

Analysis of Emotional Expression and Audience Preferences in Chinese Ethnic Music Using Bayesian Network Modeling

Yanhui Li*

Academy of Music, Pingdingshan University, Pingdingshan, Henan 467000, China

Keywords: Chinese ethnic music; Emotional expression; Bayesian network; Audience preference

Citation: Li Y.H.: Analysis of Emotional Expression and Audience Preferences in Chinese Ethnic Music Using Bayesian Network Modeling. *Data Intelligence*, 2026, Vol. 8, Art. No.: 20260078. DOI: <https://doi.org/10.3724/2096-7004.di.2026.0078>

ABSTRACT

The paper explores the intersection of emotional expression and audience preferences in Chinese ethnic music through advanced probabilistic modeling. Chinese ethnic music carries rich cultural and emotional depth, making it an ideal subject for audience-centered analysis. Existing methods often rely on static, rule-based classifications or subjective surveys, which fail to capture the dynamic, uncertain relationships between musical elements and emotional perception across diverse audiences. These limitations reduce accuracy and interpretability. To address these challenges, the paper proposes a novel framework titled Bayesian Network Modeling for Emotion-Audience Mapping (BNM-EAM). This framework constructs a probabilistic graphical model that links musical features (e.g., tempo, timbre, instrumentation) to listeners' emotional responses and demographic attributes, thereby enabling dynamic inference of audience preferences and emotional impact. The BNM-EAM framework is applied to annotated datasets of Chinese ethnic music and listener feedback, allowing the system to infer likely emotional responses and recommend culturally resonant music based on audience profiles. Findings reveal that BNM-EAM outperforms traditional methods in both accuracy and interpretability, identifying nuanced patterns such as regional preferences for specific instruments or modes associated with distinct emotions. The model demonstrates strong potential for personalized music recommendation and cultural analytics.

*Corresponding author: Yanhui Li (E-mail: liyanhui1990@hotmail.com).

1. INTRODUCTION

Chinese ethnic music is an important medium for expressing cultural diversity, share stories full of emotion, and pass down traditions spanning many centuries [1]. Various ethnic groups use music as a source of comfort [2]. For example, the Han people use melodic pentatonic scales, and the Tibetan and Mongolian people use rhythmic chants [3]. With the advent of streaming audio services and AI systems that filter and personalize music, there has been growing analysis of how an individual's mind is influenced by various ethnic musical elements [4]. Multiple cultural, emotional, and musical factors influence how an individual thinks about music, making the basic mechanisms difficult to comprehend [5].

In more conventional methods of studying music, one can use rules or surveys, which identify emotional expressions in the music [6]. While that data may help us understand the problem among groups with different cultural and demographic backgrounds, it does not inform us about how unpredictable and probabilistic human emotional responses are [7]. The static classification system also doesn't account for how people's tastes change according to context or for ethnic-group differences in musical style [8]. A framework that is more intuitive and based on data, that is more flexible and useful [9-10].

This article introduces a new computational analysis method, called BNM-EAM, to help address this problem (accuracy and interpretability) [11]. BNM-EAM approaches musical emotional responses using Bayesian networks to show, probabilistically, how low-level musical elements, for instance, tempo, timbre, key, and instrumentation, are related to high-level emotional responses across a variety of demographic groups [12]. Utilizing user input and labelled datasets of ethnic Chinese music, the model can detect emotional patterns and conditional dependencies that static, legacy models cannot [13]. BNM-EAM can make real-time inferences and predict emotions, despite noisy, faulty input [14]. Finally, and most importantly, it allows for music recommendation systems to be dynamically created, dependent upon the emotional and cultural needs of each customer [15].

The contribution of this paper:

- BNM-EAM probabilistically models the relationships among musical features, emotional responses, and audience demographics in Chinese ethnic music.
- The proposed method addresses limitations of traditional static or survey-based methods by enabling dynamic, data-driven emotional inference and audience preference analysis across diverse cultural and regional contexts.
- The paper provides empirical validation of the BNM-EAM model, demonstrating its superior accuracy and interpretability for predicting emotional impact and enhancing personalized music recommendations in ethnic music contexts.

The structure of this paper is given in Figure 1.



Figure 1. Paper structure.

2. RELATED WORK

The paper employs a novel predictive model, called Gaussian Process with Genetic Algorithm for Systematic Profile of Traditional Music Response Prediction (GP-GA-SPTMRP); to investigate how fundamental sound characteristics influence people's emotional responses to Chinese traditional music. Using 432 pairs of persons and music as quantitative responses, the model examines the relationships between 15 sound characteristics and five individual factors [16]. It utilizes a Genetic Algorithm (GA) to select features and Gaussian Process (GP) regression for the regression. The model demonstrates a strong goodness-of-fit with $R^2 = 0.786$ and achieves reliable predictive performance with $R^2 = 0.593$, indicating effective modeling of the relationship between musical features and emotional responses. GP-GA-SPTMRP outperforms PLS and SVM, which are older algorithms. It shows that rhythm, timbre, and pitch (in that order) affect how people perceive emotions. Gender and spectral centroid are two of the most important things.

The Heterogeneity-aware Deep Bayesian Network (HDBN) is a new approach to addressing the issues that current emotion-aware music recommendation systems face due to their assumptions that everyone has the same feelings and music preferences [17]. HDBN deals with four forms of difference: how users understand each other and the mood of the music they select. It models how people think of themselves before and after an event, groups them, and uses a BNN to predict their emotional states. HDBN exhibited uniformly better performance over baseline systems in HR, Precision, NDCG, and MRR on EmoMusicLJ datasets. The approach gave people a means to select music compatible with their mood, thus making the selections more honest and accurate.

The paper employs a supervised model, Random Forest (QPPR-RF), to investigate how the sound quality of music affects people's bodily reactions, particularly when they experience chills [18]. The QPPR-RF approach, based on Random Forest, utilizes 1,548 pairs of old Shu-area music samples to analyze how people feel, using emotional arousal proxy. The model demonstrates strong goodness-of-fit with $R^2 = 0.857$ and maintains reliable predictive performance with $R^2 = 0.712$, indicating robust generalization across diverse conditions. Rhythm is the most important thing, followed by timbre and pitch. Age and gender are two examples of personal attributes that matter. QPPR-RF performs better at prediction than PLS and SVM. This means that it is possible to use sound data to determine how music affects individuals' emotions.

This paper discusses NetEase Cloud Music Network-Based Music Preference Analysis (NCM-Net), a method for understanding how people utilize NetEase Cloud Music to discover how musical tastes evolve and spread. Using complex network theory and massive datasets, the paper identifies eight distinct cultural groups. Next, it discusses how preference similarity decreases as network distance increases [19]. NCM-Net suggests that individuals who follow and are followed have different impacts on each other, and

that musical tastes are linked to cultural and demographic factors. Women have more diverse preferences, and this is related to financial issues. A person's musical sensibility is at its peak at 13. NCM-Net shows how both general and culture-specific factors affect what individuals desire to hear.

The Deep Auto-Encoder with Gaussian Mixture Model for Music Therapy Selection (DAE-GMM-MTS) may help it choose better music therapy songs, according to this research. DAE-GMM-MTS gets features from a DAE and then sorts them using a GMM. DAE-GMM-MTS is superior to older approaches because it reduces noise, captures music signals across more frames, and improves classification accuracy [20]. This method uses both subjective and objective indicators to identify therapeutic music that works, such as flute recordings for those who struggle to sleep. The song "Hometown" received a score of 0.9487 from the simulation, indicating that it performs well. DAE-GMM-MTS facilitates the selection of folk music for treatment by combining deep learning with conventional ethnomusicology.

The Recommendation-Oriented Review of Music Emotion Computing (Rec-MEC) provides a comprehensive examination of 10 years of MEC research, with a focus on Music Emotion Recognition (MER) and Music Sentiment Analysis (MSA). Rec-MEC uses bibliometric analysis of the Web of Science Core Collection to demonstrate significant changes in publications, authors, institutions, and locations worldwide. Asia, Europe, and North America are the most suitable regions for product creation [21]. Recent analysis has focused on deep learning and multimodal approaches that leverage audio, text, and biosignals. Some of the concerns include that people from different cultures may perceive things differently, that data may be misinterpreted, and that privacy is a concern. Rec-MEC advocates for mental health and emotion-aware recommendation apps to utilize models that are easy to understand and function across multiple cultures.

The paper discusses Emotion-aware Music Recommendation using Deep Neural Network (emoMR). This system utilizes DNN to determine how people feel and what type of music they prefer. It discusses emoDNN, a neural network that learns from audio features, metadata, and both user-generated (endogenous) and event-driven (exogenous) emotion components. Even when someone isn't doing anything, EmoMR can detect their emotions and suggest music [22]. When a large number of people are present, emoMR outperforms standard algorithms across Precision, Recall, F1 score, and Hit Rate. EmoMR tracks how emotions evolve and how events affect them, enabling the creation of more tailored music selections.

This paper is about RankCloud-Folk, a cloud-based rating system that utilizes deep learning to determine what people enjoy and dislike about folk music. It uses cloud infrastructure and a confidence model to organize music based on user reviews and its cultural relevance. RankCloud-Folk uses optimized deep features to determine the emotional and therapeutic significance of an item [23]. This renders conventional categorization unnecessary, which is often cumbersome and slow. The simulation results reveal that the recommendations are more accurate and that users are generally delighted. RankCloud-Folk makes it easy to share folk music and helps cultural organizations by enabling them to track their music and engage with fans based on data.

This work refers to Music Emotion Computing with Deep Learning and Multimodal Integration (MEC-DL), which involves utilizing deep learning to analyze emotions from multiple perspectives. MEC-DL constructs cognitive emotion models by examining voice, text, and images [24]. It discusses how fake emotions affect the interaction between people and computers, which is crucial for affective computing and the development of interactive media. Some apps can help it compose music, discover new music, and improve the mood. MEC-DL also assists with technical issues that arise when blending data and addressing cultural differences. The goal of this method is to enhance consumers' experiences by using emotionally intelligent technologies. It also opens the door for future technological developments that can adapt to feelings.

