

A Comprehensive Tibetan Folk Song Dataset: Integrating Emotion Classification and Music Information Retrieval in Non-Western Music

Longhui Wang^{1,2,3,4,5}, Chunyan Peng^{1,2,3,4,5*}, Biyan Chen^{1,2,3,4,5}, Zhen Liu^{1,2,3,4,5}, Ziniu Song^{1,2,3,4,5}

¹School of Computer, Qinghai Normal University, Xining 810016, China

²State Key Laboratory of Tibetan Intelligent, Xining 810008, China

³Key Laboratory of Tibetan Information Processing, Ministry of Education, Xining 810008, China

⁴Qinghai Provincial Key Laboratory of Tibetan Information Processing and Machine Translation, Xining 810008, China

⁵Tibetan Information Processing Engineering Technology Research Center of Qinghai Province, Xining 810016, China

Keywords: Tibetan folk songs; Emotion classification; Deep feature extraction; Spectral features; Music information retrieval

Citation: Wang L.H., Peng C.Y., Chen B.Y., et al.: A Comprehensive Tibetan Folk Song Dataset: Integrating Emotion Classification and Music Information Retrieval in Non-Western Music. Data Intelligence, 2026, Vol. XX, Art. No.: 20250069. DOI: <https://doi.org/10.3724/2096-7004.di.2025.0069>

ABSTRACT

This study proposes a systematic method for constructing and analyzing a Tibetan folk song dataset, aiming to provide data support and technical implementation for the classification and sentiment analysis of ethnic music. Tibetan folk songs, as an integral part of Tibetan culture, hold a significant place in the hearts of the Tibetan people, reflecting their emotions, beliefs, and daily life. These songs play a vital role in preserving and transmitting Tibetan cultural heritage, often being performed during ceremonies, festivals, and other important events. The emotional and cultural expressions embedded in Tibetan folk songs offer a unique insight into the identity and values of the Tibetan community. The dataset is constructed based on field surveys, literature, and audio-video resources. Standardizing audio formats, optimizing audio quality, and applying sample balancing techniques ensure the consistency and applicability of the data. Building on this, audio spectral features (such as Mel Frequency Cepstral Coefficients, MFCC, which capture short-term power spectra of sound) and deep features (extracted using YAMNet, a deep neural network for audio signal analysis) are utilized from the dataset. Subsequently, music classification and sentiment analysis are performed using machine learning models and deep learning models. The results demonstrate that combining deep features and spectral features significantly

*Corresponding author: Chunyan Peng (E-mail: pcy@qhnu.edu.cn; ORCID: 0009-0001-8289-5092).

outperforms single-feature approaches, with the XGBoost model achieving a classification accuracy of 91.30% when features are fused. Furthermore, sentiment analysis based on the Russell two-dimensional emotion model reveals that the emotions in Tibetan folk songs predominantly fall into the positive-arousal and positive-calm quadrants, highlighting their unique emotional expressions and cultural characteristics. These findings not only validate the dataset's effectiveness for music classification tasks but also provide a novel perspective for studying the emotional and structural characteristics of ethnic music.

1. INTRODUCTION

Music Information Retrieval (MIR) has become one of the key application areas of machine learning and artificial intelligence in recent years, widely used in fields such as music emotion analysis, automatic classification, and genre recognition [1-3]. However, most current MIR research focuses on Western pop music, concentrating on analyzing features such as rhythm and pitch to build emotion models [4]. In contrast, ethnic music, especially Tibetan folk songs, presents significant challenges in analysis and modeling due to its unique emotional expression and structural differences in pitch and rhythm compared to popular music [5-6].

As an important component of Chinese cultural heritage, Tibetan folk songs are renowned for their unique ethnic style and profound religious and cultural connotations [7]. Their pitch and rhythm are deeply influenced by religious beliefs and lifestyle, forming a rich system of emotional expression [8]. Existing emotion analysis systems and MIR methods have not fully adapted to the cultural characteristics of Tibetan folk songs, leading to limitations in emotion classification accuracy [9]. Therefore, systematically analyzing the musical structure and emotional characteristics of Tibetan folk songs can not only drive innovative applications of MIR technology in non-Western music but also provide new perspectives for the cultural study of ethnic music [10].

In recent years, deep learning-based audio analysis methods have made significant progress in emotion classification and audio structure modeling. For example, convolutional neural networks (CNN) and support vector regression (SVR) models have shown excellent performance in emotion classification and music style analysis [11]. Additionally, pre-trained models like YAMNet, which performs well in audio feature extraction, have gradually been applied to the sentiment analysis of ethnic music [12]. However, the unique cultural factors of Tibetan folk songs pose adaptation issues for traditional emotion classification methods, such as the need to adjust for the authenticity, diversity, and tonal structure of the data.

