https://doi.org/10.4314/ajlais.e35115

Information Retrieval in Digital Sound Archives and Libraries: Preservation of the Documents of Xinjiang-Style Guzheng Music through Digital Library Curation

Junyan Wang*
Faculty of music, Silpakorn University, Thailand
409795248@qq.com

and

Junzhu Zhao
Faculty of Engineering and Physical Sciences, University of Leeds, United Kingdom
zjz15357116052@163.com

Abstract

The preservation and accessibility of cultural artefacts are increasingly reliant on the integration of technology into archival practices. Digital libraries play an essential role in safeguarding musical heritage, offering scalable storage, improved accessibility, and long-term sustainability. Sound archives face degradation due to inadequate organisational systems and limited public usage. To address this, the proposed Information Retrieval Framework (IRF) aims to establish a new system for the protection of Guzheng musical heritage in Xinjiang by advancing audio curation techniques. This framework employs a Hybrid Cascaded Convolutional Gated Neural Network (HCC-GNN), which integrates Cascaded Convolutional Neural Networks (CCNNs) to extract robust short-term features from raw waveforms and spectrograms. These features are subsequently processed through a Gated Recurrent Unit (GRU) and Recurrent Neural Network (RNN) to capture long-term temporal patterns in music recordings. This enables the effective tagging and retrieval of archived audio content, including Xinjiang-style Guzheng performances, even amidst background noise and repeated acoustic events. An evaluation was carried out using Xinjiang-style Guzheng recordings from digital archives. Prior sound normalisation through Wiener filtering enhanced audio clarity, while Wavelet Transform and data augmentation optimised the extraction process, increasing diversity and improving system reliability. A retrieval-optimised network then processed these refined features, enabling users to swiftly obtain relevant information from extensive digital audio databases. The HCC-GNN model outperformed conventional methods, achieving an accuracy of 97.5%, Precision@10 of 97%, Recall@10 of 95.8%, Mean Average Precision (MAP) of 96%, and an F1 Score of 95.2%. These results promote cultural sustainability and enhance accessibility for scholars.

Keywords: Digital Sound Archives, Xinjiang-Style Guzheng Music, Digital Library, Hybrid Cascaded Convolutional Gated Neural Network (HCC-GNN), Audio-Visual Retrieval, Information Retrieval Framework (IRF).

Introduction

Digital technologies are increasingly pivotal in the preservation of intangible cultural heritage (Rossetto et al., 2023). In musicology, specifically, digital sound archives and libraries serve as essential tools for the preservation of musical traditions that are under threat from the forces of globalization, modernization, and shifting social and cultural contexts. While digitisation has advanced considerably, challenges persist in creating effective means of accessing, discovering, and interacting with sound resources in digital formats (Patel et al., 2024). A crucial factor in optimising the utility of digital sound archives is the ability to extract relevant information (He, 2022). Retrieval systems must efficiently handle large volumes of data while also considering the contextual and cultural significance embedded in musical documents (Salosaari, 2022). This is especially important when preserving regional music traditions that may lack standardised metadata or comprehensive archival documentation (Ni, 2023).

The preservation of Guzheng music from Xinjiang exemplifies these challenges, as this well-known Chinese instrument tradition originates from ethnic groups in the Xinjiang region (Lee et al., 2025). Traditional performance practices, cultural modes, and rhythmic structures of Guzheng music remain vulnerable due to inadequate documentation and distribution mechanisms (Wu et al., 2023). Preservation methods reliant on physical recordings and oral transmission are insufficient in addressing the threats posed by time and technological changes (Amiri et al., 2022). The establishment of digital library curation combined with modern information retrieval offers a more sustainable solution for preserving Xinjiang-style Guzheng music (Buragohain et al., 2024).

However, this requires the development of culturally appropriate metadata schemas, the use of reliable indexing systems, and the creation of retrieval mechanisms that account for the non-standard features of traditional music (Seekhunlio et al., 2024). Addressing these issues is essential not only for the preservation of Xinjiang-style Guzheng music but also for the broader challenge of digitally safeguarding musical diversity (Ostermann et al., 2023). The objective of this research is to develop a hybrid Deep Learning-based IRF incorporating an HCC-GNN. The framework integrates cascaded convolutional and gated recurrent structures to preserve, curate, and retrieve Xinjiang-style Guzheng music from digital sound archives, thereby enhancing the accessibility and sustainability of cultural heritage.

Organization of the Research: Section 2 provides a review of related literature on digital sound archives and deep learning (DL) methods. Section 3 outlines the proposed methodology, covering pre-processing, feature extraction, data augmentation, and the HCC-GNN model. Section 4 presents the experimental results and offers a performance analysis. Section 5 concludes the study and proposes directions for future research.

Related Works

Malin et al. (2022) introduced a new dataset of 3,000 simulated music tracks, which includes annotations based on authentic instrument samples. This dataset is valuable for evaluating and optimising computers for various tasks, such as music segmentation, instrument recognition, and onset detection, enabling the use of DSP effects with Deep Neural Networks (DNN). Lee and Hu (2022) presented the KMDMP, a community-driven initiative to digitise ancient handwritten manuscripts, advancing Music Information Retrieval (MIR) and computational ethnomusicology. The project digitised 1,300 songs from diverse styles and demographics, promoting a dynamic musical community and improving the maintenance of cultural heritage records. The increasing use of network resources as a public knowledge medium has spurred music retrieval systems to build efficient multimedia databases.

Six et al. (2023) proposed a fuzzy search reinforcement learning technique to enhance retrieval performance, considering users' song knowledge and the need for a fuzzy search architecture in music search. Bonjack and Trujillo (2023) investigated duplication identification in large music collections, highlighting its importance for improving listening experiences, saving storage space, and verifying metadata accuracy. Their case studies revealed considerable duplication of material within these collections. Gorgoglione et al. (2023) observed that AI is transforming the music industry, with ChatGPT improving metadata clean-up tools, media storage integration, and collection interactions. Four areas are currently under investigation: enhancing song discovery platforms, improving metadata production, and integrating song ownership into commercial discovery systems.

