each song is 30 seconds in duration and categorized into five fairly balanced groups. Table 2 presents the emotion labels for each category.

Table 2. Dataset Categories

Category	Number of music (pieces)	Emotion Labels
Category 1	127	Affectionate, Pleasant, Amusing, Joyful, Sweet
Category 2	127	Intense, Confident, Passionate, Lively, Noisy
Category 3	152	Restless, Humorous, Foolish, Eccentric, Witty, Distorted
Category 4	152	Sad, Thoughtful, Troubled, Graceful, Bitter, Reluctant
Category 5	151	Strong-willed, Intense, Tense, Anxious, Vulgar, Capricious

4.2. Experimental Environment

In this experiment, the open-source DL framework TensorFlow is used. Training for the support vector machine (SVM) is conducted using the SVM package from Scikit-learn. The MIREX-like music emotion dataset is utilized, and the Librosa software package is used for audio processing. The audio signals are sampled at 22050Hz, with a frame size of 1024 samples per second and a frameshift of 512 samples per second for Fourier transformation. This results in the extraction of the time-domain spectrogram of the audio signal. Logarithmic amplitude spectrograms are then obtained through logarithmic operations. Using a frequency of every 3 seconds, the processed spectrograms are fed into a mobile-based model as128×128 input data. Accuracy remains the metric for all experimental comparisons [52].

4.3. Parameters Setting

The established transfer model includes initializing model parameters. In the experiment, a very small random number (0.001* randn ()) is used to allocate connectivity weights, and the bias of each node is set to 0. For each training set, the learning rate is set to 0.001, and termination terminates after 30 repetitions. After training, the final hidden layer (h1) and the state vectors of each node are retained as the input vectors for the risk-based monitoring (RBM) model. The termination condition for the experiment's learning process is either a maximum of 500 iterations or a change in the mean square error of 0.001 per iteration [53].

Table 3 presents the parameter settings for the experimental validation platform:

Table 3. Parameter Settings

Model Type	Parameters
DL Model	Model Architecture
	Hidden Layer (h1)
	Activation Function
	Weight Initialization Method
	Bias Initialization
	Learning Rate
	Number of Training Epochs
SVM	Kernel Function
	Penalty Coefficient (C)
	Kernel Function Parameters
	Maximum Number of Iterations
	Tolerance (tol)

4.4. Performance Evaluation

The performance of the three methods is first compared based on word precision, word accuracy, and sentence accuracy. Among them, word precision and word accuracy refer to the system's performance in word recognition. The word precision is the ratio of correctly recognized words to the total number of words. The word accuracy means the ratio of the number of correctly recognized words, excluding insertion errors. Sentence accuracy is similarly calculated by the ratio of correctly recognized words, but it excludes any additional inserted words from the total count. Figure 4 presents the experimental results.

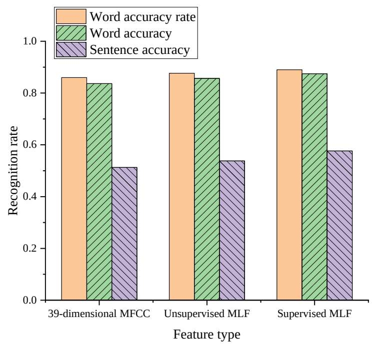

Fig. 4. Comparison of Recognition Rates

The experimental simulation results show that, compared with the traditional MFCC feature extraction method, the speech feature extraction method based on a deep auto-encoder mainly transcribes speech data and outputs the results or the model. It means the Master Label File (MLF), which can effectively improve speech recognition performance. Considering word recognition accuracy, the unsupervised training method’s a feature accuracy reaches 1.96%, exceeding that of MFCC, while the supervised training method performs 3.46% higher than MFCC. However, for sentence accuracy, the performance of all three methods is relatively suboptimal, mainly due to the lack of a complete acoustic model.

The parameters of the DNN in the experiment are as follows. The input and output layer vectors are 143-dimensional, while the hidden layer consists of 1024×5 dimensions. The input layer adopts 11 frames of MFCC superframe features. Numerical experiments are conducted on the sentence error rate (SER), word error rate (WER), and model training time to evaluate the basic performance of the model. Table 4 displays the experimental results, comparing the performance of the GMM-HMM and the DNN-HMM in the task of timbre recognition.