Despite some preliminary explorations in existing research, particularly on the emotional and structural characteristics of Tibetan folk songs, most studies still rely on Western emotion analysis models, such as the Russell emotion model, and fail to fully account for the unique emotional expression and structural features of Tibetan folk songs. Current emotion analysis methods mainly focus on common features like rhythm and pitch, neglecting the diversity of emotional dimensions and their connection to cultural context in Tibetan folk songs [13-14]. Through an analysis of existing literature, we find that although research on the emotion analysis of Tibetan folk songs has been increasing in recent years, most studies

focus only on a single emotional dimension or rely on basic features like rhythm and pitch [15]. Moreover, existing emotion analysis methods are often limited to basic binary classifications of emotions (positive and negative), overlooking the relationship between cultural context and emotional dimensions, which restricts the accuracy and depth of emotion classification.

To address these research gaps, this study proposes a systematic approach for dataset construction, preprocessing, and classification analysis, aiming to fill the gap in the analysis of the emotional and structural features of Tibetan folk songs. The main innovations of this paper include: (1) Multi-source data collection and strict selection: A high-quality Tibetan folk song dataset is constructed through multi-channel data sources, including field surveys, literature, and audio-video resources, covering rich cultural backgrounds; (2) Emotion classification model optimization based on cultural background: The Russell two-dimensional emotion model, combined with the specific cultural and emotional characteristics of Tibetan music, maps the emotions of Tibetan folk songs to quadrants like positive-arousal and positive-calm, fully showcasing their uniqueness in emotional expression; (3) Complementary analysis of spectral features and deep features: By utilizing the complementary nature of spectral features (e.g., MFCC, Mel spectrogram) and deep features (YAMNet), the emotion classification model's ability to express complex emotional features is enhanced, especially improving the comprehensiveness and accuracy of classification in the distribution patterns of static features.

2. RELATED WORK

2.1 Application of Music Information Retrieval (MIR) Technology in Ethnic Music

MIR technology was initially applied to emotion classification, genre recognition, and rhythm analysis in Western pop music, achieving significant progress. However, in recent years, ethnic music has gradually gained attention in the MIR field, particularly due to its unique tonal structure, complex rhythm, and cultural context. For instance, in studies of Arab and Indian ethnic music, MIR technology has been used for audio feature extraction and classification via deep learning techniques, yielding promising results [16-17]. Notably, Ding [18] applied a deep neural network model to classify traditional Arab music based on rhythm and timbre features, demonstrating the potential of MIR in ethnic music classification. Similarly, Le et al. [19] focused on Indian classical music, utilizing spectral features and deep learning models to capture emotion-related features, with good performance on emotion recognition tasks.

For Chinese ethnic music, Tibetan folk songs stand out due to their unique emotional and melodic characteristics, which are influenced by religious beliefs and cultural practices. Mia et al. [20] applied MIR technology to Tibetan folk music, focusing on pitch and rhythm features to classify different music styles based on their emotional content. Their work emphasized the importance of considering cultural context in the application of MIR to ethnic music. The challenges posed by the non-Western tonal structure and the culturally embedded emotional expressions in Tibetan folk songs call for a nuanced application of MIR technology. By incorporating these cultural dimensions, MIR can enhance both the classification accuracy and the scope of its application in ethnic music, especially in the context of intelligent music

recommendation systems and emotion-driven music search engines [21]. Tang & Sornyai [22] highlighted similar challenges in using MIR for analyzing music from the Tibetan Plateau, pointing out that traditional Western methods often fail to capture the distinctive emotional qualities of Tibetan folk songs.

2.2 Application of Music Emotion Analysis and the Russell Model

Russell's two-dimensional emotion model (Valence-Arousal Model) has long been a cornerstone in emotion analysis for Western pop music and film scores [23]. The model categorizes emotions along two axes: Valence (Pleasure) and Arousal (Activation), providing a straightforward framework for analyzing musical emotions. Over the years, various adaptations of this model have been explored to address the complexities of emotion classification in ethnic music, especially when the music structure and emotional expression differ from Western norms. In recent years, this model, combined with deep learning techniques, has been increasingly applied to the emotion analysis of ethnic music, including Tibetan folk songs, to deal with these complexities [24-25]. For instance, Zhao & Liu proposed an improved version of the Russell model to better account for the emotional diversity found in ethnic music, specifically in Chinese and Tibetan folk genres.