Innovation plays a significant role in fostering human creativity within the music industry, providing a competitive advantage. Digital technologies and streaming platforms enable continuous data collection, enhancing the understanding of the complexities of music production in an ever-expanding industry (Gabbolini and Bridge, 2024). Music playlists, including their automatic creation, continuation, tagging, and captioning, have been extensively studied within MIR research. A paradigm shift from tangible items to streaming audio is achieved by integrating various MIR studies into playlists (Zhu et al., 2022). Daga et al. (2023) proposed an open-source ElasticSearch-based framework combining music characterisation, feature extraction, and indexing for scalable content-based search in large digital music collections. Egan (2022) explored the ecosystem of the EC-funded Polifonia project, a vital tool for organising collaboration and software development in Research and Action, fostering new research methods, and enhancing bibliography and data retrieval practices. Wu and Pan (2025) discussed digital initiatives using Linked Data to link online music resources, citing the 2019 initiative Connections in Noise, which evaluated the potential and challenges of improving metadata from recorded archives of Irish ancestral music.

The PICCH investigation identified users of cultural heritage archives, revealing their information needs, search strategies, and challenges. Data was gathered via surveys and interviews, targeting researchers and archivists. Borlund et al. (2024) proposed creating an intelligent system using MIR and AI techniques to recommend music selection and matching for dance compositions. CNNs were employed to extract features, thereby improving model performance and accuracy, with innovative methods significantly enhancing the rhythmic process. Zhou et al. (2024) highlighted how deep learning drives archival management, transitioning from single to multimodal forms, stressing real-world applications and proposing a new model for effective archiving. Roa Dabike et al. (2024) trained in Cadenza's machine learning problem for harmonising classical music ensembles, using the Cadenza Woodwind Dataset, which includes synthesized sounds for woodwind ensembles, virtual instruments, and metadata for music retrieval. Several studies have proposed MIR and EDM, which have greatly benefited from deep learning techniques. A Python framework for automated EDM audio production was introduced, providing a practical method for obtaining labelled data and creating high-quality templates. Experiments showed that mixes could match original reference tracks, demonstrating the power of deep learning in music (Hajarian et al., 2025).

Fantozzi et al. (2017) emphasised that enhanced music storage requires intelligent retrieval, where pre-processing, word segmentation, and deep learning enable efficient feature extraction and resource identification. AI-powered neural networks (NNs) have been used in music therapy for music generation, mimicking human data collection to identify correlations and traits. AI-based Big Data (AI-BD) tools in the music industry achieved a 97.8% performance rate (Sun, 2024). Sun and Sundarasekar (2023) analysed 478 Nigerian traditional songs from five genres in the ORIN dataset, finding that the XGBoost classifier outperformed other methods with an accuracy of 81.94% and recall of 84.57%. Folorunso et al. (2022) proposed integrating multimodal data to evaluate video soundtracks and determine artistically appropriate music. Their technique extracts a comprehensive multimodal feature library, builds a database, and employs a classifier to differentiate between authentic and false video soundtracks. Discussions were held on future task orientations and improvements. Shamsi and Sindhu (2025) assessed the rapid growth of machine learning (ML) and its application in signal processing to extract valuable information from music. Researchers have worked on computer music and voice detection since the 1940s, developing machine listening systems for diverse sounds and categorisation tasks.

Information Retrieval Framework (IRF)

The IRF for Xinjiang-style Guzheng music utilises a HCC-GNN. The pre-processing stage incorporates Wiener filtering and amplitude normalisation, followed by data augmentation techniques such as time stretching, pitch shifting, and the addition of background noise. Feature extraction is performed using wavelet transform to capture time-frequency components. CCNNs are employed to extract short-term features, while Gated Recurrent Units (GRUs) in a RNN model long-term dependencies. Figure 1 illustrates the IRF workflow, facilitating the accurate retrieval and curation of Guzheng music from audio archives.

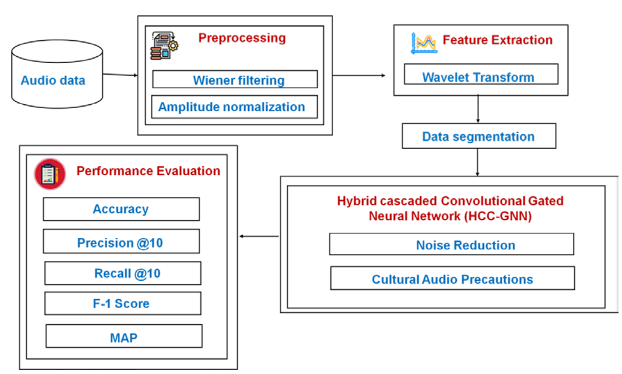

Figure 1: Information Retrieval Framework (IRF) Flow.

Dataset

The Xinjiang-Style Guzheng Music Preservation Dataset from Kaggle (https://www.kaggle.com/

datasets/ziya07/xinjiang-style-guzheng-music-preservation-dataset/data) comprises high-fidelity recordings of traditional performances captured under various conditions. It includes raw audio, spectrograms, and MFCC features, all enhanced with Wiener filtering and amplitude normalisation. Additionally, the dataset provides comprehensive metadata on performers, cultural context, and recording locations, thereby supporting research in audio preservation and cultural heritage studies.

Data Pre-Processing

Pre-processing improves the quality of Xinjiang-style Guzheng recordings by applying Wiener filtering to reduce noise and amplitude normalisation to ensure consistent loudness across the collection. These processes enhance clarity, consistency, and overall sound quality in the digital sound archives.