Table 4. Comparison of Timbre Recognition Rates

Model	GMM-HMM	DNN-HMM
SER	30.4%	23.1%
WER	5.0%	3.6%
Training Time	12h	47h

The experiments indicate that AI-based DL algorithms outperform GMM algorithms, as GMM has a lower hierarchy and may not adequately simulate the brain's basic requirements for external environments. Considering the practical application context, there is a demand for intelligent speech recognition technology that enables the brain to process external sounds. With the assistance of DNNs, multi-level nonlinear mappings can abstract and simplify complex speech data, thereby obtaining features that meet practical requirements.

4.5. Discussion

Madzík et al. focused on leveraging DL technology to extract and analyze the tonal, rhythmic, and structural features of ethnic music. They innovatively applied this technology to the specific practice of ethnic music cultural inheritance. They investigated how DL models could accurately capture and interpret the emotions and cultural connotations embedded in ethnic music. Also, they designed and implemented a comprehensive DL-driven system for ethnic music inheritance aimed at music education. However, their research did not directly address the practical application of these findings in music education and cultural inheritance [54]. Bai et al. explored the transformative impact of AI on music education, particularly in personalized teaching and resource recommendation. However, they did not specifically examine ethnic music cultural inheritance, nor did they delve deeply into the role of DL in this context [55]. Unlike previous research that focused solely on technical feature extraction or was limited to general music education applications, this work tightly integrates DL technology with ethnic music cultural inheritance. It explores specific strategies and methodologies to facilitate personalized teaching, emotional education, and cultural inheritance on intelligent music education platforms. Additionally, this work emphasizes the use of AI technology to promote the global popularity and recognition of ethnic music while simultaneously protecting and disseminating the diversity of ethnic music cultures. The goal is to foster a deep integration of technology and humanities in the realms of music education and cultural inheritance.

5. Conclusion

5.1. Research Contribution

This work, through the multi-level and comprehensive analysis of ethnic music elements, integrates DL into timbre and emotion recognition. This effectively uncovers the internal mechanism of emotional expression in ethnic music while deepening the understanding of its uniqueness and diversity. Empirical research has verified the effectiveness and feasibility of the proposed AI and DL model in the educational inheritance of ethnic music culture. The experimental results show that, compared with traditional teaching methods, DL-driven intelligent music education significantly improves students' cognitive understanding and appreciation of ethnic music. Consequently, it effectively promotes the inheritance and development of ethnic music culture.

The main contribution of this work lies in systematically introducing DL technologies into the field of ethnic music culture inheritance, demonstrating how DL can be utilized for precise feature recognition, emotion analysis, and personalized teaching. By establishing a multi-level and three-dimensional analytical model, this work deeply explores subtle elements in ethnic music such as timbre, rhythm, and emotional expression. Hence, it provides a practical and scalable framework for the transmission and education of ethnic music culture. This framework outperforms traditional feature extraction methods (e.g., MFCC) in terms of recognition accuracy, offering empirical evidence for the effectiveness of DL in recognizing and classifying the intrinsic emotional and cultural characteristics of ethnic music. Furthermore, the work finds that AI-based personalized teaching remarkably increases learners' engagement and appreciation of different musical traditions. Meanwhile, it overcomes limitations in resource allocation and teacher capabilities, thereby greatly promoting the popularization and education of ethnic music culture. The method proposed here also demonstrates significant improvements in emotion and timbre recognition accuracy compared to traditional methods and supports personalized teaching, enhancing learners' engagement and learning outcomes in ethnic music. Additionally, this method holds broad practical application potential, particularly in smart music education platforms, intangible cultural inheritance protection projects, and other related fields. Concurrently, it fosters a deeper integration of technology and the humanities, providing innovative ideas for ethnic music education and cultural inheritance. Although the method shows good performance, its model generalization ability still needs validation due to limitations in dataset size and diversity. The high training costs may also restrict its practical deployment. In addition, the model's ability to capture the complex emotions of ethnic music is insufficient, and the lack of multimodal integration could affect its comprehensive understanding of music culture. These issues need further optimization and resolution in subsequent research.

Regarding practical applications, the method proposed in this work has significant potential, especially in smart music education platforms, online learning systems, and ethnic music culture preservation projects. The findings demonstrate the feasibility of AI technology in music education while promoting broader discussions about the digital protection and dissemination of intangible cultural inheritance. By integrating technologies such as virtual reality (VR) and AR with DL, future research could further enrich the learning experience of ethnic music culture. Moreover, its interactivity and immersiveness could be enhanced, thus opening up broader space for the inheritance and development of ethnic music.