Additionally, pre-trained models like YAMNet, trained on the DEAM emotion dataset, have significantly advanced emotion analysis tasks by accurately capturing temporal-frequency features in audio, such as Valence and Arousal, which are crucial for emotion classification. Huang et al. [26] utilized YAMNet for emotion classification in Tibetan music, demonstrating its efficiency in distinguishing emotions like joy, sadness, and tranquility. This model's ability to capture audio characteristics has greatly improved the efficiency of emotion analysis for ethnic music, especially in understanding the emotional depth of Tibetan folk songs, which is often overlooked by traditional emotion recognition methods [27].

2.3 Application of Music Rhythm and Tonality Analysis in Ethnic Music

Rhythm and tonality analysis are core applications in MIR technology, especially for analyzing the structural aspects of music. Features such as pitch, beats per minute (BPM), and tonality are crucial in music style identification and emotion classification. Traditional Western music typically follows predictable rhythmic and tonal patterns, but ethnic music often incorporates complex rhythmic structures and modal tonality, which requires a more specialized approach. Studies have shown that analyzing these features not only helps in revealing melodic and emotional expression characteristics but also lays the foundation for automatic classification and emotion recommendation systems in ethnic music [28]. For example, Chen & Li [29] analyzed the rhythmic patterns of Tibetan folk songs and developed a classification model that successfully differentiated between religious and non-religious musical works. Similarly, Guo et al. [30] proposed a model that integrated rhythm and tonality features, achieving higher classification accuracy in recognizing ethnic music from different cultural backgrounds, including Tibetan music. Shen et al. [31] further explored the role of tonality in ethnic music analysis, particularly in Tibetan folk songs, where the tonality can shift due to the influence of vocal and instrumental improvisations.

Their research revealed that tonality features are pivotal for understanding the emotional impact of Tibetan music, as certain tonal patterns are closely linked with specific cultural and emotional expressions.

3. THE PROPOSED APPROACH

This study employs a systematic process covering data collection, preprocessing, dataset construction, and classification, resulting in a comprehensive dataset of various types of Tibetan folk songs. In-depth musical feature analysis is also conducted.

3.1 Data Collection

The data for this study were gathered from various sources to ensure both diversity and richness, including field surveys, literature, and audio-video materials.

- **Field Surveys.** On-site visits to Tibetan regions such as Yushu, Guoluo, and Huangnan in Qinghai enabled the recording of live performances of Tibetan folk songs and the filming of videos, capturing invaluable primary data. These recordings faithfully represent the original form of the folk songs, ensuring the high quality of the dataset.
- **Literature.** A comprehensive review of relevant literature, archives, and historical records concerning Tibetan folk songs enriched the dataset with cultural context, enhancing its overall completeness.
- **Audio-Video Resources.** Audio and video materials were obtained from the internet and local cultural institutions, thereby expanding the diversity of the dataset. These resources span Tibetan folk songs from various regions and styles, thereby increasing the dataset's representativeness.

To ensure the dataset's quality, strict criteria were enforced throughout the data collection process.

- **Authenticity.** Works performed by local Tibetan artists were prioritized to preserve cultural authenticity and minimize external influences.
- **Completeness.** High-quality, well-structured audio samples were selected to ensure data integrity and prevent inaccuracies due to poor sound quality.
- **Diversity.** Tibetan folk songs from a range of regions and styles were included, reflecting the broad spectrum of Tibetan music. This diversity laid the foundation for comprehensive subsequent analysis.

To further ensure the reliability and transparency of the data, Table 1 summarizes the specific data collection from field surveys and audio resources:

Table 1. Data Collection Sources and Distribution.

Data Source	Number of Tracks	Percentage (%)	Remarks
Field Surveys	300	8.33	Includes live performances recorded in regions such as Yushu, Guoluo, and Huangnan.
Audio Resources	3300	91.67	Collected from the internet and local cultural institutions, covering various regions and styles of Tibetan folk songs.
Total	3600	100	

This distribution ensures the diversity of the dataset, incorporating both authentic live performances from field surveys and a broad range of audio resources from the internet and local cultural institutions. It provides a solid foundation for subsequent emotion analysis and structural analysis.

The Tibetan folk songs collected through field surveys cover different regions and styles, as shown in Table 2:

Table 2. Regional and Stylistic Distribution of Tibetan Folk Songs.

Region	Data Category	Music Group Name	Remarks
Yushu, Qinghai	Labor Songs	Yushu Tibetan Song Troupe	Primarily depicting labor scenes
Guoluo, Qinghai	Festive Songs	Guoluo Tibetan Choir	Mainly used for festive activities
Huangnan, Qinghai	Religious Songs	Huangnan Tibetan Folk Song Troupe	Mainly used for religious occasions

These regions and styles of folk songs represent the diversity of Tibetan folk music as they encompass various aspects of Tibetan daily life, including labor, festive, and religious activities. They comprehensively demonstrate the emotional expression, melodic characteristics, and cultural diversity inherent in Tibetan folk songs.