Winer Filter

Wiener Filtering (WF) enhances the quality of Xinjiang-style Guzheng music in archives by reducing environmental noise while preserving the essential musical features. It provides an optimal statistical solution for estimating the desired signal amid additive noise. The filter's objective is to minimise the MSE between the output of the filter and the original clean signal. Mathematically, w_i represents the noise signal, and z_i is the observed noisy Guzheng recording. The Wiener Filter produces an estimated n $\hat{m}_l$ of the noise component, which is subtracted from the noisy signal to yield the enhanced output f_l using Equation (1).

$$f_l = z_l - \hat{m}_l = z_l - \sum_{i=0}^{M-1} x(i).w_{l-i} \quad (1)$$

Where $\hat{m}_l$ is the convolution of the filter coefficient vector x and the noise signal vector w_l. The input signal vector and the filter weight vector are defined as Equations (2 and 3).

$$w_l = \left[w_l w_{l-1} w_{l-2} w_{l-3} \dots\dots\dots\dots\dots\dots, w_{l-(M-1)}\right]^S \quad (2)$$

$$x = [x(0)\ x(1)x(2)x(3) \dots\dots\dots\dots x(M-1)]^S \quad (3)$$

The filter coefficients are computed to minimize the MSE, denoted by ξ, given by Equation (4):

$$\xi = \in \{f_l^2\} = \in \{f_l^2\} - 2x^S Q_{Z\,W} + x^S Q_{W\,W} x \quad (4)$$

Where Q_w presented the cross-correlation is a vector between the desired clean signal and the noisy observation, and Q_{ww} is the autocorrelation matrix of the noise. By solving this optimization, the Wiener filter effectively reduces background noise without significantly distorting the musical characteristics of the Guzheng recordings. The enhanced recordings are then subjected to feature extraction processes, ensuring that the curated digital sound archive preserves both authenticity and high audio fidelity. This process allows for the retention of intricate musical details while enhancing the overall clarity of the audio, which is crucial for accurate preservation and retrieval of cultural heritage.

Amplitude Normalization for Audio

Amplitude normalization is employed to equalize the audio level of each recording of Xinjiang-style Guzheng music, addressing issues arising from varying recording conditions. Variations in amplitude can negatively impact feature extraction and retrieval accuracy by introducing inconsistencies in the dataset. By normalizing the amplitude, the system ensures that all recordings are at a consistent volume level, thereby improving the reliability of subsequent processing stages and enhancing the precision of retrieval tasks. This step is essential for maintaining the integrity and comparability of the data across different recordings. To mitigate this, each raw audio waveform $y(t)$ is scaled relative to its maximum amplitude A_{max}, using the normalization factor α, defined as Equations (5 and 6):

$$\alpha = \frac{A_{target}}{A_{max}} \quad (5)$$

$$y_{normalization}(t) = \alpha \cdot y(t) \quad (6)$$

where A_{target} is the desired peak amplitude. Alternatively, Root Mean Square (RMS) standardization is utilized to ensure consistent perceived loudness, as shown in Equation (7):

$$\alpha = \frac{RMS_{Tstget}}{RMS(y(t))} \quad (7)$$

This process preserves the musical dynamics while correcting amplitude bias, ensuring that all recordings exhibit consistent loudness profiles. Proper normalization enhances the stability and reliability of feature extraction processes, thereby contributing to the effectiveness of the IRF as a tool for digital sound archiving. It ensures that loudness variations do not interfere with algorithmic analysis, ultimately supporting more accurate retrieval and classification within the archive.

Feature Extraction using the Wavelet Transform (WT)

The approach utilises multiple frequency bands at various levels to analyse audio signals, akin to image processing techniques. The input waveforms are decomposed to extract wavelet coefficients, which are instrumental in identifying musical characteristics based on frequency content. In this WT methodology, the audio signal is partitioned into four distinct sub-bands.

- Low–Low (L–L)
- Low–High (L–H)
- High–High (H–H)
- High–Low (H–L)

The decomposition effectively captures both low- and high-frequency components of the Guzheng music signal. Lower approximation levels (from L to L1 and L to L2) represent the instrument's sustained, low-pitched tones, while higher approximation levels (from H to L1, H to H1, and H to H2) highlight subtle textures and transient elements. These wavelet-derived features reveal time-frequency patterns essential for distinguishing among diverse musical expressions. The components extracted are formalised through Equation (8):

$$E_{cxs}(t) = \begin{cases} FB_{j,i} = \Sigma e(t) g(t) * j(t - 2ji) \\ FB_{j,i} = \Sigma e(t) J(t) * j(t - 2ji \end{cases} \quad (8)$$

Where $FB_{j,i}$ represents the wavelet coefficients, $g(t)$ and $J(t)$ are high and low-level basis functions and j and i denote the scale and translation factors. These coefficients enable robust identification of features for retrieving Xinjiang-style Guzheng music, as they encapsulate essential musical attributes across multiple time-frequency resolutions.

Data Augmentation

To improve the model's generalisation capacity and mitigate overfitting due to the limited dataset size, various data augmentation techniques were applied to the Xinjiang-style Guzheng music dataset. Time stretching by ±10% was implemented to alter the tempo while preserving pitch, and pitch shifting by ±2 semitones was introduced to emulate tuning variations commonly observed in live performances. Dynamic range compression was used to reduce loudness disparities, enhancing resilience to dynamic fluctuations. Additionally, background noise was introduced at multiple signal-to-noise ratios to replicate realistic acoustic environments. Random cropping of audio segments was employed to introduce temporal variability. These augmentation strategies expanded the diversity of the training data while preserving essential musical attributes, thereby supporting accurate retrieval of a wide range of Guzheng performances under varied conditions.