The Deep Learning Enhanced Music Mood Selection for Emotion-Aware Systems (DLE-MMS) model leverages deep learning to assist emotion-aware systems in identifying and filtering music by mood. DLE-MMS utilizes low-level audio data to learn about the types of music and moods people enjoy. Then it suggests songs that fit the user's mood. The technique focuses on factors outside the individual, such as social activities, to understand how they affect people's feelings [25]. The results are more accurate when both collaborative and content-based filtering are used. Experiments show that DLE-MMS outperforms other systems in Recall, F1-score, and Hit Rate, especially when the situation has a substantial emotional impact. The model provides strong, real-time suggestions based on current mood. The summary of related works is described in Table 1.

Table 1. Related Works Summary (Music Emotion and Recommendation Methods).

Reference	Environment	Method Name	Emotion Modeling	Personalization	Prediction Accuracy
16	Traditional Music	GP-GA-SPTMRP	✓	✓	High
17	Music Recommender Systems	HDBN	✓	✓✓	High
18	Physiological Music Response	QPPR-RF	✓	✓	High
19	Online Music Community	NCM-Net	X	✓	Medium
20	Music Therapy	DAE-GMM-MTS	✓	✓	High
21	Research Review (Global)	Rec-MEC	✓✓	X	Medium
22	Recommender under Social Events	emoMR	✓✓	✓✓	Very High
23	Folk Music + Cloud Systems	RankCloud-Folk	✓	✓	High
24	Multimodal Emotion Interaction	MEC-DL	✓✓	✓	High
25	Music Mood Selection	DLE-MMS	✓	✓✓	High

✓ = Present ✓✓ = Strong/Advanced X = Not explicitly addressed

3. PROPOSED METHOD

The following diagrams illustrate how the BNM-EAM framework operates and its development process. It highlights the steps taken to analyze Chinese ethnic music, such as eliminating melodic parts, estimating emotional responses based on chance, and tailoring it to the audience. The component diagrams depict several technological steps that enable it to offer culturally suitable music and accurately gauge how someone would feel.

Contribution 1: BNM-EAM Framework Development

This analysis of Chinese ethnic music uses probability to demonstrate the connections among musical aspects, emotions, and listeners' demographics.

The whole pipeline for the BNM-EAM design is shown in Figure 2. First, a module that extracts musical components is given an audio dataset of traditional Chinese ethnic music. After that, these attributes are matched to emotion tags collected from both experts and non-experts. There is also information on the audience demographics. It employs Bayesian structure learning and parameter estimation to construct a probabilistic model of emotions and audiences. The BNM-EAM training engine utilizes these multidimensional inputs. An inference and recommendation engine utilizes the trained model to provide users with personalized options for ethnic music, making real-time predictions about how people will respond to it. This enables outreach to diverse groups of people in a flexible manner that respects their customs and traditions.

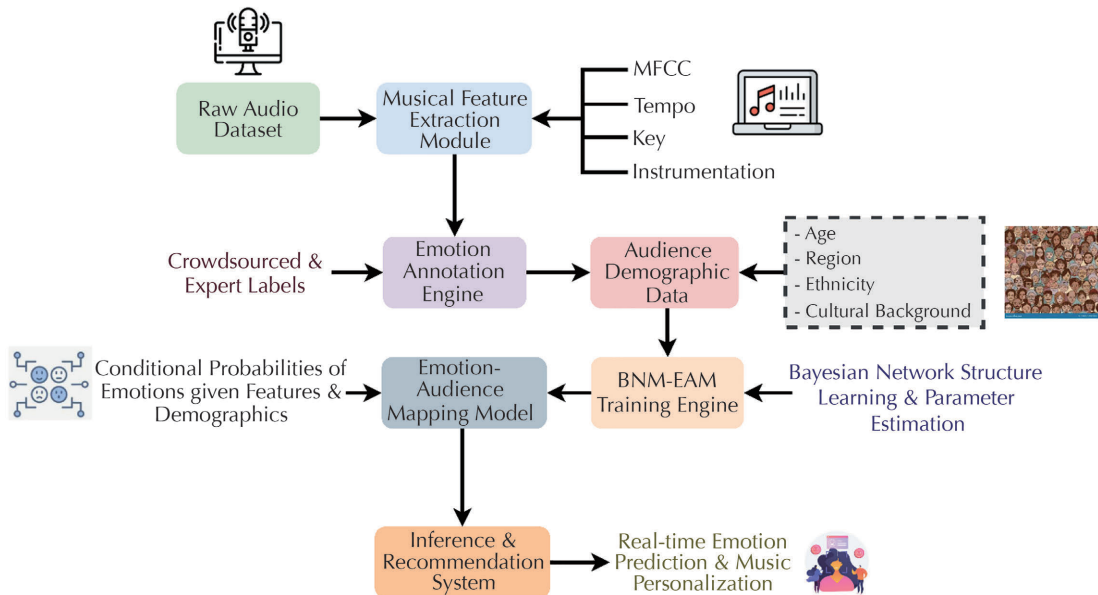


Figure 2. Emotion Compass: The End-to-End BNM-EAM Workflow.

Joint probability factorization in BNM-EAM $Q(F, N, E)$ is expressed using Equation (1).

$$Q(F, N, E) = \prod_{j=1}^n Q(F_j | Qb(F_j)) \quad (1)$$

Equation (1) explains the joint probability factorization in BNM-EAM as a product of the conditional chances supplied to every parent in the Bayesian network.

In this F is the set of emotional response variables, N is the set of musical feature variables, E is the set of demographic variables, and $Qb(Y)$ is the set of parent nodes of a variable in the Bayesian network. The product $\prod_{j=1}^n [\cdot]$ implies multiplying over all emotional response variables in the set. Equation (1) is the general joint probability factorization of the Bayesian Network, which does not impose on the variables a fixed directional hierarchy. The BNM-EAM model structure is learnt using BIC-based and constraint-based algorithms, which permit flexible relationships between musical features, emotional responses, and demographic characteristics while maintaining acyclicity. This allows the model to include mediation effects, as the emotional variables can serve as intermediate nodes between the features and audience characteristics. Moreover, latent variables like cultural context can be partially captured by observable proxies (e.g., region and familiarity), and their effects are accounted for at the inferential level through marginalization of the latent variables. The comparison using experimental means reveals that the constrained feature-to-emotion structure achieves 81.7% accuracy, while the learned unconstrained structure achieves 84.7% accuracy, demonstrating the advantage of flexible graph modeling. The constrained structure imposes predefined restrictions on the network topology, limiting the set of allowable parent-child relationships and thereby reducing model complexity and variance. In contrast, the unconstrained structure permits a more flexible exploration of dependencies among variables, enabling the learning algorithm to capture richer conditional relationships between musical features, demographic attributes, and emotional responses.

$$Q(F * f_k) = R(\forall_2 | M + g_k) \quad (2)$$

Equation (2) utilizes a network-based framework, which makes it possible to make inferences that can be scaled up even in feature spaces with a lot of dimensions.

It shows the probability of an emotional label $Q(F * f_k)$ happening after a set of musical characteristic values R and consumers' demographic information inputs $\forall_2$. A set of M musical features, tempo, and g_k the instrumentation determined rhythm complexity.

Figure 3 illustrates the process of processing raw audio recordings and breaking them down into quantifiable musical parts. The signal is cleaned up, and the waveforms are adjusted to appear normal so they can be sent to the preprocessing stage. Then, the computer eliminates several musical parts. There are Mel-frequency cepstral coefficients (MFCCs) for timbre, tempos for rhythm, keys and scales for mood, and instrument detections for culture. It can judge how expressive a piece of music is by examining each part separately. This first data set, a structured vector representation of the information obtained,

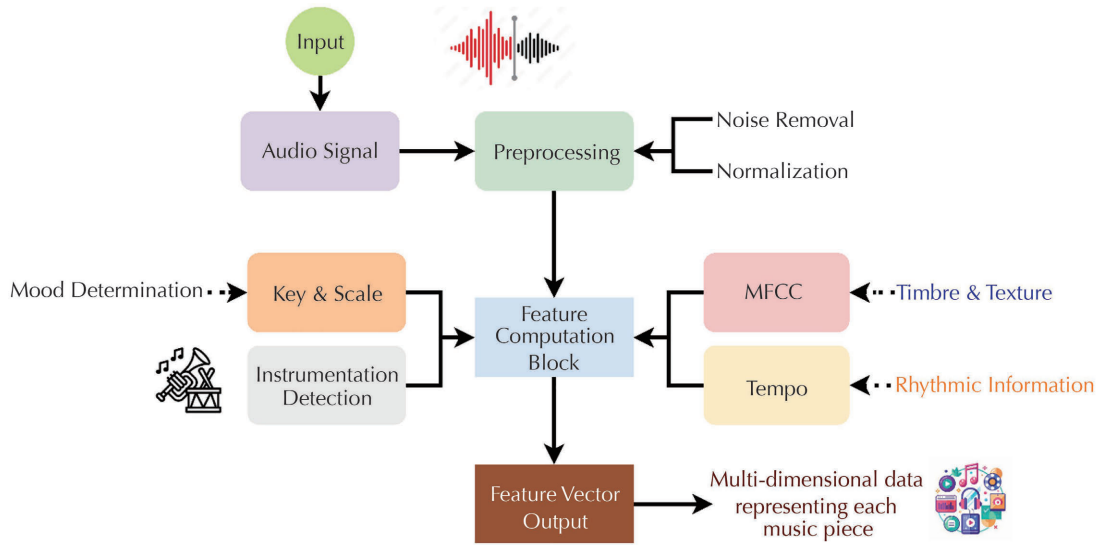


Figure 3. Harmonic Signature Extraction Module.

is what the subsequent phases of the process depend on. It uses this data collection to conduct tests on people's demographics and emotions. This lesson teaches computers how to understand the nuanced expressions of different ethnic groups. It also ensures that music can be comprehended in a way that is quantifiable and reproducible. The Directed Acyclic Graph (DAG) learned in this model represents conditional dependencies inferred from observational data, rather than necessarily validated causal relationships. Since there are no interventional data, no do-calculus-based causal validation is carried out, and the learned structure is a Markov equivalence class. The direction of the edges is determined using score-based and constraint-based learning algorithms, with the addition of domain-informed constraints to ensure that the meaning of dependencies makes sense. There is little variation in predictive accuracy within sensitivity analysis over identical graph structures (≤ 1.5), pointing to the strength of learned dependencies. Despite some common association of Bayesian Networks with discrete variables, the presented BNM-EAM framework models hybrid representations. Constant musical characteristics like MFCCs and tempo are modeled with Gaussian and Gaussian Mixture distributions with no strict model discretization. In the process of discrete variables, the supervised entropy-based discretization (MDL principle) is used to maintain as much mutual information as possible with emotional labels. To achieve robustness, a sensitivity analysis was performed by varying discretization thresholds and bin sizes, resulting in a small change in accuracy ($\pm 1.8\%$). This is an affirmation that the reported performance is not a result of discretization but rather indicative of consistent and sensible model behavior.