5.2. Future Works and Research Limitations

With the ongoing advancements in AI and DL technologies, the integration research of AI and ethnic music cultural inheritance from a DL perspective holds significant promise. In the future, DL models can be further optimized and improved to enhance their ability to accurately understand and explore the complex and subtle emotional expressions and cultural connotations in ethnic music. Through the successful application of DL in timbre and emotion recognition, this work has laid the foundation for constructing an immersive, AI-driven ethnic music teaching platform. In addition, more application scenarios can be explored. For example, by combining VR and AR technologies with DL, rich and interactive learning experiences in ethnic music culture can be created. This can promote the digitalization, intellectualization, and personalization of ethnic music education.

Although this work has made valuable contributions to integrating DL and ethnic music cultural inheritance, certain limitations remain. First, the current DL models may have insufficient recognition accuracy when dealing with complex and highly localized features of ethnic music, necessitating optimization of model structure and training strategies. Second, limited by the dataset size and diversity, the model's generalization ability still needs to be verified through more comprehensive ethnic music data. Subsequent research should focus on expanding the dataset, particularly by collaborating with cultural institutions to collect under-represented music samples. Additionally, it should explore combining VR or AR technologies to provide a richer and more interactive learning experience.

References

1. Huang, L., Song, Y.: Intangible Cultural Heritage Management Using Machine Learning Model: A Case Study of Northwest Folk Song Huaer. Scientific Programming, 2(17). (2022)

2. Yao, M., Liu, J.: The Analysis of Chinese and Japanese Traditional Opera Tunes with Artificial Intelligence Technology Based on Deep Learning. IEEE Access, 11(1), 16-21. (2024)
3. Zhang, N.: Informatization Integration Strategy of Modern Vocal Music Teaching and Traditional Music Culture in Colleges and Universities in the Era of Artificial Intelligence. Applied Mathematics and Nonlinear Sciences, 5(11), 5-9. (2023)
4. Tang, Z.: Application Model Construction of Emotional Expression and Propagation Path of Deep Learning in National Vocal Music. International Journal of Advanced Computer Science & Applications, 14(11). (2023)
5. Zhou, W.: The Development System of Local Music Teaching Materials Based on Deep Learning. Optik, 273, 170421. (2023)
6. Li, P., Liang, T., Cao, Y., et al.: A Novel "An" Drum Music Generation Method Based on Bi-LSTM Deep Reinforcement Learning. Applied Intelligence, 54(1), 80-94. (2024)
7. Hui, F.: Transforming Educational Approaches by Integrating Ethnic Music and Ecosystems through RNN-Based Extraction. Soft Computing, 27(24), 19143-19158. (2023)
8. Başarır, L., Çiçek, S., Koç, M.: Local Intelligence: Time to Learn from AI. Architectural Science Review, 1-16. (2024)
9. Zhang, A.: Optimization Simulation of Match between Technical Actions and Music of National Dance Based on Deep Learning. Mobile Information Systems, 2023. (2023)
10. Huang, R., Holzapfel, A., Sturm, B., et al.: Beyond Diverse Datasets: Responsible MIR, Interdisciplinarity, and the Fractured Worlds of Music. Transactions of the International Society for Music Information Retrieval. (2023)
11. Lei, X.: Analysing the Effectiveness of Online Digital Audio Software and Offline Audio Studios in Fostering Chinese Folk Music Composition Skills in Music Education. Journal of Computer Assisted Learning, 40(5), 2339-2350. (2024)
12. Zhang, X.: Digital Communication of Folk Music in Social Music Culture. Frontiers in Art Research, 5(14), 322-356. (2023)
13. Zhao, X., Xu, D.: Research on Public Digital Cultural System of Huizhou Folk Music from the Perspective of Cultural Identity. Academic Journal of Humanities & Social Sciences, 4(10), 1028-1031. (2021)
14. Zhang, B., Romainoor, N. H.: Research on Artificial Intelligence in New Year Prints: The Application of the Generated Pop Art Style Images on Cultural and Creative Products. Applied Sciences, 13(2), 1082. (2023)
15. Jing, S., Lei, L.: Construction and Implementation of Content-Based National Music Retrieval Model under Deep Learning. International Journal of Information System Modeling and Design (IJISMD), 15(1), 1-17. (2024)
16. Li, P., Liang, M. T., Cao, M. Y., et al.: A Novel Xi'an Drum Music Generation Method Based on Bi-LSTM Deep Reinforcement Learning. Applied Intelligence, 54(1), 80-94. (2024)
17. Qin, L. N., Junyan, S.: Artificial Neural Network for Folk Music Style Classification. Mobile Information Systems, 2022(12), 1035-1038. (2022)
18. Reshma, M. R., Kannan, B., Raj, V. P. J., et al.: Cultural Heritage Preservation through Dance Digitization: A Review. Digital Applications in Archaeology and Cultural Heritage, 28, e00257. (2023)
19. Catalina, M., Marcelo, M., Aarón, C., et al.: An End-to-End DNN-HMM Based System with Duration Modeling for Robust Earthquake Detection. Computers and Geosciences, 179. (2023)
20. Hai, J. Y., Guofa, L., Jialong, H., et al.: Health Condition Evaluation Method for Motorized Spindle on the Basis of Optimised VMD and GMM-HMM. The International Journal of Advanced Manufacturing Technology, 124(11-12), 4465-4477. (2022)