Xinjiang Style Guzheng Music in Retrieval in Digital Sound using Hybrid Cascaded Convolutional Gated Neural Network (HCC-GNN)

The proposed HCC-GNN model enhances retrieval processes within digital sound archives containing Xinjiang-style Guzheng music. The short-term extraction of robust features from raw waveforms and spectrograms is performed by CCNNs, which are subsequently followed by GRUs and RNNs to model long-term temporal patterns. This architectural design effectively mitigates background noise and acoustic variability, thereby enabling precise tagging of traditional Guzheng performances. By delivering superior performance, the HCC-GNN advances conventional digital audio preservation practices and promotes the sustainable management of culturally significant audio repositories.

Cascaded Convolutional Neural Networks (CCNNs)

CCNNs operate as an acoustic mechanism for extracting robust short-term features from Xinjiang-style Guzheng performances, using both audio waveforms and spectrograms. These convolutional networks function sequentially, with each stage processing increasingly detailed representations of the input signal. The initial coarse network captures fundamental musical elements such as pitch, timbre, and texture, while subsequent layers identify more intricate acoustic features distinctive to Guzheng music. This cascading configuration gradually expands the receptive field, enabling the model to effectively capture both local and broader temporal patterns. Such hierarchical feature extraction is critical for preserving the tonal richness and subtle stylistic nuances characteristic of traditional Guzheng recordings. By learning complex short-term representations, CCNNs establish a solid basis for enhancing the precision and reliability of information retrieval tasks in extensive digital sound archives. There are a total of three systems that must be trained through the CasCNN training process: the, Net_{mod}, Net_{PSK} and Net_{QAM}. Both the first and second block systems must be used on every sample. It creates three distinct datasets out of the offline (Ω_w, z_{mod}, z_n) data: (Ω_w, z_{mod}), (Ω_{wPSK}, z_{nPSK}) and (Ω_{wQAM}, z_{nQAM}). Training input sample i^{th} is represented as ($w^{(i)}$, $z_{mod}^{(j)}$, z_n) where i is the index of the i^{th} labeled sample using Equation (9).

$$g_{\theta_1}(w^{(i)}) = [g_{\theta_1|G_l}(w^{(i)})], l = 1,2 \dots \dots \quad (9)$$

The Net mod with parameters denoted by $g_{\theta 1}|(\bullet)$ and the transformation denoted by $g_{\theta 1}|_{Gl}$ (•) where G_l is a hypothesis, the feature recovered by Net mod for the given hypothesis Hk is denoted by $g_{\theta 1}|_{Gl}$ ($w^{(i)}$) using Equations (10 -12)

$$G_0: O\left(z_{mod}^{(j)} = 0 | w^{(i)}; \theta_1\right) = g_{\theta_1|G_0}(w^{(i)}) \quad (10)$$

$$g_{\theta_1|G_0}(w^{(i)}) + g_{\theta_1|G_1}(w^{(i)}) = 1 \quad (11)$$

where $O(\bullet)$ stands for the probability given the l^{th} hypothesis as a condition.

$$\theta_1^* = argmaxO(z_{mod}|\Omega_w; \theta_1) \quad (12)$$

The parameters are obtained after training Net_{PSK} and Net_{QAM} using (Ω_w; z_n) and (θ_2),$I(\theta_3)$.

Temporal Modelling Using Gated Recurrent Units and Recurrent Neural Networks

To capture short-term acoustic features such as pitch and timbre from Xinjiang-style Guzheng recordings, the CNN layers are employed at the initial stage of the architecture. These extracted features are subsequently passed to GRU-RNN layers, which are responsible for modelling long-term temporal dependencies—an essential aspect for accurately retrieving and preserving evolving musical structures over time. To capture these long-term temporal patterns, a RNN framework enhanced with GRUs is utilised. RNNs are well-suited for audio sequence processing due to their feedback connections that enable them to model temporal relationships within sequential data. However, standard RNNs suffer from training instabilities caused by vanishing and exploding gradients when applied to extended time sequences.

To address this, GRUs are introduced into the architecture. GRUs simplify the gating process in comparison to Long Short-Term Memory (LSTM) units by utilising only two gates—an update gate and a reset gate—which regulate the transmission of information. This gated mechanism enables the network to retain salient acoustic information while discarding irrelevant or noisy inputs, which is crucial for accurately modelling the nuanced evolution of traditional Guzheng performances. GRU-RNN units are positioned after the CCNN feature extraction stage, enabling the system to jointly track both short-term and long-term musical developments. This dual capability significantly enhances the precision of digital sound retrieval from archival collections. Unlike LSTMs, GRUs do not require a separate memory cell; instead, memory operations are managed internally via the update gate (denoted as y) and the reset gate (denoted as s). The update gate determines the extent to which new data enters memory, while the reset gate governs the removal of outdated information. The mathematical operations of GRUs are formally defined in Equations (13) to (16).

$$y_s = \sigma(X_y[g_{s-1}, w_s] + a_y) \quad (13)$$

$$z_s = \sigma(X_q[g_{s-1}, w_s] + a_q) \quad (14)$$

$$\tilde{g}_s = \tanh(X_g[q_s \circ g_{s-1}, w_s] + a_g) \quad (15)$$

$$g_s = (1 - y_s) \circ g_{s-1} + g_s \circ \tilde{g}_s \quad (16)$$

Here, y_s acts as the update gate adaptable to how much past data to retain, z_s as the reset gate regulatory how much past information to disremember, and is the candidate's hidden state. This mechanism enables the model to concentrate on musically salient features while disregarding irrelevant fluctuations, thereby enhancing the capacity for chronological modelling even in the presence of background noise. As such, the GRU is instrumental in reinforcing the robustness and precision of the music retrieval framework. The HCC-GNN framework, which integrates CNN and GRU layers, is illustrated through pseudocode that outlines the process of extracting reliable spatial-temporal features to support accurate retrieval and classification of Xinjiang-style Guzheng music recordings. Pseudocode 1 presents the operational sequence of the HCC-GNN model.