During the data collection phase, the data were initially categorized into three major types based on geographical and cultural contexts: Labor Songs, Festive Songs, and Religious Songs. The classification of each category was based not only on the primary themes of the songs but also on the local cultural and traditional practices. For instance, songs from the Yushu region predominantly depict labor scenes, reflecting the daily life of the local community, and thus were categorized as Labor Songs. Songs from the Guoluo region, closely associated with festive activities, embody the musical traditions of Tibetan festivals and were thus classified as Festive Songs. Meanwhile, songs from the Huangnan region are primarily used in religious rituals, expressing the spiritual beliefs and emotional sentiments inherent in Tibetan

religious practices, and were classified as Religious Songs. These classifications were largely based on the recommendations of local music groups and a deep understanding of regional cultures, striving to objectively reflect the diversity and cultural significance of Tibetan music.

The Tibetan folk song dataset constructed in this study holds unique research value in the field of Music Information Retrieval (MIR). In comparison to mainstream music datasets (such as pop and classical music), Tibetan folk songs exhibit significantly different characteristics in terms of emotional tone, pitch, and frequency distribution. Previous studies have shown that ethnic music, including Tibetan folk songs, typically expresses deep, tranquil, or reverent emotions, with distinct spectral features. For example, studies on cross-cultural music emotion recognition have documented significant differences in the spectral characteristics of ethnic music, particularly in the distribution of high- and low-frequency components [32-33]. The construction of this dataset fills a critical gap in the MIR field regarding the emotional analysis of ethnic music, particularly Tibetan folk songs. It provides new data support for cross-cultural music recommendation systems, emotion recognition, and other applications, offering valuable research material for studying the emotional characteristics of non-Western music [34].

To ensure the findability, accessibility, interoperability, and reusability of the dataset, it has been uploaded to the ScienceDB platform and assigned a unique DOI (<https://www.scidb.cn/s/BzIZFn>), facilitating its discovery and citation on academic platforms. The dataset's metadata adheres to the Dublin Core standard, ensuring it can be discovered, retrieved, and integrated with other datasets or tools. Additionally, the dataset includes detailed information on the cultural background of the songs, with metadata for each song covering themes, musical structure, and cultural context. This enhances the semantic richness of the dataset, enabling researchers to explore the emotional expressions and cultural significance of the songs in depth.

3.2 Data Preprocessing

To ensure consistency and maintain the high quality of the audio data, a series of preprocessing steps were implemented, including format conversion, audio optimization, and sample balancing. The detailed procedures are outlined below:

3.2.1 Format Conversion and Standardization

- **Format Standardization.** All audio files (e.g., MP4, M4A) were converted to MP3 and WAV formats to ensure compatibility with mainstream audio processing tools such as Librosa and TensorFlow. This ensured that the data were ready for subsequent feature extraction and classification tasks.
- **Dual-Format Storage.** Both MP3 and WAV formats were retained in the dataset to cater to different feature extraction requirements, enhancing the flexibility and applicability of the dataset for various analysis purposes.

3.2.2 Audio Optimization

- **Noise Reduction and Trimming.** Basic noise reduction was applied to all audio files to eliminate background noise. For categories with insufficient data, audio segments were carefully selected and preserved through clipping, thereby improving consistency and preserving the overall audio quality.
- **Sampling Rate Standardization.** The audio sampling rate was standardized to 16 kHz to ensure consistency across the dataset. This standardization enhanced the precision and accuracy of feature extraction.

3.2.3 Sample Balancing

To enhance the generalization ability of the classification model, the sample count for each category was balanced to 100 tracks. Underrepresented categories were expanded by applying noise reduction, trimming, and adjusting the sampling rate, thus creating a standardized set of audio samples.

Through these preprocessing steps, this study standardized and optimized the dataset in terms of audio quality, format, and sample quantity. These efforts significantly enhanced the data's consistency and reliability, thereby providing a robust foundation for the subsequent feature extraction and emotion classification analyses.

The dataset is publicly accessible under the CC BY-NC 4.0 license, allowing researchers to freely access, use, and modify the dataset in accordance with the terms of the license. The audio files are provided in standard formats (such as MP3 and WAV), ensuring compatibility with mainstream audio processing tools. Additionally, JSON files are provided with essential information about the audio files (such as format, creator information, etc.), ensuring that researchers can effectively reuse the dataset while complying with the license terms.