21. Ning, H., Chen, Z.: Fusion of the Word2vec Word Embedding Model and Cluster Analysis for the Communication of Music Intangible Cultural Heritage. Scientific Reports, 13(1), 22717. (2023)
22. Zhou, Y., Huang, F.: Navigating Knowledge Dynamics: Algorithmic Music Recombination, Deep Learning, Blockchain, Economic Knowledge, and Copyright Challenges. Journal of the Knowledge Economy, 1-25. (2024)
23. Wang, J.: Music Personalization Imputation Method Based on Deep Transfer Learning. Applied Mathematics and Nonlinear Sciences, 9(1), 3215-3218. (2024)
24. Yuan, Y. Y., Samaneh, S.: Differentiated Analysis for Music Traffic in Software Defined Networks: A Method of Deep Learning. Computers and Electrical Engineering, 107. (2023)
25. Deng, J.: A Brief Analysis of the Path of Intangible Cultural Heritage Inheritance and Innovative Development under Digital Technology. Journal of Innovation and Development, 3(3), 29-32. (2023)
26. Xie, Y.: Consumer Preference Measurement of Folk Culture Based on Confidence Rule Base Model. Journal of Electrical Systems, 19(4), 211-226. (2023)
27. Deng, M., Liu, Y., Chen, L.: AI-Driven Innovation in Ethnic Clothing Design: An Intersection of Machine Learning and Cultural Heritage. Electronic Research Archive, 31(9), 5793-5814. (2023)
28. Wang, T., Chen, J., Liu, L., et al.: A Review: How Deep Learning Technology Impacts the Evaluation of Traditional Village Landscapes. Buildings, 13(2), 525. (2023)
29. Wang, T., Ma, Z., Yang, L.: Creativity and Sustainable Design of Wickerwork Handicraft Patterns Based on Artificial Intelligence. Sustainability, 15(2), 1574. (2023)
30. He, F.: Research on the Application of Artificial Intelligence Aesthetics in the Cultivation of Aesthetic Literacy of Art-Normal Students. International Journal of Social Sciences and Public Administration, 2(1), 160-168. (2024)
31. Foka, A., Eklund, L., Løvlie, A. S., et al.: Critically Assessing AI/ML for Cultural Heritage: Potentials and Challenges. Handbook of Critical Studies of Artificial Intelligence, 815-825. (2023)
32. Breit, A., Waltersdorfer, L., Ekaputra, F. J., et al.: Combining Machine Learning and Semantic Web: A Systematic Mapping Study. ACM Computing Surveys, 55(14s), 1-41. (2023)
33. Li, P., Wang, B.: Artificial Intelligence in Music Education. International Journal of Human–Computer Interaction, 1-10. (2023)
34. García-Madurga, M. Á., Grilló-Méndez, A. J.: Artificial Intelligence in the Tourism Industry: An Overview of Reviews. Administrative Sciences, 13(8), 172. (2023)
35. Rane, N., Choudhary, S., Rane, J.: Sustainable Tourism Development Using Leading-Edge Artificial Intelligence (AI), Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) Technologies. Blockchain, Internet of Things (IoT), Augmented Reality (AR) and Virtual Reality (VR) Technologies (October 31, 2023), 2023.
36. Rane, N.: Role and Challenges of ChatGPT and Similar Generative Artificial Intelligence in Arts and Humanities. Available at SSRN 4603208, 2023.
37. Ning, H.: Analysis of the Value of Folk Music Intangible Cultural Heritage on the Regulation of Mental Health. Frontiers in Psychiatry, 14, 1067753. (2023)
38. Zhou, E., Shen, Q., Hou, Y.: Integrating Artificial Intelligence into the Modernization of Traditional Chinese Medicine Industry: A Review. Frontiers in Pharmacology, 15, 1181183. (2024)
39. Sarkar, C., Das, B., Rawat, V. S., et al.: Artificial Intelligence and Machine Learning Technology Driven Modern Drug Discovery and Development. International Journal of Molecular Sciences, 24(3), 2026. (2023)
40. Ahmed, N. M. F., Elwakeel, L. M., Marzuk, A. A., et al.: Using Artificial Intelligence Applications to Create Contemporary Heritage Designs to Enrich Typography to Preserve Saudi Heritage. Kurdish Studies, 11(3), 39-57. (2023)

