Research on the Inheritance and Innovation of Traditional Ethnic Vocal Art in the Context of Multiculturalism based on Neural Network Algorithm

JUN MA
School of Music,
Dalian Art College,
Dalian 116600,
CHINA

Abstract: - During the present period of cultural integration worldwide, there's a heredity and innovation conundrum for traditional ethnic vocal arts. The current evaluation system is obviously too subjective and lacks a clear road. There is no clear direction for development. This paper uses an improved BP neural network and LSTM (Long-Short-Term-Memory Network) algorithm to construct a whole evaluation model. To accurately understand the effectiveness and innovativeness of traditional ethnic vocal art inheritance, and to transform the qualitative indicator values into feature vector sets through quantization and to input them into the neural network system so as to achieve an accurate judgment of the status quo of its development and a forecast of the future trends. Collecting multiple ethnic group data and constructing a model database based on traditional multiple school vocal databases. This is compared and analyzed with single-layer BP neural network and classic indicators. From the experimental test results, it can be seen that compared with the single architecture, the proposed integrated architecture has the advantages of faster convergence and higher prediction accuracy. And it also greatly reduces the data bias caused by human factors, which provides scientific basis for the spread and development of ethnic vocal cultural. It also expands the application area of technology for the digital protection of intangible cultural heritage.

Key-Words: - Neural network algorithm, Multiculturalism, Traditional ethnic vocal music, Inheritance and innovation, Evaluation model.

Received: November 19, 2025. Revised: January 15, 2026. Accepted: February 17, 2026. Published: June 2, 2026.

1 Introduction

The important carrier of excellent traditional Chinese culture, the traditional ethnic vocal Art, has gathered many ethnic cultural accumulations, aesthetic tastes, and spiritual connotations, showcasing different characteristics such as Han folk tunes, Mongolian long tunes, Tibetan nangma, Uyghur muqam, etc. Under the trend of globalization getting deeper and culture exchange growing bigger between Eastern and Western countries, the impact of pop music and modern consumerism culture by Western countries is more significant. Resulted in a gradual decline of its audience, few young successors, even the disappearance of certain skills. Compatibility with current style is very low. There is no clear planning for creative activities. There is a lack of competitiveness in the current cultural industry pattern, [1].

Inheritance and innovation studies on traditional national vocal music at present are mostly studied from the perspective of chemical culture and art, [2]. The research is mainly qualitative, and there are shortcomings like: First, in the inheritance efficiency evaluation, experts rely on subjective judgment for the evaluation, with hazy evaluation criteria making it hard to gauge the real results generated by different inheritance pathways; Second, in determining innovative direction, there's no data to back up decisions, leading to "too Western" or "too traditional" extremes; Lastly, controlling the amount of multicultural elements is tough, without good quantitative analytical means.

Artificial intelligence tech advances fast, thus it gives traditional Chinese art studies modern technical means. Neural networks are algorithms that have strong non-linear mapping and adaptive learning abilities, they have also been used initially in fields like art evaluation, pattern recognition, and cultural heritage protection. As one of the currently most commonly used neural network models, the BP neural network is able to complete complex functions mapping between inputs and outputs, but the problems such as relatively slow training speed and being easily trapped into local minima; LSTM networks have unique strengths in handling temporal data, due to their ability to capture long-term dependencies and being suitable for analyzing the inheritance and evolution laws of traditional vocal art, [3], [4].

This article combines the core requirements of the inheritance and innovation of traditional ethnic vocal music in the context of multiculturalism and establishes an improved BP-LSTM fusion neural network model. Inheritance efficiency and innovation potential all enter as a vector which is made of evaluation indicators. By model training and validation, objective analysis is conducted on the current situation of the inheritance of traditional ethnic vocal arts, and scientific prediction for future innovation direction is made. The purpose of this study is to overcome the existing limitations of qualitative research such as being time-consuming, expensive, and low efficiency, and then provide analysis of quantity and technical support for the protection, inheritance and improvement of traditional ethnic vocal music, and integrate traditional art and technology

1.1 Core Characteristics of Traditional Ethnic Vocal Music in the Context of Multiculturalism

Under the backdrop of cross-cultural communication, the inheritance and development of traditional ethnic vocal music involves factors like culture, art, acceptance, and so on, [5]. This system mainly consists of three levels: Firstly, it is essential to attach great importance to protecting the cultural characteristics and continuity of each ethnic group, ensuring that the cultural heritage of each ethnic group can be fully preserved; Secondly, on the basis of integration and modernization, various forms of expression should be enriched and diversified to make them more suitable for the modern aesthetic demands resulting from the combination of multiple elements, [6]. A systematic evaluation system should be formed with respect to a work, from many aspects to take into consideration about the artistic level, meaning and dissemination effect.