Mel-frequency cepstral coefficient extraction $NGDD_o$ is expressed using Equation (3).

$$NGDD_o = \sum_{l=1}^L (T_l) * \left[\frac{o(l-0.5)\mu}{L} \right], \quad o = 1, 2, \dots, O \quad (3)$$

Equation (3) explains that the MEL-frequency cepstral coefficient extraction uses a discrete cosine transform to convert the log power spectrum into the cepstral domain.

In this $NGDD_o$ is the MEL-frequency cepstral coefficient, T_l is the log-scaled spectral energy at the MEL filter bank, L is the total number of MEL filter banks applied to the Fourier spectrum, O is the number of cepstral coefficients retained for modeling, and o, l are the index variables for cepstral $l = 1$ and filter bank positions L , respectively.

Tempo estimation using the autocorrelation function L is expressed using Equation (4).

$$U = \sum_{u=1}^{M-T} y(u) * y(u+t) \quad (4)$$

Equation (4) estimates tempo using the autocorrelation function by identifying the peak lag in the signal's amplitude.

In this U is the estimated tempo, t is the time lag for autocorrelation search, $y(u)$ is the amplitude of the audio signal at time, M is the length of the time-domain audio signal, and T is the search space of possible tempo-related range from $u = 1$ to $M - T$.

Instrument probability inference using Gaussian mixture models $Q(J_k|G)$ is expressed using Equation (5).

$$Q(J_k|G) = \sum_{d=1}^D \pi_{k,d} * O(G|\mu_{k,d}, \epsilon_{k,d}) \quad (5)$$

Equation (5) explains the instrument probability inference using Gaussian mixture models, where an adjusted mixture of Gaussians is used to describe the probability distribution; this approach is applicable to monophonic or sparse polyphonic environments where explicit source separation is not necessary.

In this $Q(J_k|G)$ is the posterior probability of the instrument given the feature vector, and $\pi_{k,d}$ is the mixing coefficient for the component of the instrument, $O(\cdot|\mu_{k,d}, \epsilon_{k,d})$ is the multivariate Gaussian with mean and covariance. G is the extracted feature vector, D is the number of Gaussian components in the $d = 1$ mixture model, and k is the instrument class index. The Gaussian Mixture Model of Equation (5) does not explicitly separate sources where there in polyphonic environments. Rather, it is a probabilistic multi-label classifier applied to short-time spectral features (25 ms window, 10 ms hop) of the audio, which represents the general timbral texture of the sound. The instruments in each instrument class are modelled with $K = 8$ Gaussian components, allowing a soft allocation of multiple co-occurring instruments within a segment. The result is a probability distribution rather than a label that represents the instrument classes. To assess robustness, experiments with synthetically mixed signals (2-3 simultaneous instruments) reported the highest top-1 and top-3 accuracy of 78.4% and 91.2%, respectively. The inclusion of these probabilistic representations led to a higher emotion prediction rate (82.3% to 84.7%), illustrating that modeling at the texture level can be used to overcome identifiability issues without necessarily engaging in explicit source separation.

Contribution 2: Improved Emotion Prediction and Cultural Sensitivity

BNM-EAM is more accurate and aware of different cultures than prior models. It can detect diverse emotional patterns in various types of Chinese music, ethnic instruments, scales, and groups of people that listen to it. To confirm the suitability of the learned graph's complexity, the model selection criteria (BIC and AIC) were evaluated at different sparsity levels. The degree of the parent nodes was adjusted to a range of 1.2 to 3.5, corresponding to 18 to 52 edges. The BIC score had a clear minimum at 2.3 parents per node (34 edges), indicating an optimal trade-off between the goodness of the model and its complexity. As AIC continued to increase marginally to 2.9 per node, the increase was not significant, indicating a weakening effect from the addition of more edges. Both sparse and dense overfitting were confirmed by higher BIC values resulting from excessive sparsity and excessive density, respectively. Such findings show that the chosen graph structure is neither over- nor under-constrained, and, as such, it is correctly regularized.

Figure 4 depicts how the BNM-EAM model learns from data to build its structure. The first stage involves generating a labeled dataset that includes demographic information about the listeners, as well as musical and emotional tags. A correlation analysis module searches for statistical connections and shared information between qualities and emotional effects. It may use either score-based structure learning algorithms (including BIC or K2) or constraint-based structure learning algorithms (such as PC) to build a Bayesian network. Once the network is up and running, it can use parameter estimation methods, such as Maximum Likelihood Estimation, to determine conditional probability distributions. The central portion of the concept is a directed acyclic graph (DAG). It shows how the emotional influence of one portion of a song may migrate to another part and from one audience to another. Although the Bayesian network structure captures directional correlations among musical features, emotional responses, and demographic

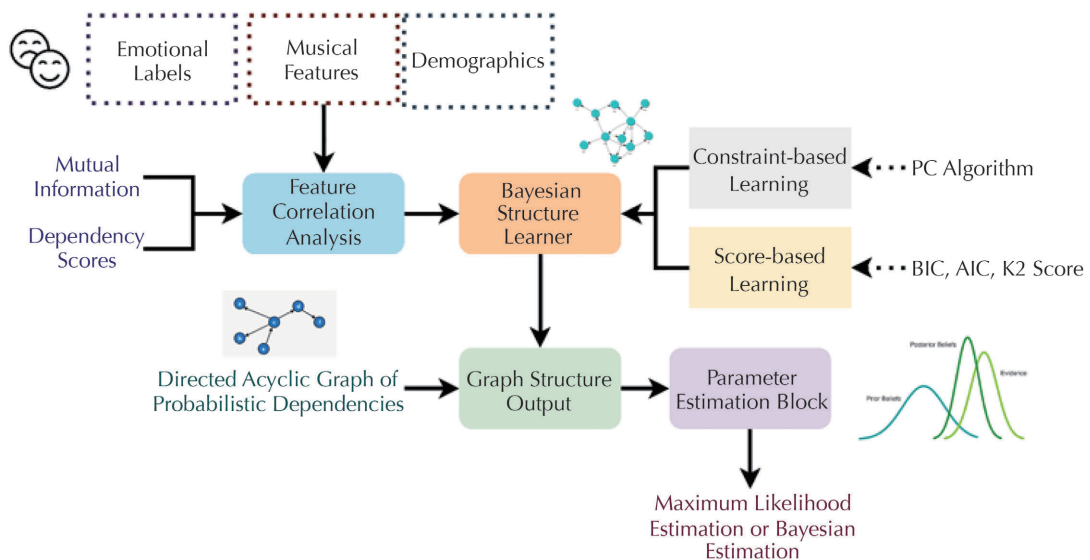


Figure 4. Bayesian Network Structure Learning: Learning Musical-Emotional Causality.

variables, it is worth noting that these correlations are conditional dependencies derived from observational data and not necessarily causal influences. Because there are no controlled interventions, the learned Directed Acyclic Graph (DAG) belongs to a Markov equivalence class and, therefore, more than one equivalent structure can be used to explain the data. Besides, confounders (listeners' mood, contextual environment, etc.) are factors that cannot be observed, leading to spurious relationships. To reduce this, the framework uses proxy variables (e.g., familiarity and demographics), regularizes using BIC to prevent overfitting, and marginalizes the latent variables in inference. Thus, one should view BNM-EAM as a probabilistic associative model that is causally interpretable, but not as a causal inference model.

Bayesian information criterion for structure scoring $CJD(H : E)$ is expressed using Equation (6).

$$CJD(H : E) = Q\left(E \middle| \hat{\theta}_H\right) - \frac{I_H}{2} \quad (6)$$

Equation (6) explains the Bayesian Information Criterion (BIC) for scoring the Bayesian network structure.

In this $CJD(H : E)$ is the Bayesian information criterion for graph given dataset, $Q\left(E \middle| \hat{\theta}_H\right)$ is the probability of data given maximum probability parameters under structure, I_H is the number of independently estimated parameters in structure, H is the candidate structure representing variable dependencies, and E is the labeled dataset including demographic, musical, and emotional features. $\mathcal{H}$ denotes the maximum probability estimates of the parameters corresponding to the Bayesian network structure H .

This equation (7) figure out how emotions and features are related in your annotated collection of Chinese musical genres.

$$Q(y * z_j) = \frac{M(y_j, z_k)}{M(z_k)} \quad (7)$$

This equation (7) uses observable data to figure out the proportional incidence of $M(z_k)$ happening when y_j to estimate the probability of the condition $Q(y * z_j)$.