41. Zheng, X.: Integration of Multiple Features in Chinese Landscape Painting and Architectural Environment Using Deep Learning Model. International Journal of Intelligent Systems and Applications in Engineering, 12(6s), 593-606. (2024)
42. Chen, Y., Wang, L., Liu, X., et al.: Artificial Intelligence-Empowered Art Education: A Cycle-Consistency Network-Based Model for Creating the Fusion Works of Tibetan Painting Styles. Sustainability, 15(8), 6692. (2023)
43. AYAZ, A., IMAMOGLU, M. B., BOZTAS, G. D., et al.: Artificial Intelligence and Immersive Reality: A Systematic Literature Review on Research in the Social Sciences. Current Studies in Technology, Innovation and Entrepreneurship, 50. (2023)
44. García-Esparza, J. A., Pardo, J., Altaba, P., et al.: Validity of Machine Learning in Assessing Large Texts through Sustainability Indicators. Social Indicators Research, 166(2), 323-337. (2023)
45. Zhu, M., Wang, G., Li, C., et al.: Artificial Intelligence Classification Model for Modern Chinese Poetry in Education. Sustainability, 15(6), 5265. (2023)
46. Xu, W.: Research on the Communication Opportunities of Intangible Cultural Heritage under the Background of Big Data and AI. Journal of Artificial Intelligence Practice, 6(8), 12-17. (2023)
47. Dong, L.: Using Deep Learning and Genetic Algorithms for Melody Generation and Optimization in Music. Soft Computing, 27(22), 17419-17433. (2023)
48. Liu, J.: An Automatic Classification Method for Multiple Music Genres by Integrating Emotions and Intelligent Algorithms. Applied Artificial Intelligence, 37(1), 2211458. (2023)
49. Spennemann, D. H. R.: Will Artificial Intelligence Affect How Cultural Heritage Will Be Managed in the Future? Responses Generated by Four GenAI Models. Heritage, 7(3), 1453-1471. (2024)
50. Tan, S. N., Ng, K. H.: Gamified Mobile Sensing Storytelling Application for Enhancing Remote Cultural Experience and Engagement. International Journal of Human–Computer Interaction, 40(6), 1383-1396. (2024)
51. Adjeisah, M., Asamoah, K. O., Yeboah, M. A., et al.: Adinkra Symbol Recognition Using Classical Machine Learning and Deep Learning. arXiv preprint arXiv:2311.15728, 2023.
52. Zou, H., Ge, J., Liu, R., et al.: Feature Recognition of Regional Architecture Forms Based on Machine Learning: A Case Study of Architecture Heritage in Hubei Province, China. Sustainability, 15(4), 3504. (2023)
53. Singh, A., Kanaujia, A., Singh, V. K., et al.: Artificial Intelligence for Sustainable Development Goals: Bibliometric Patterns and Concept Evolution Trajectories. Sustainable Development, 32(1), 724-754. (2024)
54. Madzík, P., Falát, L., Copuš, L., et al.: Digital Transformation in Tourism: Bibliometric Literature Review Based on Machine Learning Approach. European Journal of Innovation Management, 26(7), 177-205. (2023)
55. Bai, N., Ducci, M., Mirzikashvili, R., et al.: Mapping Urban Heritage Images with Social Media Data and Artificial Intelligence, a Case Study in Testaccio, Rome. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 48, 139-146. (2023)

Wenbo Chang was born in Tai Yuan,Shan xi.P.R. China, in 1996. He received his master's degree from Belarusian State Academy of Music,Belarus. Now, he studies in The Catholic University of Korea. His research interest include culture and art sociology, ethnomusicology. changwenbo2023@163.com

Received: November 19, 2024; Accepted: Januar 15, 2025.