1.2 Principles and Improvements of BP Neural Network

Basic BP Neural Network: BP neural network contains three parts of layers: input layer, hidden layer, and output layer, [7], [8]. Network weights and thresholds are modified via error backpropagation algorithm to approach the objective function, [9]. Learning process has two stages: forward propagation and backward propagation. Forward propagation, the input vector is passed through the hidden layer activation function, and the hidden layer is sent to the output layer to calculate the error of the output value with the expected output; Backward propagation, the error returns to where it came from, and the weights and thresholds at each layer are updated with gradient decent to minimize the error function, [10].

Taking into account the efficiency of gradient propagation and the requirement for probability mapping at the output layer in model training, the Sigmoid function (as shown in Equation (1) is adopted as the activation function for the output layer of the BP neural network. Its output value ranges between [0,1], allowing for direct mapping to the probability distribution of evaluation scores. The ReLU activation function (as shown in Equation (2) is employed for the hidden layer. This function effectively mitigates the vanishing gradient problem and enhances the training efficiency of deep networks. The combination of the two functions balances output rationality and training stability.

$$f(x)=\frac{1}{1+e^{-x}} \tag{1}$$

$$ReLU(x) = max(0, x) \tag{2}$$

The error function adopts mean square error (MSE), expressed as Equation (3):

$$E=\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{3}$$

Among which, y_i refers to achieve the desired output, $\hat{y}_i$ refers to the actual output of the network, n refers to the sample size.

Improved BP Neural Network: In order to overcome the two defects of slower training speed and easy fall into local minima of traditional BP neural networks, this paper makes the following improvements:

Introduce momentum factor α, The formula for adjusting weight updates is shown as Equation (4):

$$\Delta w_{ij}(k+1)=\eta\delta_j a_i+\alpha\Delta w_{ij}(k) \tag{4}$$

Among which, η: learning rate, δ_j: To output layer error signals, a_i: To output the hidden layer, $\alpha \in [0\middle, 1]$. choice $\alpha = 0.9$ Based on the following theoretical and empirical evidence: when$\alpha = 0.9$ When the momentum term contributes significantly to weight updates, it can effectively smooth the gradient descent path and reduce oscillations. Mathematical derivation shows that the effective step size for gradient update under this value is $\eta/(1-\alpha)$, compared to $\alpha < 0.9$ It can cross the local minimum area faster; Through controlled variable experiments, it has been verified that when$\alpha = 0.9$ The ratio of error values of the model at the same number of iterations $\alpha = 0.7$ and $\alpha = 0.8$, reduced by 18.3% and 9.7% respectively, with the most significant improvement in convergence stability.

Adopting an adaptive learning rate adjustment strategy, the mathematical expression is shown in Equation (5):

$$\eta(k+1) = \begin{cases} \eta(k) \times \gamma_{inc}, & E(k+1) < E(k) \times \theta \\ \eta(k) \times \gamma_{dec}, & E(k+1) > E(k) \\ \eta(k), & \text{other conditions} \end{cases} \quad (5)$$

Among which, $\gamma_{\text{inc}} = 1.05$ (Learning rate improvement coefficient), $\gamma_{\text{dec}} = 0.95$ (Learning rate reduction coefficient), $\theta = 0.98$ (Error improvement threshold). When the error decreases by more than 2% after iteration, increase the learning rate appropriately to accelerate convergence; When the error increases, reduce the learning rate to avoid oscillation; If the error changes within a reasonable range, keep the learning rate unchanged. The mathematical essence of this strategy is to dynamically adjust the step size through error feedback, so that the gradient descent increases the step size in the gentle area of the error surface and decreases the step size in the steep area, thereby balancing convergence speed and stability.