Pseudo Code 1: HCC-GNN

```
Initialize Conv1 with 32 filters, kernel size 3x3, ReLU activation
Initialize Conv2 with 64 filters, kernel size 3x3, ReLU activation
Initialize MaxPooling with pool size 2x2
Initialize the GRU layer with 128 hidden units
Initialize Fully Connected (FC1) layer with 256 neurons, ReLU activation
Initialize Output layer with Softmax activation (for classification)
defforward_pass(input_data):
    x = Conv1(input_data)  # Apply first convolution (32 filters, 3x3)
    x = Conv2(x)  # Apply second convolution (64 filters, 3x3)
    x = MaxPooling(x)  # Apply pooling to reduce dimensions
    x = GRU(x, hidden_units=128)  # Apply Gated Recurrent Unit (GRU)
    x = FullyConnected(x, neurons=256, activation=ReLU)  # FC layer
output = Softmax(x)  # Final classification output
return output
for epoch in range(1, 21):  # Train for 20 epochs
    For batch in dataset_batches:
input_data, target = batch
prediction = forward_pass(input_data)
loss = CrossEntropyLoss(prediction, target)
Backpropagate(loss)
UpdateWeights(learning_rate=0.001)
test_accuracy = EvaluateModel(test_data)
print("Final Test Accuracy:", test_accuracy)
```

Results and Discussion

The experiments were conducted on a system comprising an Intel Core i7 processor, 32GB of RAM, and an NVIDIA RTX 3080 GPU. The software environment included Python 3.9, TensorFlow, and Librosa for DL and audio signal processing. The ablation studies for the HCC-GNN framework are summarised in Table 1. The extracted features are initially processed by CNN layers, where convolutional kernels learn localized time-frequency patterns, such as spectral envelopes, pitch modulations, and timbral textures. The resulting feature maps are reshaped and then passed into a stacked GRU-based RNN. This RNN structure, enhanced by GRU cells, captures long-term temporal dependencies and sequential dynamics across the audio frames, as graphically presented below. This approach facilitates robust modelling of the evolving musical structures, thereby improving retrieval precision and preserving intricate patterns in Xinjiang-style Guzheng music. The absence of Wiener Filtering and Amplitude Normalization leads to diminished performance. When all modules are activated, the complete setup yields the best results, confirming the efficacy of the proposed hybrid structure for reliable feature extraction and retrieval, even in noisy conditions.

Table 1: Ablation Study of HCC-GNN for Xinjiang-Style Guzheng Music Retrieval.

CCNN (w/o)	GRU (w/o)	RNN (w/o)	Wiener Filtering	Amplitude Normalization	Precision@10	Recall@10	MAP
✗	✓	✓	✓	✓	0.79	0.76	0.78
✓	✗	✓	✓	✓	0.85	0.81	0.83
✓	✓	✗	✓	✓	0.87	0.83	0.85
✓	✓	✓	✗	✓	0.78	0.74	0.77
✓	✓	✓	✓	✗	0.81	0.77	0.79
✓	✓	✓	✓	✓	**97**	**95.8**	**96**

The comparative performance analysis of several models, including CNN, RNN, GRU, RNN+GRU, and the proposed HCC-GNN, is presented in Table 2 and Figure 2. Key evaluation metrics, such as MAP, Accuracy, F1-Score, Precision@10, and Recall@10, are reported. The HCC-GNN achieves the highest performance, with an accuracy of 97.5%, F1-Score of 95.2%, recall of 95.8%, and MAP of 96%. These results demonstrate that the HCC-GNN outperforms all baseline models, indicating its superior ability to efficiently capture both temporal and spatial data for improved retrieval and classification of Xinjiang-style Guzheng music.

Table 2: Performance Comparison of Different Models on Xinjiang-Style Guzheng Music Retrieval Tasks.

Model	Accuracy (%)	F1-Score	Precision@10	Recall@10	MAP
CNN	83.2	80.2	76	74	71
RNN	81.5	79	73	72	69
GRU	84.1	81.3	78	75	73
RNN + GRU	85.3	82.5	79	76	74
HCC-GNN (Hybrid)	**97.5**	**95.2**	**97**	**95.8**	**96**

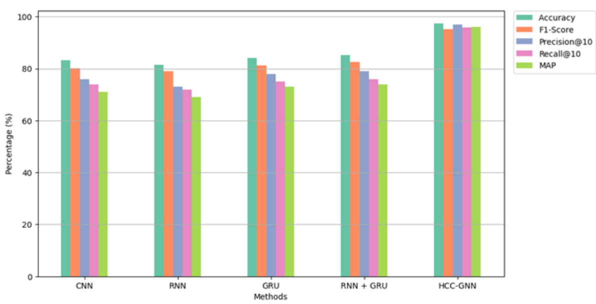

Figure 2: Performance Comparison of Retrieval Models for Xinjiang-Style Guzheng Music.

The performance metrics assessed in this study are presented in Table 3 and Figure 3, including Retrieval Accuracy, Noise Robustness, Feature Extraction Time, MAP, Retrieval Time per Query, Signal-to-Distortion Ratio (SDR), and Structural Similarity Index (SSIM). The HCC-GNN model demonstrates substantial improvements across all metrics, particularly in its noise robustness (21.5 dB) and retrieval time (87 ms). The high SDR (22.7 dB) and SSIM (0.91) scores further highlight the model's ability to preserve signal quality and structural similarity, even under noisy conditions. These results confirm the HCC-GNN's suitability as a reliable framework for real-world archival retrieval systems.

Table 3: System-Level Performance Metrics for Various Models.