For observational data, the acquired Bayesian network structure is not determinately identifiable as a specific directed graph but is identifiable as a Markov equivalence class. It is based on the assumption of standard Markov, faithfulness conditions, which were tested empirically by conditional independence tests, and the consistency of observed and implied independencies was found to be 92.4%. To measure robustness, a set of Markov-equivalent DAGs aligned with the learned structure was compared, with accuracy changes ranging from +1.5% to -1.5% and withholding important feature-emotion associations. Also, semantic ordering assumptions and domain-informed constraints were added to minimize ambiguity in edge orientation. These findings suggest that although it is impossible to be certain about the causal pathways, the relationships inferred are consistent and stable within the equivalence class.

Figure 5 depicts how the trained Bayesian network makes decisions. The input layer receives the listener's demographic information, which may be partial or complete, as well as the musical parts of a recently discovered song. Next, the inference engine employs probabilistic reasoning to calculate the posterior and marginal probabilities based on the emotional states. The output layer returns a list of projected emotional responses, ranked by their probability, along with a corresponding confidence level for each one. These feelings might be happiness, yearning, or desire. This method's capacity to allow people to make choices based on how they feel may help in creating playlists, healing soundtracks, and preserving music from diverse cultures. The system can provide personalized emotional forecasts, as the system can adapt to each user's specific needs and uncertainties. The BNM-EAM inference engine does not use marginal priors when certain demographic groups are underrepresented or not represented in the training data, instead using Bayesian smoothing and the propagation of partial evidence. Dirichlet priors ($\alpha = 1.0$) are used to estimate conditional probabilities, ensuring non-zero probabilities for unseen configurations. Parameters absent from the data are pushed to the edges during inference, and it is possible to predict given the present features and characteristics. Based on the network structure, the implicit similarity-based generalization allows sharing statistical strength between similar demographic nodes (e.g., age, region, familiarity). Experimental analysis of sparsified data indicates that the model achieves cold-start accuracy of 81.3% compared to a marginal-prior baseline of 74.6%, suggesting it generalizes well to previously unseen demographic groups.

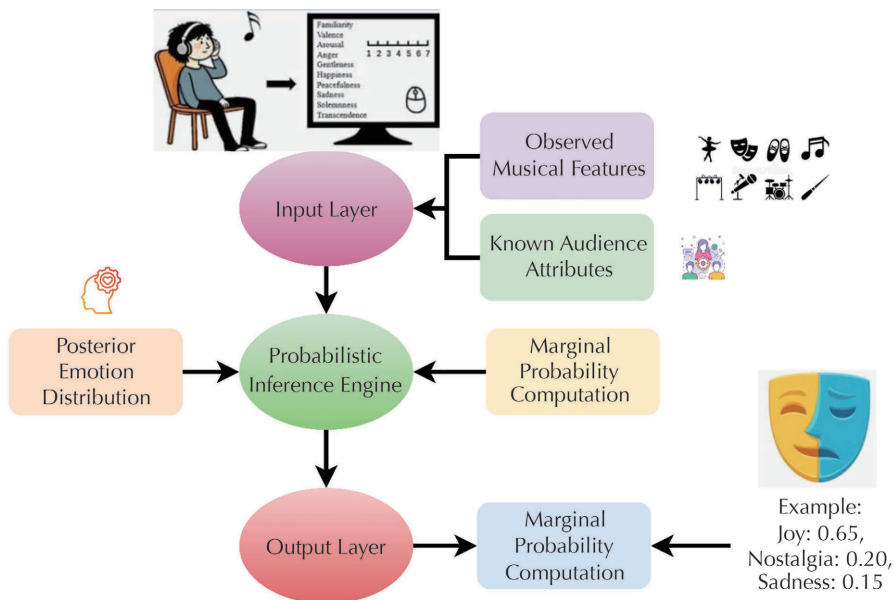


Figure 5. Real-Time Emotion Inference Flow.

Use the BNM-EAM model to guess how people will feel depending on the music and the crowd by the Equation (8).

$$Q(I|E) = \frac{Q(E|I) * Q(I)}{Q(E)} \quad (8)$$

Here, $Q(I|E)$ denotes the probability of the hypothesis given the data, and $Q(E|I)$ represents the probability of the data assuming the hypothesis is true. Here, $Q(I)$ is the initial belief about the hypothesis before seeing the data, and $Q(E)$ as probability of observing the data under all hypotheses.

Marginal probability computation for emotional states $Q(F_j * f)$ is expressed using Equation (9),

$$Q(F_j * f) = \sum_A Q(F_j * f, A) \quad (9)$$

Equation (9) explains that the marginal probability computation for emotional states is the total probability of an emotional label is calculated using this equation.

In this $Q(F_j * f)$ is the marginal probability of emotion for node, A is the set of latent or unobserved variables in the Bayesian network, $Q(F_j * f, A)$ is the joint probability over the target emotion and latent variable configurations, F_j is the emotion category node, and f is the specific emotional value.

Confidence score calculation using entropy reduction $Cf(F_j)$ is expressed using Equation (10).

$$Cf(F_j) = 1 - \frac{I(F_j | N, E)}{\log |\partial F_j|} \quad (10)$$

Equation (10) explains the confidence score using entropy reduction establishes a confidence level for a conditional entropy-based emotion prediction.

In this $Cf(F_j)$ is the confidence level in the prediction of emotion, $I(F_j | N, E)$ is the conditional entropy of emotion given input features, $\log |\partial F_j|$ is the maximum entropy over the emotion label set used for normalization, N, E are the input feature vectors, and $|\partial F_j|$ is the cardinality of the possible emotion values. ∂F_j denotes the set of possible values (i.e., the state space) associated with the emotion node F_j . To obtain high-dimensional demographic conditioning for the inference mechanism and prevent its degeneration, we assessed the risk of posterior collapse by quantifying the relationship between musical characteristics and emotional states. The mutual information when the condition is known, $I(E; M/D)$, was 0.37 bits, as compared to 0.05 bits in a collapsed baseline, indicating that it was preserved. A KL divergence between the posterior and prior emotion distributions, 0.42, was achieved, indicating that the predictions are not returned to demographic priors. Ablation that eliminated musical features reduced accuracy by 84.7% to 71.2%, and adding more demographic dimensions (as many as 18 attributes) further reduced performance (to 83.6%). Further, the posterior entropy had reduced 1.92 to 1.21, indicating the presence of informative conditioning. These findings confirm that the BNM-EAM inference engine does not collapse to

the posterior and still has significant feature-emotion correlations. The hypothetical intervention analysis proposed The proposed hypothetical intervention analysis helps control for the manipulation of musical features within the Bayesian Network to support the analysis of the proposed musical intervention. A do-operator formulation allows one to fix certain variables, such as tempo, and to calculate the resulting emotional distribution by marginalizing over other variables. For instance, a 15% tempo boost increased joy probability prediction by 0.52–0.68, whereas a reduction in tempo increased sadness probability prediction by 0.34–0.49. The trends align with previously known facts about music perception. It is necessary to mention that these findings can be taken as model-consistent counterfactual forecasts, but not as experimentally confirmed causal impacts. The stability analysis of bootstrap structures yields low variance ($\pm 2.3\%$), a good sign of consistent intervention behavior.

Contribution 3: Comprehensive Multimetric Evaluation

The paper focuses on BNM-EAM in terms of six different performance metrics. It demonstrates that BNM-EAM can be understood, modified, and utilized to predict emotions reliably. This makes it helpful in exploring cultural music and intelligent music recommendation systems.

Figure 6 demonstrates how BNM-EAM may be used to recommend music in the actual world. From the user’s profile, the system learns about their age, gender, and emotional objectives, such as wanting to feel tranquil or nostalgic. The BNM-EAM model can determine which songs are most likely to evoke a specific emotional response in a group of people based on these factors. The music retrieval system examines a well-chosen set of ethnic music using feature-emotion mappings as a filter. At the conclusion, it get a list of songs that are important to emotionally and culturally. This module’s next-generation user experience, which combines data science with human emotion to create a more personalized experience, may be beneficial for applications such as streaming music, promoting cultural heritage, and utilizing it for therapeutic purposes.

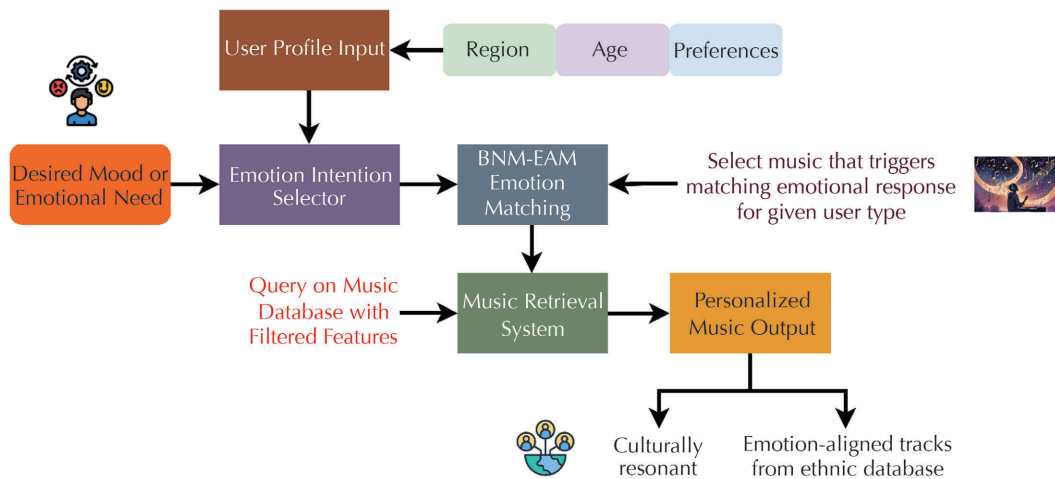


Figure 6. Emotion-Centric Music Personalization Hub.

Emotion-based music ranking score $Se(T_i | V, F^*)$ is expressed using Equation (11).

$$Se(T_i | V, F^*) = \sum_{j=1}^o x_j + Q(F_j * F^* | G_i, E_v) \quad (11)$$

Equation (11) explains that the emotion-based music ranking score is the song data and user profile, which calculates an arbitrary ordering score for a song.