1.3 LSTM Network Principle

LSTM Network is an improved type of Recurrent Neural Network (RNN); LSTMs deal with the issue of vanishing or exploding gradients for long sequence data by using a forget gate, an input gate, and an output gate, [11], [12]. Its basic structure comprises cellular state and 3 gating units:

Forgotten gate: determines which information to discard from the cellular state, expressed as Equation (6):

$$f_t=\sigma\left(W_f \cdot \left[\mathrm{h}_{t-1}, x_t\right]+b_f\right) \quad (6)$$

Input gate: determines which new information is stored in the cell state, expressed as Equations (7)-(8):

$$i_t=\sigma\left(W_i \cdot \left[\mathrm{h}_{t-1}, x_t\right]+b_i\right) \quad (7)$$

$$\widetilde{C}_t=\tanh\left(W_C \cdot \left[\mathrm{h}_{t-1}, x_t\right]+b_C\right) \quad (8)$$

Cell state update: Combining the outputs of the forget gate and input gate to update the cell state, as shown in Equation (9):

$$C_t=f_t \odot C_{t-1}+i_t \odot \widetilde{C}_t \quad (9)$$

Output gate: determines which parts of the cell state are output to the hidden state, expressed as Equations (10)-(11):

$$o_t=\sigma\left(W_o \cdot \left[\mathrm{h}_{t-1}, x_t\right]+b_o\right) \quad (10)$$

$$\mathrm{h}_t=o_t \odot \tanh(C_t) \quad (11)$$

Among which, σ refers to activate the Sigmoid function, tanh refers to the hyperbolic tangent activation function, $\odot$ refers to element wise multiplication, W refers to the weight matrix, b refers to the bias vector, h_t refers to hide the state, C_t refers to cellular state.

1.4 Construction of 1.4BP-LSTM Fusion Model

1.4.1 Model Training Mechanism

The BP-LSTM fusion model constructed in this article adopts an end-to-end joint training method. During the training process, the weights of the BP network and LSTM network are synchronously updated to ensure the collaborative optimization of the two components. The gradient backpropagation path is as follows:

(1) Gradient of LSTM network parameters with output layer loss

$\nabla_{\mathrm{W_o,b_o,W_f,b_f,W_i,b_i,W_C,b_C}} \mathrm{L} = \nabla_{\mathrm{h_T}} \mathrm{L} \cdot \frac{\partial \mathrm{h_T}}{\partial \mathrm{C_T}} \cdot \frac{\partial \mathrm{C_T}}{\partial \cdot}$, among which T To maximize the time step, propagate the gradient in reverse along the time step using the chain rule.

(2) LSTM output gradient of BP network parameters

$\nabla_{\mathrm{W_{BP},b_{BP}}} \mathrm{L} = \nabla_{\mathrm{s}} \mathrm{L} \cdot \frac{\partial \mathrm{s}}{\partial \mathrm{a_{BP}}} \cdot \frac{\partial \mathrm{a_{BP}}}{\partial \mathrm{W_{BP},b_{BP}}}$, among which, s: The static features output by the BP network, $\mathrm{a_{BP}}$: Output for the hidden layer of BP network.

The loss function of the fusion model introduces L2 regularization term, as shown in Equation (12):

$$\mathrm{L} = \mathrm{MSE} + \lambda \cdot \left(\sum_{\mathrm{W} \in \Theta_{\mathrm{BP}}} |\mathrm{W}|^2 + \sum_{\mathrm{W} \in \Theta_{\mathrm{LSTM}}} |\mathrm{W}|^2\right) \quad (12)$$

Among which, $\lambda = 0.001$. For the regularization coefficient, Θ_{BP} and Θ_{LSTM} The sets of weight parameters for BP network and LSTM network respectively. The regularization term effectively suppresses model overfitting by penalizing excessive weight values. At the same time, dropout layers are added to the hidden layer of the BP network and the LSTM network, respectively, with a dropout probability of 0.2, expressed mathematically as $a_{drop} = a \cdot m$, among which, m: To follow the Bernoulli distribution of mask vectors, $P(m_i = 1) = 0.8$, Further enhance the model's generalization ability by randomly deactivating some neurons.

The mathematical definition of the early stop criterion is as follows: Set the validation set error windowwindow $= 10$, The decrease in validation set error during 10 consecutive iterations $\Delta E = (E_k - E_{k+10})/E_k < 10^{-4}$. Stop training and use the current iteration's model parameters as the optimal solution to avoid overfitting caused by overtraining.