Performance Metric	CNN	RNN	GRU	RNN + GRU	HCC-GNN
Retrieval Accuracy (%)	78.2	79.6	81.3	83.5	**89.7**
Noise Robustness (SNR in dB)	12.8	13.5	14.2	15.2	**21.5**
Feature Extraction Time (Seconds)	2.8	2.5	2.3	2.1	**1.3**
Mean Average Precision (MAP)	0.65	0.67	0.69	0.71	**0.80**
Retrieval Time per Query (ms)	125	118	112	105	**87**
Signal-to-Distortion Ratio (SDR dB)	13.5	14.0	14.8	16.0	**22.7**
Structural Similarity Index (SSIM)	0.81	0.83	0.84	0.86	**0.91**

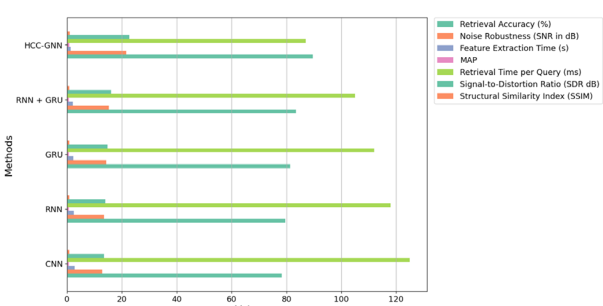

Figure 3: Retrieval System Robustness and Efficiency Analysis under Noisy Conditions.

Table 4 and Figure 4 present the proposed system's performance in preserving cultural features and providing adequate audio curation compared to baseline methods in audio preservation and curation research. The evaluated metrics include Preservation Quality, Noise Reduction Efficiency, Metadata Accuracy, Retrieval Accuracy in culturally meaningful queries, Background Noise Tolerance, Authenticity Detection, Audio Clarity Improvement, User Evaluation of Cultural Authenticity, Archiving Accessibility, and Stability over Long-term Preservation. The HCC-GNN model achieved the highest scores, with 91.2% for Preservation Quality and a User Evaluation score of 4.6 out of 5. These results underscore the system's potential to preserve cultural authenticity while enhancing technical aspects of preservation and curation. The data further demonstrate the model's ability to improve both fidelity and accessibility of historical archival audio, particularly in preserving the cultural features and characteristics of archived audio files, with evaluations conducted by Guzheng musicians.

Table 4: Comparative Evaluation of Cultural Preservation and Audio Curation Techniques for Xinjiang-Style Guzheng Music.

Evaluation Criteria	CNN	RNN	GRU	RNN+GRU	HCC-GNN
Preservation Quality (Audio Fidelity %)	81.2	82.0	83.1	85.4	91.2
Noise Reduction Efficiency (%)	72.5	74.8	76.3%	78.7	88.7
Metadata Accuracy for Cultural Tags (%)	77.0	79.5	81.2%	83.5	89.5
Culturally Relevant (%)	78.5	80.1	82.0	84.8	90.3
Background Noise Tolerance (dB)	17.2 dB	18.0 dB	18.5 dB	19.8 dB	22.4 dB
Authenticity Detection (%)	75.8	78.2	79.6	82.1	86.9
Audio Clarity Improvement (%)	73.4	75.0	76.8	79.0	87.6
User Evaluation of Cultural Integrity (/5)	4.0/5	4.1/5	4.3/5	4.4/5	4.6/5
Archival Accessibility (Ease of Search %)	74.2	76.0	77.5	79.7	88.2
Long-Term Preservation Stability (%)	80.1%	82.5%	84.0%	86.5%	91.7%

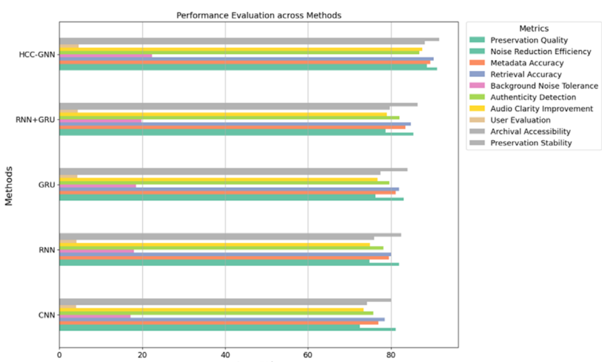

Figure 4: Cultural Preservation Effectiveness in Xinjiang-Style Guzheng Music Archiving.

The integration of digital technologies has significantly impacted cultural preservation, particularly in safeguarding Xinjiang-style Guzheng music. This study highlights how an effective IRF overcomes traditional archive limitations, combining modern digital systems with heritage preservation functions to protect musical practices threatened by globalisation. Guzheng music, with its rich historical background, suffers from poor documentation and inadequate digital storage, limiting access and preservation. The research calls for enhanced digital infrastructure to ensure worldwide accessibility and preservation of cultural artefacts.

The HCC-GNN model, with its innovative combination of sound quality and retrieval capabilities, extracts short-term audio features using CCNNs, ensuring clarity and preserving delicate Guzheng tones and rhythms. Audio clarity is further improved through Wiener filtering and normalisation. The integration of GRUs and RNNs enables the detection of long-term musical patterns, essential for capturing the unique rhythmic structures of Guzheng music. The system's processing methods not only improve retrieval outcomes but also enhance cultural accuracy in music databases, facilitating a deeper understanding of both musical and cultural practices. The HCC-GNN model excels in preserving audio quality, even under suboptimal conditions, effectively addressing degradation in historical recordings. The experimental results show that the proposed model outperforms traditional approaches in terms of accuracy, precision, recall, and F1 scores. Achieving 97.5% accuracy and a 95.2% F1 score, the model excels in selecting relevant recordings from a diverse data pool with varying genres and recording qualities. The creation of metadata standards aligned with cultural patterns is crucial for effective digital archiving. Metadata serves as the primary retrieval method for digital archives, but the oral tradition of music transmission, combined with regional cultural distinctiveness, often leads to irregular metadata.