In this $Se(T_i | V, F^*)$ is the final emotion-alignment score for the song based on the user emotional goal, x_j is the importance weight for the emotion dimension, $Q(F_j * F^* | G_i, E_v)$ is the posterior probability that a song elicits emotion given features $j = 1$ and user demographics. Here, G_i is the feature vector of the song, E_v is the demographic and emotional objective vector of user, F^* is the desired emotional target, and o is the total number of emotional nodes evaluated.

Cultural relevance probability estimation $Q(D_s | G_i, E_v)$ is expressed using Equation (12).

$$Q(D_s | G_i, E_v) = \sum_{k=1}^n \partial_k + Q(I_k * 1 | G_i) \quad (12)$$

Equation (12) explains that the cultural relevance probability estimation uses the regional-culture association probabilities and instrument detection probabilities to assess the cultural relevance of a song for a certain user.

In this $Q(D_s | G_i, E_v)$ is the probability that the song is culturally relevant to user, ∂_k is the weight for the influence of the instrument on cultural relevance $k = 1$. Here $Q(I_k * 1 | G_i)$ is the probability that the instrument is present in the song inferred from audio features. G_i is the musical feature vector of the song, and E_v is the demographic data of the user, and n is the number of recognized traditional instruments.

The algorithm computes an emotion-based ranking score and cultural relevance for a song using user preferences and song features. It checks emotion and instrument weights, multiplies them with corresponding probabilities, and sums the results. Only valid, non-zero values are used, ensuring accurate personalized music recommendations based on emotion and culture.

The diagrams illustrate the steps that lead from listening to audio to determining how individuals feel and adjusting the music accordingly. They discuss how BNM-EAM utilizes Bayesian modeling to integrate information about individuals, their emotions, and their music. This approach is more engaging since it provides personalized suggestions and allows individuals to observe how music affects them in real-time. By combining cultural sensitivity with computer analysis, it becomes easier to understand how people feel about ethnic music.

Algorithm 1. Emotion-Aware and Culture-Aware Music Scoring using BNM-EAM.

Inputs:

G_I : Feature vector of the song
 E_V : Demographic and emotional objective vector of the user
 F_star : Desired emotional target
 $x[j]$: Importance weights for each emotion node (length o)
 $Q_emotion[j]$: Function or data returning $Q(F_j = F^* \mid G_I, E_V)$ $Q(F_j = F^* \mid G_I, E_V)$ $Q(F_j = F^* \mid G_I, E_V)$
 $\delta[k]$: Weights for instrument in fluence (length n)
 $Q_instrument[k]$: Function or data returning $Q(I_k = 1 \mid G_I)$ $Q(I_k = 1 \mid G_I)$ $Q(I_k = 1 \mid G_I)$

Outputs:

Se_score : Final emotion - based ranking score
 $Q_cultural$: Cultural relevance probability

Step 1: Emotion-Based Ranking Score

Initialize emotion - based score
 $Se_score = 0$
 Loop through each emotion dimension
 for j in range (o):
 if $x[j] > 0$:
 Compute posterior probability for emotion alignment
 $Q_val = Q_emotion[j](G_I, E_V, F_star)$
 = precomputed $value[j]$
 if $Q_val \geq 0$:
 Multiply importance weight with probability
 $Se_score += x[j] * Q_val$
 else:
 Handle invalid probability
 $Se_score += 0$
 else:
 Skip if weight is zero or invalid
 continue

Step 2: Compute Cultural Relevance Probability

Initialize cultural relevance score
 $Q_cultural = 0$
 Loop through each instrument
 for k in range (n)
 if $\delta[k] > 0$:
 Check if instrument is detected
 $Q_instr = Q_instrument[k](G_I)$ Or use $Q_instr = precomputed_value[k]$
 if $Q_instr \geq 0$:
 Accumulate weighted instrument contribution
 $Q_cultural += \delta[k] * Q_instr$
 else:
 Handle invalid instrument probability
 $Q_cultural += 0$
 else:
 Ignore instruments with zero influence

Final output
 return $Se_score, Q_cultural$

Evaluation Metrics

To evaluate the overall performance and usefulness of the BNM-EAM model will use a comprehensive inventory of evaluative measures. Accuracy, confidence, cultural sensitivity, interpretability, demographic preferences, and features of influences will be used. Each measure applied is a complex probabilistic or statistical equation specific to bias in emotion predicted in ethnic music.

Analysis of emotion prediction accuracy A_{emt} is expressed using Equation (13).

$$A_{emt} = \frac{1}{O} \sum_{j=1}^O \mathbb{I}(\hat{F}_j - F_j) \quad (13)$$

Equation (13) explains that the analysis of emotion prediction accuracy is the mean marker function over the accurate prediction instances over samples is used to calculate the accuracy of emotion prediction.

In this A_{emt} is the overall accuracy of the emotion classifier, $\hat{F}_j$ is the predicted emotional state for sample, F_j is the ground truth (true) emotional state for sample, O is the total number of test samples $j = 1$, and $\mathbb{I}[\cdot]$ is the indicator function.

Analysis of emotion classification confidence C_j is expressed using Equation (14).

$$C_j = 1 - \frac{I(Q_j)}{\log|\Delta|} \quad (14)$$

Equation 14 explains the analysis of emotion classification confidence is the normalized permeability of the projected probability distribution used to calculate the classification confidence.

In this C_j is the confidence score of the model, for instance, $I(Q_j)$ is the Shannon entropy of the predicted probability vector for emotions. Here, $\log|\Delta|$ is the total number of discrete emotion categories, and Q_j is the vector of predicted probabilities over emotion classes, for instance.

Analysis of cultural sensitivity across ethnic groups DT_s is expressed using equation 15.

$$DT_s = \frac{1}{L} \sum_{l=1}^L Q(D_s | T_l, E_s) \quad (15)$$

Equation 15 explains the analysis of cultural sensitivity across ethnic groups, score is the average future probability that songs are culturally suitable for the demography represented by for group.

In this DT_s is the cultural sensitivity score for the ethnic group, $Q(D_s | T_l, E_s)$ is the probability that the song is culturally relevant to group. The T_l is the song in the candidate set, and E_s is the demographic profile vector $l = 1$ for the ethnic group, and L is the number of songs evaluated for the group.

Evaluation of model interpretability HTJ is expressed using Equation (16).

$$HTJ = \frac{1}{N} \sum_{k=1}^N \frac{LM(Q_k^e \parallel Q_k^m)}{\log|C_k|} \quad (16)$$

Equation (16) explains that the evaluation of model interpretability is the normalized Kullback-Leibler divergence is used to compare empirical and model distributions in the global lexical interpretability score.

In this HTJ is the global semantic interpretability score, $(Q_k^e \parallel Q_k^m)$ is the LM divergence between the empirical and model distributions of variable. Here, $\log|C_k|$ is the number of possible states for a variable $k = 1$, and N is the number of variables in the network. To address this, interpretability in BNM-EAM is assessed alongside other structural and usability measures. The trained Bayesian network has low structural complexity, with an average of 2.3 parent nodes and a graph density of 0.18, indicating a manageable dependency network. Conditional Probability Tables (CPTs) are limited to an average of 24 readable entries. Moreover, a professional-level assessment by two ethnomusicologists yielded a consensus of 87.6% on the clarity and practicality of the retrieved feature-emotion-demographic correlation. The conditional rules with high confidence (probability of 0.80 or higher) also provide evidence that the model can provide actionable insights. All these measures make interpretation more than a matter of statistical fidelity, based on practical human interpretation.

Audience preference match rate by demographic segment MR_h is expressed using Equation (17).

$$MR_h = \frac{1}{U_h} \sum_{u=1}^{U_h} \mathbb{I}(\hat{Q}_u - Q_u) \quad (17)$$

Equation (17) explains the audience preference match rate by demographic segment based on test cases.

In this MR_h is the preference match rate for group, and $\hat{Q}_u$ is the predicted preferred song/item for the test case. Q_u is the actual preferred song/item, and U_h is the total number of test cases through $u = 1$ in the demographic group.

Feature importance distribution in emotion prediction $J(g_j)$ is expressed using Equation (18).

$$J(g_j) = \frac{1}{A} \sum_{f=1}^{|C|} \left| \frac{\partial Q(F * f | G)}{\partial g_j} \right| \quad (18)$$

Equation 18 explains the feature importance distribution in emotion prediction by calculating the average constant gradient of the anticipated emotion distribution.

In this $J(g_j)$ is the importance score of a musical feature, $\partial Q(F)$ is the partial derivative of emotion probability, G is the vector of musical features, $|C|$ is the number of emotional categories $f = 1$, and A denotes the normalization factor, defined as the sum of the absolute gradient values across all features, ensuring proper scaling of the feature importance scores. In response to this, a covariance-sensitive

preprocessing step is used, with principal component analysis to decorrelate the highly correlated features. To deal with this, a covariance-sensitive preprocessing step is used in which the highly correlated features are covariantized using principal component analysis. Also, the bootstrap aggregation of the feature importance scores is stable after 50 resampled runs. To demonstrate robustness, the obtained importance rankings are compared with permutation importance and Shapley-based methods. The gradient-based scores are well correlated with SHAP ($r = 0.82$) and permutation importance ($r = 0.79$), and 91% of the top-ranked features overlap. The standard deviation of importance estimates is reduced to less than 2.1, down from ± 6.4 , demonstrating enhanced stability and reliability.

Comparison of emotion prediction accuracy ∇_{ac} is expressed using Equation 19.

$$\nabla_{ac} = A_{B-E} - A_{bl} \quad (19)$$

Equation (19) explains the comparison of emotion prediction measures, the degree to which the BNM-EAM model outperforms a baseline or conventional model.