1.4.2 Feature Fusion and Dimension Conversion

BP network outputs 1D static featuress(Range of values [0,1]), with 3D dynamic temporal characteristics $d_1(t), d_2(t), d_3(t)$ Perform dimension concatenation to form a 4-dimensional input for the LSTM network $x_t = [s, d_1(t), d_2(t), d_3(t)]^T$ .The mathematical definitions and calculation methods of the three dynamic temporal features are as follows:

Inheritance and evolution rate $d_1(t)$:

$d_1(t) = \frac{1}{k}\sum_{i=1}^{k} \frac{f(t)-f(t-i)}{f(t-i)}$, among which f(t):

For the th t The comprehensive score of inheritance indicators during the period, $k = 5$ The size of the time window reflects the rate of change in the inheritance status over the past five periods.

Innovation volatility coefficient $d_2(t)$:

$d_2(t) = \sqrt{\frac{1}{k-1}\sum_{i=1}^{k}(c(t-i) - \bar{c})^2}$, among which,c(t): For the tht The innovation index score of the period,$\bar{c}$ Measure the stability of innovation level by averaging the score within the window.

Spread index $d_3(t)$:

$d_3(t) = \frac{N(t)-N(t-1)}{N_{max}}$, among which, N(t):For the tht The audience coverage scale during the period, N_{max} To determine the maximum audience size in the sample set and reflect the growth trend of the dissemination scope after standardization.

The mathematical description of the feature fusion mechanism is: $z_t = \beta \cdot s + (1 - \beta) \cdot \frac{1}{T}\sum_{t=1}^{T} h_t$, among which, $\beta = 0.3$: To integrate weights, h_t : For the hidden states of each time step in the LSTM network, z_t To achieve the final fusion of features, the effective integration of multi-level information is achieved through the weighted sum of static and dynamic features.

The advantage of this fusion model is that the BP neural network can handle the nonlinear mapping relationship between static features well, while the LSTM network can capture the long-term dependency relationship of dynamic temporal features. The combination of the two achieves multi-level and all-round evaluation of the inheritance and development of traditional ethnic vocal music.

2 Construction of Evaluation System for Inheritance and Innovation of Traditional Ethnic Vocal Music

2.1 Design Principles for Evaluation Indicators

In accordance with the main content of inheriting and improving traditional ethnic singing based on multiculturalism, and the design is carried out, the evaluation indicators should follow the principle of

(1) Systematic principle. Covering four dimensions of inheritance foundation, inheritance efficiency, innovation potential, and dissemination effect, it comprehensively reflects the current development status of traditional ethnic vocal music, [13].

(2) Principle of scientificity. The selection of indicators is based on relevant theoretical research and practical experience, ensuring clear and quantifiable connotations of the indicators.

(3) Principle of operability. Indicator data is easy to collect and avoids selecting abstract indicators that are difficult to quantify, [14], [15].

(4) Principle of dynamism. Considering the dynamic development of multiculturalism, the indicator system has a certain degree of flexibility and adaptability.

2.2 Composition of Evaluation Indicator System

The evaluation index system constructed in this article includes 4 primary indicators, 12 secondary indicators, and 36 tertiary indicators, as shown in Table 1 (Appendix).

2.3 Quantification of Indicators and Data Preprocessing

2.3.1 Qualitative Indicators Quantification

For qualitative indicators, the 5-point rating scale (1-5) is adopted, with 10 experts in vocal music and

5 cultural research scholars invited to rate. After removing the highest and lowest scores, the average is taken as the indicator score.

2.3.2 Standardization of Quantitative Indicators

For quantity-based indicators like the number of inheritors and the size of the audience, because there are large variances in the indicator units and values, the MinMax standardization approach is utilized to transform the data into the [0,1] interval, which is formulated as Equation (13):

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{13}$$

Among which, x refers to raw data, $x_{\min}$ refers to the minimum value of the indicator, $x_{\max}$ refers to the maximum value of the indicator, x' refers to standardized data.

2.3.3 Determination of Indicator Weights

To apply Analytic Hierarchy Process (AHP) to determine the weights of different evaluative indicators, first establish a judgment matrix, and then ask experts in the relevant field to pairwisely compare and give scores to all indicators, to finally conduct consistency test procedures on it, after which the standardization yields a standardized weight distribution plan to guarantee its scientific basis.