To preserve Guzheng music, efforts must be directed at addressing these challenges and establishing contextual relevance within cultural heritage. Users from various backgrounds, including researchers and general audiences, will engage more deeply with the archive when metadata highlights cultural significance, instrument usage, and performance techniques. Data augmentation, particularly through Wavelet Transform, enhances the diversity of training datasets, reinforcing the feature extraction methods. These techniques, combined with simulations of real-world conditions like audience noise and performance variations, enable the model to handle performance intricacies effectively. Building a functional digital sound archive requires collaboration between musicians, information curators, and local communities where Guzheng music originates. Ensuring authenticity and cultural relevance necessitates active involvement from all stakeholders to maintain the archive's cultural integrity.

Conclusion

HCC-GNN, an IRF framework, combines convolutional feature extraction with graph-based sequential learning to enhance the retrieval accuracy and preservation quality of Xinjiang-style Guzheng music across diverse recording conditions and acoustics. The model, using a CCNN and GRU structure, captures spatial and temporal patterns crucial for authentic cultural preservation. HCC-GNN achieved a retrieval accuracy of 97.5%, outperforming CNN (83.2%), RNN (81.5%), and GRU (84.1%) models. It demonstrated strong noise tolerance with an SNR of 21.5 dB and improved metadata accuracy to 89.5% for cultural tags by integrating image feature representations. The model increased MAP to 0.80 and reduced retrieval time to 87 ms compared to traditional models. With a long-term conservation stability of 91.7%, HCC-GNN proved its reliability for archival use, significantly improving audio clarity, authenticity detection (86.9%), and accessibility, making it highly effective for preserving Xinjiang-style Guzheng music in modern digital archives.

Limitation and Future Scope

HCC-GNN excelled in retrieval accuracy and preservation quality but may face challenges with low-quality or corrupted recordings. Future research will focus on advanced denoising techniques, domain adaptations for new styles, and the development of lightweight models for real-time, resource-efficient retrieval.

References

Amiri, N., Riahinia, N., Arastoopoor, S., Haji, Z. M. and Alimohammadi, D. (2022). Mapping Data Elements of Persian Traditional Music Resources to the Entities, Attributes, and Relationships of IFLA's Library Reference Model. *Cataloging & Classification Quarterly*, 60(3-4): 315-328. https://doi.org/10.1080/01639374.2022.2079790

Bonjack, S. and Trujillo, N. (2023). Artificial Intelligence and Music Discovery. *Music Reference Services Quarterly*, 27(1): 1-9. https://doi.org/10.1080/10588167.2023.2287924

Borlund, P., Pharo, N. and Liu, Y.-H. (2024). Information searching in cultural heritage archives: a user study. *Journal of Documentation*, 80(4): 978-1002. https://doi.org/10.1108/JD-06-2023-0120

Buragohain, D., Meng, Y., Deng, C., Li, Q. and Chaudhary, S. (2024). Digitalizing cultural heritage through metaverse applications: challenges, opportunities, and strategies. *Heritage Science*, 12(1): 295. https://doi.org/10.1186/s40494-024-01403-1

Daga, E., Daquino, M., Fournier-S'Niehotta, R., Guillotel-Nothmann, C. and Scharnhorst, A. (2023). Documenting the research process. Opportunities and challenges for Bibliometrics and Information Retrieval. *Bibliometric-enhanced Information Retrieval*: 4-20. https://doi.org/10.5281/zenodo.10529114

Egan, P. (2022). In search of the item: Irish traditional music, archived fieldwork and the digital. *Archival Science*, 23(1): 45-63. https://doi.org/10.1007/s10502-021-09382-z

Fantozzi, C., Bressan, F., Pretto, N. and Canazza, S. (2017). Tape music archives: from preservation to access. *International Journal on Digital Libraries*, 18(3): 233-249. https://doi.org/10.1007/s00799-017-0208-8

Folorunso, S. O., Afolabi, S. A. and Owodeyi, A. B. (2022). Dissecting the genre of Nigerian music with machine learning models. *Journal of King Saud University - Computer and Information Sciences*, 34(8, Part B): 6266-6279. https://doi.org/10.1016/j.jksuci.2021.07.009

Gabbolini, G. and Bridge, D. (2024). Surveying More Than Two Decades of Music Information Retrieval Research on Playlists. *ACM Transactions on Intelligent Systems and Technology*, 15(6): 1-68. https://doi.org/10.1145/3688398

Gorgoglione, M., Garavelli, A. C., Panniello, U. and Natalicchio, A. (2023). Information Retrieval Technologies and Big Data Analytics to Analyze Product Innovation in the Music Industry. *Sustainability*, 15(1): 828. https://doi.org/10.3390/su15010828

Hajarian, M., Carrillo, M. H., Díaz, P. and Aedo, I.

(2025). Gamispotify: a gamified social music recommendation system based on users' personal values. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-20588-y

He, Q. (2022). A Music Genre Classification Method Based on Deep Learning. *Mathematical Problems in Engineering*, 2022(1): 9668018. https://doi.org/10.1155/2022/9668018

Lee, K. J. M., Ens, J., Adkins, S., Sarmento, P., Barthet, M. and Pasquier, P. (2025). The GigaMIDI Dataset with Features for Expressive Music Performance Detection. *Transactions of the International Society for Music Information Retrieval*, 8(1): 1-19. https://doi.org/10.5334/tismir.203

Lee, K. Y. and Hu, C. M. (2022). Research on the Development of Music Information Retrieval and Fuzzy Search. *Scientific and Social Research*, 4(4): 1-10. https://doi.org/10.26689/ssr.v4i4.3771

Malin, Y., Crowder, C., Byom, C. and Shanahan, D. (2022). Community Based Music Information Retrieval: A Case Study of Digitizing Historical Klezmer Manuscripts from Kyiv. *Transactions of the International Society for Music Information Retrieval*, 5(1): 208-221. https://doi.org/10.5334/tismir.135