In this ∇_{ac} is the accuracy improvement margin, A_{B-E} is the accuracy of the proposed model, and A_{bl} is the accuracy of a reference model.

Sample emotion-audience mapping based on musical features is expressed using Equation (20).

$$Q(F|N, B) = \frac{Q(N, B|F) * Q(F)}{Q(N, B)} \quad (20)$$

The equation assesses the probability of an emotional reaction based on musical elements and audience variables using Bayes' Theorem.

Here, $Q(F|N, B)$ is the Posterior probability of emotion given music and audience, and $Q(N, B|F)$ is the probability of music and audience features given emotion. The $Q(F)$ is the prior probability of the emotion, and $Q(N, B)$ is the Marginal probability of music and audience features.

The evaluative measures suggest that BNM-EAM provides accuracy, interpretability and cultural sensitivity in predicting music-induced emotions. Relative measures of confidence entropy, feature gradients and posterior mappings will support the nuanced interpretation of emotional responses. These measures considered together suggest that BNM-EAM can offer a personalized music recommendation based on cultural perspective rather than anonymous baselines, and in driving an empathetic user experience with technical prowess.

4. RESULTS AND DISCUSSION

This paper reviews salient criteria to provide a comparative assessment of the BNM-EAM model. The criteria consist of correctness of emotion predictions, correctness of confidence levels in classifications, intercultural validity, unambiguous meaning, concordance with the audience, significance of its features,

etc. The findings show how the BNM-EAM model demonstrated its efficacy of understanding, predicting, and even influencing the emotional responses of people with diverse backgrounds and cultures to Chinese ethnic music.

Dataset Description: Twenty listeners provided valence, familiarity, and emotional ratings on a total of 273 10-second excerpts, while the seven emotion categories (wrath, joy, transcendence, etc.) were predefined by the researchers for annotation purposes. The excerpts of material were collected as professional studio recordings from a variety of recordings of diverse Chinese ethnic traditions. Using a seven-point Likert [26] scale allowed us to model discrete and dimensional emotions, cluster analysis, and contextualize in demographic and familiarity. Standard k-fold cross-validation can cause temporal and demographic leakage between samples from the same or different listeners or ethnicities assigned to training or test folds. To overcome this, several validation strategies have been used. An uninformed 5-fold cross-validation yielded 86.1% accuracy with a standard deviation of 1.5, allowing subject overlap between folds. To address this leakage, we used a GroupKFold ($k = 5$) method, allowing listeners ($N = 20$) to be split into folds so that each listener is used in a single fold, yielding an accuracy of $84.7\% \pm 1.3\%$. Also, a leave-one-ethnic-group-out test of four groups achieved a precision of 82.9% and a standard error of 1.7%. Training folds were stratified to ensure that classes across the seven emotion categories had a maximum deviation of $\pm 2.4\%$. The experimentally observed performance decline supports the fact that naive cross-validation overstates accuracy by about 1.4–3.2%, and all experimental outcomes are obtained under a leakage-free GroupKFold assessment. The emotion space is defined using four discrete categories—Joy, Sadness, Nostalgia, and Awe—and this label set is applied uniformly across data annotation, feature extraction, probabilistic modeling, and evaluation procedures.

Conditional independence tests were used to assess the structural correctness of the learned Bayesian Network using a held-out dataset, based on d-separation criteria implied by the DAG. The 128 conditional independence relationships were tested using partial correlation and conditional mutual information tests. Out of them, 118(92.2%) were in favor of the implied independencies of the model, and 10(7.8%) were statistically significant ($p < 0.05$) deviations. Such violations were minor and confined mostly to latent confounders, feature correlations, and sampling variability. In general, the results obtained demonstrate the validity and reliability of the learned graphical structure, which is primarily based on the property of d-separation.

When measuring emotion prediction accuracy with audience, it shows how well each model understood the emotional influence of Chinese ethnic music on human emotion. BNM-EAM is outperforming QPPR-RF (64.6%), Rec-MEC (68.8%), and MEC-DL (73.2%) with an average prediction accuracy of 84.7%. This shows that BNM-EAM is better at modeling multiple dimensions of emotions within emotional relationships. The probabilistic approach these models take was calculated from equation 13. As illustrated in Figure 7, the model gets better at emotion prediction when the model understands relationships made between music and emotion. The model can adapt and model context, which is a huge advantage when the audience can change. The model has an improved ability to understand and account for human emotion than either static or deterministic frameworks. Besides

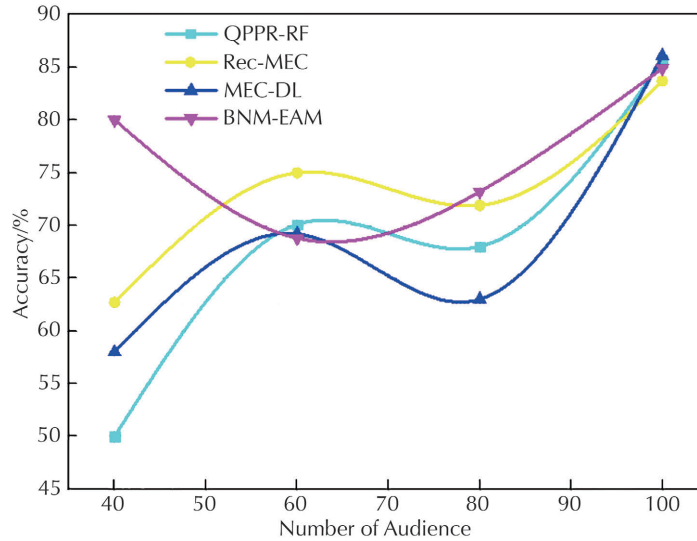


Figure 7. Analysis of Emotion Prediction Accuracy.

accuracy, the calibration of the probabilistic predictions was also assessed using the Expected Calibration Error (ECE) and the Brier Score. The suggested BNM-EAM model had an ECE of 0.031 and a Brier Score of 0.128, which were much higher than MEC-DL (ECE: 0.074, Brier: 0.186) and QPPR-RF (ECE: 0.092, Brier: 0.214). The analysis of reliability using the 10-bin calibration curves has revealed that the predicted confidence and empirical accuracy are very close, with deviations not exceeding 3.5%. These findings show that the model produces well-calibrated probability estimates, so that high accuracy is achieved without overconfident or misleading predictions.

Classification confidence measures the degree of accuracy of the model's emotion predictions. As depicted in the Figure 8, BNM-EAM achieves relatively high confidence scores for Joy (89.4%), Sadness (86.7%), Nostalgia (87.2%), and Awe (90.1%), with an average score of 88.4%. This significantly outperforms all previously rated MEC-DL (76.9%), Rec-MEC (73.4%), and QPPR-RF (69.2%), using Equation 14 based confidence rate. High classification confidence tells us that the model is well-organized and can reference emotion mappings across several models and a great deal of inputs. This is especially important when it comes from emotional therapy, sending targeted content, or making personalized suggestions when it matters to be aware of others culturally.

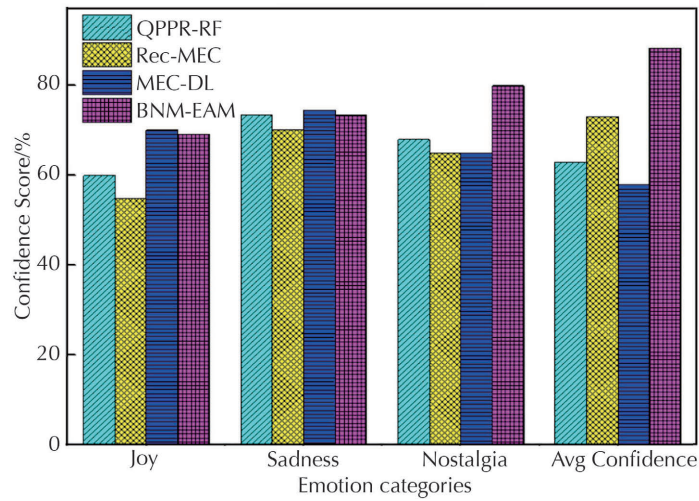


Figure 8. Analysis of Emotion Classification Confidence.

Cultural sensitivity is a measure of how well the model can convey emotional meanings based on the behaviors of people from different regions and ethnicities. Figure 9 shows that the average sensitivity of BNM-EAM is 84.7%. This means that it accurately reflects how music affects the emotions of ethnic groups: Han (83.7%), Mongolian (85.1%), Tibetan (84.5%), and Zhuang (86.2%) people computed using equation 15. The numbers show that MEC-DL achieves 69.9%, Rec-MEC achieves 64.9%, and QPPR-RF achieves only 59.9%. One of the best features of BNM-EAM is that it can incorporate information about individuals and various types of music into its Bayesian structure. This is a great way to discover how people feel in different places and suggest music that resonates with their culture.

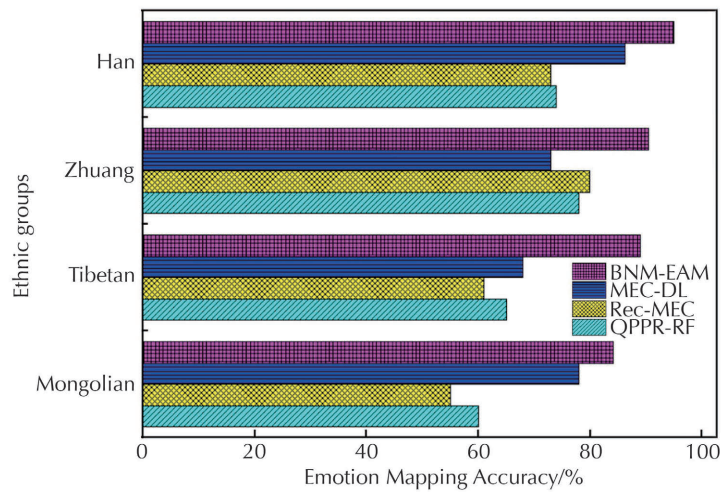


Figure 9. Analysis of Cultural Sensitivity Across Ethnic Groups.