2.4 Sample Data Collection

To make sure we got good training performance and that the model was good enough for the generalization task, we gathered materials from traditional ethnic vocal music of 5 major ethnic music groups in China plus 8 well-known vocal schools. It is also the same as including information records of inheritors and singing videos, audience record of reactions, recording of media comments, etc. During this process a total of 300 high quality samples were collected, 60 independent test samples were selected which form classic work and innovative developmental achievement at different time stage and have different type of distribution and wide applications.

3 Experimental Design and Result Analysis

3.1 Experimental Environment and Parameter Settings

The hardware environment of this experiment adopts Intel Core i7-10700K processor, 32GB memory, and NVIDIA GeForce RTX 3080 graphics card; The software environment adopts Python 3.8 and calls libraries such as TensorFlow 2.5, Keras 2.4.3, Matplotlib 3.4.3, sklearn 0.24.2 for development.

The basis for setting and selecting model parameters is as follows:

3.1.1 Improving BP Neural Network

The neural network model constructed in this study adopts a three-layer structure design, with 36 neurons corresponding to 36 three-level evaluation indicators in the input layer. Optimization testing was conducted using grid search method within the range of 32 to 128 neurons, and the final number of hidden layer neurons was determined to be 64. At this point, the validation set error reached its minimum value. The output layer adopts a single neuron structure for result output. The model training parameters are set as follows: the initial learning rate is determined to be 0.01 through sensitivity analysis, which refers to the standard empirical value of neural network training, and the momentum factor is set to 0.9. The training process is set to 500 iterations and 0.001 is set as the error threshold to control the training accuracy.

3.1.2 LSTM Network

The neural network model designed in this study adopts a 4-dimensional input feature space, which includes 1D static features and 3D dynamic temporal features. By using cross validation method to optimize parameters within the range of 16 to 64 neurons, the final number of hidden layer neurons was determined to be 32. The time step of the model is set to 10, which fully considers the time span characteristics of the inheritance data and can fully cover one inheritance cycle. In terms of training parameter settings, the batch size is set to 32, which ensures memory efficiency while balancing the stability and convergence speed of the training process. After multiple rounds of testing verification, the model exhibits optimal convergence characteristics and generalization performance when the batch processing volume is 32. The entire training process is set to 300 iterations to ensure that the model learns fully.

3.1.3 Fusion Model

Adam optimizer (learning rate 0.001), loss function is mean square error with L2 regularization, and the hidden layer uses ReLU activation function; The output layer uses the Sigmoid activation function.

Taking batch size and time step as an example, when the batch size increases from 16 to 32, the model training time is shortened by 27% and the validation set R^2 is improved by 0.03. When the training time increases from 32 to 64, the training time increases by 42%, and the R^2 only increases by 0.01, indicating a decreasing marginal benefit. When the time step increases from 5 to 10, R^2 increases by 0.04. When increasing from 10 to 15, there is no significant change in R^2, and the training complexity increases. Therefore, the optimal parameter combination is determined.

3.2 Experimental Plan Design

This experiment focuses on the current situation and development of the inheritance and innovation of traditional ethnic vocal music, and designs a three part research plan with hierarchical promotion and logical closed loop, balancing scientificity and practicality:

The first step is to conduct model performance comparison experiments. Train improved BP neural network, LSTM network, and BP-LSTM fusion model using the training set, evaluate performance using the test set, and perform statistical significance analysis using t-test (significance level α=0.05). Simultaneously conducting ablation experiments, constructing control models with "no momentum factor", "no adaptive learning rate", and "no regularization" to quantify the contributions of each improved component.

The second step is to conduct an evaluation experiment on the effectiveness of inheritance. Select the BP-LSTM fusion model to quantitatively evaluate the inheritance effectiveness of test samples, calculate Pearson correlation coefficient (r) and Spearman correlation coefficient (p), and verify the consistency between the model and expert evaluation through correlation analysis; Use single sample t-test to verify the statistical significance of "error less than 2%".