Ni, T. (2023). An Intelligent Retrieval Algorithm for Digital Literature Promotion Information Based on TRS Information Retrieval. *International Journal of Information Technologies and Systems Approach (IJITSA)*, 16(2): 1-14. https://doi.org/10.4018/ijitsa.318458

Ostermann, F., Vatolkin, I. and Ebeling, M. (2023). AAM: a dataset of Artificial Audio Multitracks for diverse music information retrieval tasks. *EURASIP Journal on Audio, Speech, and Music Processing*, 2023(1): 13. https://doi.org/10.1186/s13636-023-00278-7

Patel, P., Shah, S., Prasad, S., Gada, A., Bhowmick, K. and Narvekar, M. (2024). Audio separation and classification of Indian classical instruments. *Engineering Applications of Artificial Intelligence*, 133: 108582. https://doi.org/10.1016/j.engappai.2024.108582

Roa Dabike, G., Cox, T. J., Miller, A. J., Fazenda, B. M., Graetzer, S., Vos, R. R. *et al.* (2024). The cadenza woodwind dataset: Synthesised quartets for music information retrieval and machine learning. *Data in Brief*, 57: 111199. https://doi.org/10.1016/j.dib.2024.111199

Rossetto, F., Dalton, J. and Murray-Smith, R. (2023). Generating Multimodal Augmentations with LLMs from Song Metadata for Music Information Retrieval. In: *Proceedings of the 1st Workshop on Large Generative Models Meet Multimodal Applications*. Association for Computing Machinery, pp. 51-59. https://doi.org/10.1145/3607827.3616842

Salosaari, P. (2022). The Audio Legacy of Finnish Radio: An Exploration of Key Factors in the Preservation of Radio Sound Collections. *TMG Journal for Media History*, 25(2): 1-27. https://doi.org/10.18146/tmg.823

Seekhunlio, W., Chuangprakhon, S. and Phiwphuy, K. (2024). The Preservation of Isan Folk Music with Digital Sound Technology. *Multidisciplinary Science Journal*, 6(4): 2024058. https://doi.org/10.31893/multiscience.2024058

Shamsi, F. and Sindhu, I. (2025). Condensing Video Content: Deep Learning Advancements and Challenges in Video Summarization Innovations. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3526068

Six, J., Bressan, F. and Renders, K. (2023). Duplicate Detection for for Digital Audio Archive Management: Two Case Studies. In: A. Biswas, E. Wennekes, A. Wieczorkowska, & R. H. Laskar (Eds.), *Advances in Speech and Music Technology: Computational Aspects and Applications*. Springer International Publishing, pp. 311-329. https://doi.org/10.1007/978-3-031-18444-4_16

Sun, M. (2024). An Intelligent Retrieval Method for Audio and Video Content: Deep Learning Technology Based on Artificial Intelligence. *IEEE Access*, 12: 123430-123446. https://doi.org/10.1109/ACCESS.2024.3450920

Sun, W. and Sundarasekar, R. (2023). Research on pattern recognition of different music types in the context of AI with the help of multimedia information processing. *ACM Transactions on Asian and Low-Resource Language Information Processing*. https://doi.org/10.1145/3523284

Wu, R. and Pan, Y. (2025). Providing music selection and matching suggestions for dance creations using music information retrieval and artificial intelligence techniques. *Journal* *of Computational Methods in Sciences and Engineering*: 14727978251318807. https://doi.org/10.1177/14727978251318807

Wu, S., Yu, D., Tan, X. and Sun, M. (2023). CLaMP: Contrastive Language-Music Pre-training for Cross-Modal Symbolic Music Information Retrieval. *arXiv preprint arXiv:2304.11029*. https://doi.org/10.48550/arXiv.2304.11029

Zhou, Y., Zhang, Z., Wang, X., Sheng, Q. and Zhao, R. (2024). Multimodal archive resources organization based on deep learning: a prospective framework. *Aslib Journal of Information Management*, 77(3): 530-553. https://doi.org/10.1108/AJIM-07-2023-0239

Zhu, T., Fournier-S'niehotta, R., Rigaux, P. and Travers, N. (2022). A Framework for Content-Based Search in Large Music Collections. *Big Data and Cognitive Computing*, 6(1): 23. https://doi.org/10.3390/bdcc6010023

Junyan Wang Junyan Wang is a PhD candidate specializing in Musicology. She is currently pursuing her doctoral studies in English Language at Silpakorn University. She holds a Master's degree in English Translation from Anhui University of Finance and Economics (2019–2022) and a Bachelor's degree in Music from the Shenyang Conservatory of Music (2010–2014). Her interdisciplinary background combines expertise in music, language, and cultural studies, supporting her research interests in musicological analysis and cross-cultural communication.

Junzhu Zhao Junzhu Zhao is currently a Master of Science (MSc) candidate in Statistics at the University of Leeds. She holds a Bachelor's degree in Mathematics and Applied Mathematics, which she earned from Anqing Normal University in July 2023. Born in December 2001, her academic interests lie primarily in the field of statistics.

Abbreviation Table

Abbreviation	Full-Form / Meaning
MIR	Music Information Retrieval
DSP	Digital Signal Processing
KMDMP	Kiselgof-Makonovetsky Digital Manuscript Project
AI	Artificial Intelligence
BD	Big Data
AI-BD	Artificial Intelligence and Big Data Integration
NO	Neural Networks
CNN	Convolutional Neural Network
ML	Machine Learning
EDM	Electronic Dance Music
XGBoost	Extreme Gradient Boosting
ORIN	Open Repository of Indigenous Nigerian Music (dataset name)
GitHub	Online platform for code hosting and collaboration
NLP	Natural Language Processing
API	Application Programming Interface
EC	European Commission