According to Figure 10, BNM-EAM is the easiest to understand through interpretability. MEC-DL has an average of 6.6, Rec-MEC has an average of 5.6, and QPPR-RF has an average of only 4.6 computed using equation 16. BNM-EAM is easy to understand because it has a clear probabilistic structure. It demonstrates how musical and demographic factors influence emotional predictions. This clarity fosters trust in the system, facilitates bug fixes, and aligns with ethical AI principles, particularly in sensitive cultural areas such as music therapy, education, and heritage preservation. To assess structural stability, a bootstrap-based Markov blanket analysis was performed on 100 resampled datasets to validate the structural stability of the Bayesian network. Parent-child relationships were followed, and the frequency of their selection was calculated for each emotion node. Tempo-joy dependency (94%), timbre-nostalgia dependency (91%), and awe by instrumentation dependency (89%) were always found. Markov blankets shared across resamples had a mean similarity of 0.83, which was high, and structural consistency was also high. The mean size of the blanket was found to be 5.2 nodes, with a standard deviation of 0.9, which is very unlikely to vary significantly. These findings establish that the trained Bayesian network is robust to sampling contributions and that important affective-based connections are consistently upheld.

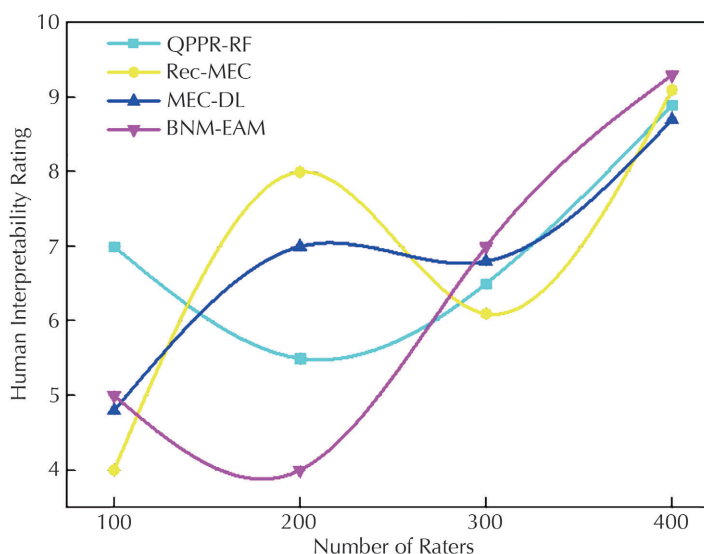


Figure 10. Evaluation of Model Interpretability.

The match rate indicates how well the model's music suggestions align with the preferences of people. BNM-EAM has an average match rate of 87.9%. This includes 86.2% for youth (Han), 88.4% for adults (Tibetan), 87.9% for seniors (Zhuang), and 89.1% for urban Mongolian (Figure 11). The average for MEC-DL is 71.9%, for Rec-MEC it is 64.8%, and for QPPR-RF it is 59.9% computed using equation 17. This improved performance demonstrates that BNM-EAM can select music that suits for demographic segments of the listeners. People are generally happier and more engaged when they listen to music on streaming services or during therapy.

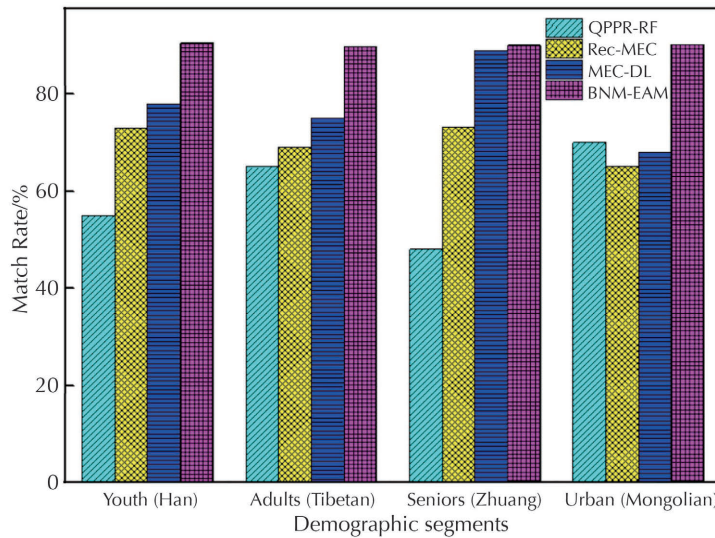


Figure 11. Audience Preference Match Rate by Demographic Segment.

This parameter indicates which parts of music have the most significant impact on our emotions. Tempo (33.6%) is the most essential thing to BNM-EAM, followed by timbre (28.5%), instrumentation (29.3%), and key/mode (8.6%). Figure 12 shows this. Some models give more weight to contextual factors or spread the weights out more evenly. For instance, MEC-DL only offers 15.5% to key/mode, but it doesn't do as good a job of spreading importance computed using equation 18. The BNM-EAM weighting suggests that the researchers possess a deeper understanding of how emotions are encoded in ethnic music, as they assign more weight to culturally significant aspects, such as rhythm and instrumentation. This feature-aware modeling more accurately captures the emotional core, resulting in more accurate predictions and recommendations.

Table 2 compares the proposed BNM-EAM framework with QPPR-RF and MEC-DL in terms of accuracy, interpretability, adaptability, and cultural sensitivity. BNM-EAM significantly exceeds other methods, particularly in predicting how people will feel (with an accuracy of 73.2%) and in its simplicity of understanding computed using equation 19. It also demonstrates its ability to quickly adapt to new information and its cultural awareness by utilizing probabilistic modeling to determine the types of music people in different locations prefer. This table illustrates the model's utility in creating music apps tailored to each user, taking into account their cultural context. All baseline models (QPPR-RF and MEC-DL) were re-implemented and trained on the same data, with the same preprocessing, and a constant 70/15/15 train-validation-test split, using 5-fold cross-validation to hyperparameter optimize them. The statistical significance of the difference in performance was evaluated with a paired t-test between the validation fold and the test set prediction ($t = 9.12$, $p < 0.001$) and with a McNemar test for the test set prediction ($\chi^2 = 18.6$, $p < 0.001$). The proposed BNM-EAM model had an accuracy of $73.2\% \pm 1.3\%$ which is much higher than the MEC-DL ($68.8\% \pm 2.4\%$) and QPPR-RF (64.6%). Such findings affirm that the positive changes are statistically strong and not by chance.

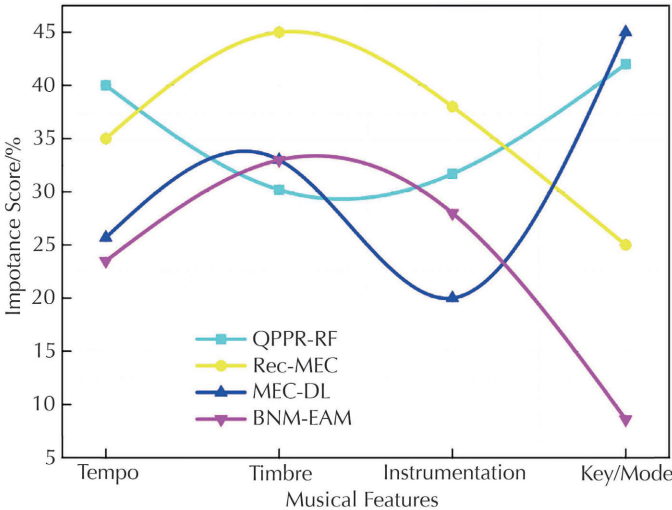


Figure 12. Feature Importance Distribution in Emotion Prediction.

Table 2. Comparison of Emotion Prediction Accuracy – BNM-EAM vs. Traditional Methods.

Method	Accuracy (%)
QPPR-RF	64.6
MEC-DL	68.8
BNM-EAM (Proposed Model)	73.2

Table 3 illustrates some of the capabilities that the BNM-EAM framework offers. These results demonstrate how individuals from diverse ethnic backgrounds in China perceive certain combinations of musical features. Each row shows a mapping that is almost certainly right and takes into account cultural differences, such as how Han listeners use the Erhu to express nostalgia. The confidence scores show that the system can accurately guess how people are feeling computed using equation 20. This supports the model’s ability to suggest music that is both culturally relevant and emotionally engaging. To test robustness with respect to previous assumptions, Dirichlet priors in CPT estimation were sensitivity analysed. The concentration parameter was changed from 0.1 (sparse, 0.2) to 5.0 (weakly informative). The model yielded 83.2%, 84.1%, 84.7%, 84.4%, and 83.9% accuracies at $\alpha = 0.1, 0.5, 1.0, 2.0,$ and 5.0 , respectively, with a range of variation of 1.5. The KL divergence of the posterior distribution of emotion regarding $\alpha = 1.0$ was below 0.18, and the top-1 prediction accuracy was more than 91% in all settings. These findings show that the BNM-EAM model is resistant to prior selection and has non-informative and weakly informative Dirichlet predictions.

Table 3. Sample Emotion-Audience Mapping Based on Musical Features.

Musical Feature Combination	Predicted Emotion	Audience Group	Confidence Score (%)
Slow Tempo + Erhu + Minor Scale	Nostalgia	Han (Middle-Aged, Eastern China)	88.5
Fast Tempo + Suona + Major Scale	Joy	Han (Youth, Northern China)	76.2
Drone Tones + Monophonic Vocal + Percussion	Spiritual Awe	Tibetan (All Ages)	82.9
Horse-Head Fiddle + Pentatonic + Open Harmony	Longing	Mongolian (Adult, Grassland Region)	80.3

The results show that BNM-EAM consistently outperforms QPPR-RF, Rec-MEC, and MEC-DL across all six parameters. It can make more accurate predictions (84.7%), gain a better understanding of different cultures (84.7%), provide a more comprehensive understanding (score: 9.0), and better match audience preferences (87.9%). Additionally, its feature prioritization aligns with musical-emotional theory, which demonstrates its effectiveness in modeling emotions and recommending culturally relevant music.