Step three, conduct innovation direction prediction experiments. Based on the multidimensional index scores output by the fusion model, a weighted comprehensive evaluation model is constructed, and the mathematical expression is shown in Equation (14):

$$I = \sum_{i=1}^{m} w_i \cdot s_i \tag{14}$$

Among which, I : For the comprehensive index of innovation potential, $\mathrm{w_i}$: For the weights of various innovation related indicators (determined based on the AHP method), $\mathrm{s_i}$: Score for the indicator. by calculationI And the deviation values of each sub indicator $\Delta \mathrm{s_i} = 1 - \mathrm{s_i}$, Prioritize selecting indicators with high weights and large bias values as innovation directions, forming a scientific prediction mechanism.

3.3 Experimental Results and Analysis

3.3.1 Model Performance Comparison

The performance comparison results of the three models are shown in Table 2 (Appendix). Through t-test analysis, it was found that there is a significant difference in MAE between the BP-LSTM fusion model and the improved BP neural network($\mathrm{t} = 4.23,\ \mathrm{p} < 0.01$), Significant difference in MAE compared to LSTM network($\mathrm{t} = 3.87, \mathrm{p} < 0.01$), Prove that the performance advantage of the fusion model has statistical significance.

The results of the ablation experiment are shown in Table 3. The momentum factor reduces MAE by 19.0%, the adaptive learning rate reduces MAE by 14.3%, and regularization and dropout reduce MAE by 11.9%. The synergistic effect of the three achieves optimal performance, demonstrating the effectiveness of each improved component.

Table 3. Results of ablation experiment

Model variant	MAE	RMSE	R^2
Fusion model (complete)	0.025	0.032	0.964
No momentum factor	0.031	0.039	0.942
No adaptive learning rate	0.029	0.037	0.948
No regularization and dropout	0.028	0.036	0.951

Source: created by the authors

Convergence rate analysis: The convergence curve of the fusion model satisfies $\mathrm{E(k)} = \mathrm{E_0} \cdot \mathrm{e}^{-\lambda \mathrm{k}}$, $\lambda = 0.012$. To improve the convergence rate coefficient, the BP neural network is improved $\lambda = 0.008$, LSTMnetwork $\lambda = 0.009$, The convergence rate of the fusion model has been improved by 33.3%-50%, demonstrating its advantage in convergence efficiency. The visualization results of the loss function surface show that the fusion model can effectively get rid of local minima and ultimately converge to the global optimal region, while traditional BP neural networks are prone to falling into local minima and the error cannot be further reduced.

3.3.2 Results of Inheritance Efficiency Evaluation

The BP-LSTM fusion model was used to quantitatively predict the inheritance effectiveness of 30 test samples, and compared with the subjective scores of 15 experts. The comparison data of some samples are shown in Table 4 (Appendix). Correlation analysis result: Pearson correlation coefficient $r = 0.972(p < 0.001)$, spearman correlation coefficient $\rho = 0.968(p < 0.001)$, This indicates a strong linear and rank correlation between model predictions and expert evaluations.

Single sample t-test for relative error, null hypothesis $H_0: \mu \geq 2\%$, alternative hypothesis $H_1: \mu < 2\%$. Inspection results: $t = -5.32, p < 0.001$, Reject the null hypothesis and prove that 'error less than 2% has statistical significance, and the model's prediction accuracy is reliable.

3.3.3 Prediction Results of Innovation Direction

The analysis results based on the weighted comprehensive evaluation model show that traditional vocal music of different ethnic groups presents differentiated innovative potential characteristics. The Han Shaanbei folk songs show weak integration of multicultural elements, with a score of 0.62 and a deviation value of 0.38. However, their overall priority is relatively high, with a weight of 0.07; In terms of the combination of traditional and modern techniques, it shows great potential for innovation, with a score of 0.81. It is recommended to enhance contemporary adaptability by integrating modern popular music arrangement methods. The Mongolian tribal chief's tone performs outstandingly in the dimension of conveying ethnic cultural connotations, with a score of 0.93, but the innovation degree of communication forms is only 0.58. It is recommended to use new media methods such as short videos to expand the audience. Uyghur Muqam has a score of 0.65 in terms of technique integration, but performs well in the dimension of multicultural integration, with a score of 0.85, making it suitable for incorporating Central Asian music elements. The innovation level of Zhuang folk song themes is the lowest, only 0.52, but the potential for optimizing audience age distribution is 0.78. It is recommended to combine modern social hot topics to attract young people. Through a 6-month innovative practice verification, the improved work achieved an average improvement of 15.3% in audience satisfaction and dissemination scope, fully demonstrating the practical guidance value of the model.

3.4 Experimental Conclusion

This study conducted a systematic analysis of traditional ethnic vocal music based on the BP-LSTM fusion model. The main conclusions are as follows: Firstly, the fusion model exhibits significant advantages in inheritance and innovation evaluation. Compared with single improved BP neural network or LSTM network, its prediction accuracy is higher and stability is stronger. Statistical significance testing and ablation experiments have verified that the improvement effect is scientific and reliable, and can avoid subjective factors interference, providing reliable objective data support for evaluation work; Secondly, the evaluation results of the model are highly consistent with the subjective evaluation of experts, with correlation coefficients exceeding 0.96, proving that the model can identify and grasp the core characteristics of the inheritance and innovation of traditional ethnic vocal music, and achieve quantitative evaluation of the effectiveness of inheritance; Finally, the innovative direction prediction based on the weighted comprehensive evaluation model is reasonable and feasible. It can not only effectively guide the innovative practice of traditional ethnic vocal music, but also promote the integration of traditional art and modern culture, promote the innovative development of ethnic music, and provide important methods and support for the digital dissemination and innovative inheritance of traditional art.

4 Conclusion

This study mainly focuses on the main bottleneck of the inheritance and development of traditional ethnic folk music in a multicultural context. To build a multidimensional evaluation system for gene inheritance based on inheritance foundation, inheritance effect, innovative inheritance, and dissemination effect, and to propose an improved BP-LSTM fusion neural network model. It is realized with a two-stage process of extracting the static features and modeling the dynamic temporal features, to make an accurate quantitative judgment on the present development status of the traditional ethnic vocal music and a scientific prediction about its innovative direction. Experimental result proves that fusion model's prediction effect is better than single BP neural network, LSTM network, and it is more stable. It is evaluated to be very similar to the subjective opinion of experts. The targeted innovation strategy is based on the result of the model, and it has been proved successful in practice. Not only can it break through the limitation of traditional qualitative research, but also provide an objective basis and technical support for the digital protection and inheritance path optimization as well as innovative practice of traditional ethnic vocal music. And it also provides a reference research idea on the deep integration of intangible cultural heritage art with modern technology in different multicultural backgrounds.

References

[1] Liu H, Guo W. Effectiveness of AI-driven vocal art tools in enhancing student performance and creativity. *European Journal of Education*, Vol.60, No.1, 2025, pp.e70037-e70037.

[2] Shen D. Modeling emotional patterns and singing styles of ancient poetry in vocal art songs based on text mining. *Applied Mathematics and Nonlinear Sciences*, Vol.10, No.1, 2025, pp.1-8.

[3] Bin F. The application and inheritance of Chinese ancient poetry and art songs in vocal teaching. *Art and Performance Letters*, Vol.5, No.4, 2024, pp.1-6. doi: 10.23977/ARTPL.2024.050401.

[4] Chen Y. The effective use of traditional music elements in Chinese vocal music art. *Research and Commentary on Humanities and Arts*, Vol.2, No.8, 2024, pp.66-68

[5] Dai Y. Integration of vocal technique and artistic expression in vocal art performances. *Frontiers in Art Research*, Vol.6, No.2, 2024, pp.7-11.

[6] Zhao X. A study on the inheritance status and enhancement path of Chinese ancient poetry art songs integrated into vocal music teaching in contemporary colleges and universities. *Pacific International Journal*, Vol.6, No.3, 2023, pp.166-171.

[7] Yao L. Research on the inheritance of Chinese traditional vocal art in contemporary opera. *Frontiers in Art Research*, Vol.4, No.6, 2022 pp.83-86.

[8] Bin F . The Application and Inheritance of Chinese Ancient Poetry and Art Songs in Vocal Teaching. *Art and Performance Letters*, Vol.5, No.4, 2024, pp.1-6.

[9] Yao L. Research on the inheritance of Chinese traditional vocal art in contemporary opera. *Frontiers in Art Research*, Vol.4.0, No.6.0, 2022, pp.83-86.

[10] Qi G. Inquiry and Reflection of the Inheritance and Reference of National Vocal Music Art. *4th International Conference on Education, Language, Art and Intercultural Communication (ICELAIC 2017). Atlantis Press*, 2017, pp.669-672.

[11] Chang W. The integration of artificial intelligence and ethnic music cultural inheritance under deep learning. *Computer Science and Information Systems,* Vol.22, No.3, 2025, pp.36-36.