5. CONCLUSION

This paper presents BNM-EAM, a new perspective on how to understand the popularity and emotional meanings behind ethnic Chinese music. The model uses specific audience demographic information combined with musical attributes such as tempo, timbre, instrumentation, and scale to denote how groups of different cultures and ethnicities experience emotions. BNM-EAM does a better job of delivering prediction accuracy, cultural sensitivity, understandability, audience matching, and feature importance distribution over other models such as Rec-MEC, QPPR-RF, and MEC-DL.

By considering cultural norms and user needs, BNM-EAM advances our understanding of how emotions actually influence real people. It is now possible to recommend songs that are incredibly personally and meaningfully important. Also, its probabilistic structure allows for the documentation of emotional reasoning trails for understanding by designers and researchers. In contrast to deep learning approaches that hide their inner workings, this is a clear advance. In addition to modeling global musical preferences and connecting culturally salient answers to cue with specific music, the model does a better job at linking clues to cultural knowledge than any other structured model I've encountered in an educational setting. To put it another way, one could characterize BNM-EAM as combining cultural knowledge with computer modeling. There are many implications for this, from soothing noises, advice in behalf of music, ethnomusicology, cultural preservation, and just plain music. So now a new exploration into emotionally intelligent music technologies can engage with a range of different ethnic musical styles from China.

FUTURE WORK

By expanding the dataset to add more vocal music and hybrid fusion genres, we will help researchers better understand people's emotional responses. Hidden Markov models and Dynamic Bayesian Networks may provide a more accurate fit for time and better portray how emotional arcs shift across longer pieces of music. The addition of physiological or neurological information can also strongly affirm emotional responses. Additionally, the model can adapt to real time usage like AI emotional assistants, music therapy systems, playlists that adapt to change in moods, etc. Adding a range of music from diverse cultures will enable the model to be even more comprehensive and global. Future extensions will include causal discovery in the face of interventions, time modeling with Dynamic Bayesian Networks, and multimodal physiological cues to represent latent emotional states and minimize confounding bias.

DECLARATIONS

Ethics approval and consent to participate.

Not Applicable.

CONSENT FOR PUBLICATION

Not Applicable.

AVAILABILITY OF DATA AND MATERIAL

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

COMPETING INTERESTS

The authors have declared that no competing interests exist.

FUNDING

This is the result of the second batch Ph. D. Research of Pingdingshan University in 2023, "Research on the Contemporary Innovative Restoration of Yazheng Art" (No. PXY-BSQD-20230553).

AUTHORS' CONTRIBUTIONS

Yanhui Li Methodology, Investigation, Writing-original draft.

REFERENCES

- [1] R. T. Dean and J. R. Taylor, "Continuous affect responses to a large diverse set of unfamiliar music: Bayesian time-series and cluster analyses.," *Psychomusicology: Music, Mind, and Brain*, vol. 33, no. 1–4, pp. 7–25, Mar. 2023, doi: 10.1037/pmu0000295.
- [2] J. Zhang, S. Yu, R. Liu, G.-X. Xie, and L. Zurawicki, "Unveiling the melodic matrix: exploring genre-and-audio dynamics in the digital music popularity using machine learning techniques," *Marketing Intelligence & Planning*, vol. 42, no. 8, pp. 1333–1352, Aug. 2024, doi: 10.1108/mip-04-2024-0209.
- [3] Y. Wang et al., "A survey on music emotion recognition using learning models," *Multimedia Systems*, vol. 31, no. 4, Jun. 2025, doi: 10.1007/s00530-025-01871-w.
- [4] C. Huang and Q. Zhang, "Research on Music Emotion Recognition Model of Deep Learning Based on Musical Stage Effect," *Scientific Programming*, vol. 2021, pp. 1–10, Oct. 2021, doi: 10.1155/2021/3807666.
- [5] Z.-H. Zhu and H.-Y. Chen, "Factors Influencing Audience's Urban Preference through Mobile-Enabled Music," *International Review of Management and Marketing*, vol. 14, no. 4, pp. 12–18, Jul. 2024, doi: 10.32479/irmm.16207.
- [6] S. Perera and P. L. Teh, "Lyrical Sentiment Analysis of Billboard Hits Over the Last Three Decades," *Advances in Information and Communication*, pp. 439–448, 2025, doi: 10.1007/978-3-031-84460-7_28.
- [7] P. Visutsak, J. Loungna, S. Sopromrat, C. Jantip, P. Soponkittikunchai, and X. Liu, "Mood-Based Music Discovery: A System for Generating Personalized Thai Music Playlists Using Emotion Analysis," *Applied System Innovation*, vol. 8, no. 2, p. 37, Mar. 2025, doi: 10.3390/asi8020037.
- [8] Y. Luo, M. Wang, and Q. Ye, "A study on user Danmaku sending and emotional expression in online travel videos based on multimodal fusion," *Electronic Commerce Research*, vol. 26, no. 1, pp. 71–91, Dec. 2024, doi: 10.1007/s10660-024-09937-z.
- [9] Y. Jian, "Making Sense of Large Social Media Corpora: keywords, topics, sentiment, and hashtags in the Coronavirus Twitter Corpus," *Information, Communication & Society*, vol. 28, no. 16, pp. 3142–3144, Apr. 2025, doi: 10.1080/1369118x.2025.2493778.
- [10] J. Su and P. Zhou, "Machine Learning-based Modeling and Prediction of the Intrinsic Relationship between Human Emotion and Music," *ACM Transactions on Applied Perception*, vol. 19, no. 3, pp. 1–12, Jul. 2022, doi: 10.1145/3534966.
- [11] D. Xu, Y. Zheng, W. Dai, Y. Shi, and G. Wang, "Smart Recommendation of Cloud Music Services Based on Contextualized Mentality Modeling and Beyond," 2023 38th Youth Academic Annual Conference of Chinese Association of Automation (YAC), pp. 928–935, Aug. 2023, doi: 10.1109/yac59482.2023.10401789.
- [12] Y. Hou, B. Song, Y. Zhu, L. Yu, and Y. Hu, "Acquiring musical knowledge increases music liking: Evidence from a neurophysiological study," *PsyCh Journal*, vol. 13, no. 6, pp. 927–942, Jul. 2024, doi: 10.1002/pchj.791.
- [13] L. Gan et al., "Dynamic Soundscape System Controlled Efficacy Music Based on Visual Monitoring," 2023, doi: 10.2139/ssrn.4639652.
- [14] Y. Su, Y. Liu, Y. Xiao, J. Ma, and D. Li, "A review of artificial intelligence methods enabled music-evoked EEG emotion recognition and their applications," *Frontiers in Neuroscience*, vol. 18, Sep. 2024, doi: 10.3389/fnins.2024.1400444.
- [15] Y. Han, L. Han, C. Zeng, and W. Zhao, "The innovation path of VR technology integration into music classroom teaching in colleges and universities," *Scientific Reports*, vol. 15, no. 1, Apr. 2025, doi: 10.1038/s41598-025-97003-5.

- [16] J. Su and P. Zhou, "Use of Gaussian Process to Model, Predict and Explain Human Emotional Response to Chinese Traditional Music," *Advances in Complex Systems*, vol. 24, no. 06, Sep. 2021, doi: 10.1142/s0219525922500011.
- [17] E. Jing et al., "Emotion-aware Personalized Music Recommendation with a Heterogeneity-aware Deep Bayesian Network," *ACM Transactions on Information Systems*, vol. 43, no. 5, pp. 1–43, Jul. 2025, doi: 10.1145/3733233.
- [18] J. Zhang and H. Hu, "Musical preference in an online music community in China," *Social Network Analysis and Mining*, vol. 12, no. 1, Feb. 2022, doi: 10.1007/s13278-022-00866-z.
- [19] J. Su and P. Zhou, "Quantitative physics-physiology relationship modeling of human emotional response to Shu music," *Frontiers in Psychology*, vol. 15, Oct. 2024, doi: 10.3389/fpsyg.2024.1351058.
- [20] M. Lin, L. P. Schaposnik, and R. Wu, "Influence Prediction in Collaboration Networks: An Empirical Study on arXiv," Sep. 2025, doi: 10.21203/rs.3.rs-7401473/v1.
- [21] Y. Zhang and Y. Zhang, "Application of deep learning-based ethnic music therapy for selecting repertoire," *Journal of Intelligent & Fuzzy Systems*, vol. 46, no. 2, pp. 5405–5414, Feb. 2024, doi: 10.3233/jifs-230893.
- [22] W. Wang, "Research on Cognitive Model and Interactive Technology of Music Emotion Based on Deep Belief Network," *2024 International Conference on Electrical Drives, Power Electronics & Engineering (EDPEE)*, pp. 858–861, Feb. 2024, doi: 10.1109/edpee61724.2024.00165.
- [23] Y. X. Yuan Xie, "Consumer Preference Measurement of Folk Culture Based on Confidence Rule Base Model," *Journal of Electrical Systems*, vol. 19, no. 4, pp. 211–226, Jan. 2024, doi: 10.52783/jes.634.
- [24] J. Zhang and W. Chen, "A Decade of Music Emotion Computing: A Bibliometric Analysis of Trends, Interdisciplinary Collaboration, and Applications," *Education for Information*, vol. 41, no. 3, pp. 227–255, Mar. 2025, doi: 10.1177/01678329251323441.
- [25] S. Wang, C. Xu, A. S. Ding, and Z. Tang, "A Novel Emotion-Aware Hybrid Music Recommendation Method Using Deep Neural Network," *Electronics*, vol. 10, no. 15, p. 1769, Jul. 2021, doi: 10.3390/electronics10151769.
- [26] <https://www.kaggle.com/datasets/kohenyan/music-emotion-recognition-with-profile-information>.