[12] Gao S, Phanlukthao P, Guo K. Exploring the Preservation, Inheritance, and Digital Development of Ethnic Dance in a Modern Technological Environment. International Journal for Housing Science & Its Applications, Vol. 47, No.1, 2025, pp.181-187.

[13] Zhang Y, Liu C. Melody prediction of vocal performance using LSTM and attention mechanism and its application in folk music innovation. Journal of Computational Methods in Sciences and Engineering, Vol. 25, No.4, 2025, pp.3608-3622.

[14] Yu H. Deep Learning-Based Artistic Inheritance and Cultural Emotion Color Dissemination of Qin Opera. Frontiers in Psychology, Vol. 13, 2022, pp.872433-872433.

[15] Li X, Lin J, Zhang X. Dynamic Transmission and Innovative Transformation of Cultural Heritage: Generative Artificial Intelligence Practices Based on Cultural Cognitive Models. Applied Sciences, 2025, Vol. 15, No.23, 2025, pp.12651.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

JUN MA: Writing – original draft, Writing –review & editing, Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

APPENDIX

Table 1. Evaluation index system for inheritance and innovation of traditional ethnic vocal music

First-level indicator	Secondary indicator	Third-level indicator	Indicator type	Weight
Inheritance Foundation (A)	Inheritance subject (A1)	Number of inheritors (A11)	quantitative	0.05
		Professional level of inheritors (A12)	qualitative	0.07
		Proportion of young inheritors (A13)	quantitative	0.06
	Heritage Resources (A2)	Integrity of Music Score Literature Preservation (A21)	qualitative	0.04
		Standardization of Skill Inheritance (A22)	qualitative	0.08
		Inheritance Base Construction Level (A23)	qualitative	0.05
Inheritance effectiveness (B)	Mastery of skills (B1)	Mastery of core vocal techniques (B11)	qualitative	0.07
		Vocal style restoration degree (B12)	qualitative	0.06
		Complete performance of the work (B13)	qualitative	0.05
	Cultural Transmission (B2)	Accuracy of Transmitting Ethnic Cultural Connotation (B21)	qualitative	0.08
		Reflectance of Regional Cultural Characteristics (B22)	qualitative	0.06
		Audience cultural identity (B23)	quantitative	0.05
Innovation potential (C)	Element Fusion(C1)	Integration degree of multicultural elements (C11)	qualitative	0.07
		Integration of Traditional and Modern Techniques (C12)	qualitative	0.06
		Adaptability of innovative elements (C13)	qualitative	0.05
	Formal Innovation (C2)	Innovation level of work theme (C21)	qualitative	0.04
		Innovation level of performance form (C22)	qualitative	0.03
		Innovation level of communication forms (C23)	qualitative	0.04
Communication effect (D)	Audience coverage (D1)	Audience size (D11)	quantitative	0.03
		Rationality of audience age distribution (D12)	quantitative	0.02
		Cross regional transmission range (D13)	quantitative	0.02
	Social Impact (D2)	Media exposure (D21)	quantitative	0.02
		Social attention (D22)	quantitative	0.02
		Cultural Industry Value (D23)	quantitative	0.03

Source: created by the authors

Table 2. Comparison of model performance indicators

Model	Mean Absolute Error (MAE)	Root mean square error (RMSE)	Coefficient of determination(R^2)	Convergence iteration times	Training time (s)
Improved BP neural network	0.042	0.058	0.876	420	186
LSTM network	0.038	0.051	0.898	280	245
BP-LSTM Fusion Model	0.025	0.032	0.964	350	312

Source: created by the authors

Table 4. Comparison of model evaluation and expert evaluation (partial samples)

Sample ID	Ethnicity/School	Model evaluation score	Expert evaluation score	Error	Relative error(%)
1	Han/Shaanbei Folk Songs	0.876	0.892	0.016	1.79
2	Mongolian/Changdiao	0.912	0.905	0.007	0.77
3	Tibetan/Tsangma	0.785	0.793	0.008	1.01
4	Uyghur/Muqam	0.853	0.847	0.006	0.71
5	Zhuang/Mountain Songs	0.721	0.735	0.014	1.90
6	Han/Jiangnan folk songs	0.897	0.901	0.004	0.44

Source: created by the authors