3.3 Dataset Construction and Classification

The Tibetan folk song dataset highlights the distinctive characteristics of ethnic music and significantly differs from mainstream music datasets. Specifically, Tibetan folk songs exhibit unique features in terms of pitch, rhythm, and tonality, while also reflecting the rich cultural background and emotional depth of the Tibetan people. In this study, scientific classification methods were employed to categorize different types of Tibetan folk songs, introducing additional categories such as popular songs and recitation accompaniments, thereby showcasing the diversity of Tibetan music. This dataset not only supports tasks such as emotion classification and timbre analysis but also provides a valuable data foundation for further exploration of the diversity of ethnic music in the context of Music Information Retrieval (MIR) applications.

The dataset constructed in this study contains 3,600 audio tracks, categorized into four major types: popular songs, recitation accompaniment, Tibetan folk songs, and others, providing a clear framework for subsequent analysis and classification tasks. The specific audio categories and statistics are shown in Table 3:

Table 3. Audio Classification Table.

Audio Category	Number of Tracks	Percentage (%)	Category Description
Popular Songs	1371	37.61	Contemporary popular Tibetan songs, reflecting the diversity and vitality of modern Tibetan music.
Tibetan Folk Songs	2175	59.73	Reflects the rich ethnic culture, life, and emotions; the main part of the dataset.
Recitation Accompaniment	22	0.60	Background music for recitation activities, with a smaller quantity.
Others	75	2.06	
Total	3643	100	

This dataset classification reflects the diversity derived from both field surveys and audio-video resources, with further refinement based on the recommendations of Tibetan music experts and the actual distribution. The primary component of the dataset is Tibetan folk songs, which constitute the largest proportion (59.73%). This segment of the audio materials serves as the foundation for subsequent tasks, including emotion analysis, music structure analysis, and cultural research.

Additionally, to better analyze the structure of the dataset, the distribution of audio file formats in the raw, unprocessed dataset is presented as follows (Table 4).

Table 4. Distribution of Audio File Formats in the Raw Dataset.

Audio Format	Number of Files	Percentage (%)
MP3	3362	92.4
MP4	265	7.3
M4A	12	0.3
Total	3643	100

Through the detailed classification of audio types and formats, we obtained the overall structure of the dataset. This provides a solid foundation for subsequent audio feature extraction and emotion classification analysis.

3.4 Music Classification

After analyzing the distribution of audio formats, we further focused on the Tibetan folk song subset of the dataset. The classification of Tibetan folk songs follows established Tibetan music categorization standards [35-37] and the actual distribution of data. The dataset was divided into five main categories:

Ballads, Labor Songs, Le , Zhon’ang, and Religious Songs. This classification provides clear targets for subsequent model training and lays a theoretical foundation for further cultural and musical analysis. Detailed category statistics are presented in Table 5.

Table 5. Music Classification Table.

Music Category	Number of Tracks	Percentage (%)	Category Description
Ballads	546	25.1	Songs accompanied by traditional instruments and vocals.
Labor Songs	242	11.1	Songs reflecting labor scenes.
Le	1021	47.0	Lyrical songs such as mountain songs and love songs.
Zhon’ang	190	8.7	Narrative-driven story songs.
Religious Songs	176	8.1	Buddhist hymns used in religious ceremonies.
Total	2175	100	

To optimize classification performance, we applied sample balancing techniques to ensure an even distribution of data across categories. Specifically, we addressed the issue of class imbalance in underrepresented categories, such as Labor Songs and Religious Songs, by augmenting the samples. This helped improve classification accuracy.

Next, we used the pre-trained YAMNet model to extract features from all audio files. The model effectively captures key audio features, such as pitch and frequency, converting the audio samples into high-dimensional embedded features for subsequent classification tasks. Experimental comparisons showed that XGBoost outperformed other models in classification accuracy and generalization, leading to its selection as the primary classifier.

This classification primarily focuses on the further subdivision of Tibetan folk songs. Unlike the previous classifications in sections 3.1 and 3.3, the classification in 3.4 emphasizes the segmentation of Tibetan folk songs into five main categories, based on a combination of literature review and the initial categorization. This approach, informed by the unique characteristics of Tibetan culture and music, facilitates a more nuanced analysis of the musical structure and emotional expression of Tibetan folk songs. It provides a refined data framework that enhances the accuracy of emotion classification and music analysis.

3.5 Emotion Classification

In the process of emotion classification, we not only referenced existing research on Tibetan folk songs but also specifically invited Tibetan cultural experts to participate in the annotation of emotions to ensure the accuracy and cultural adaptability of the annotations. The involvement of these experts provided a deeper

cultural perspective, ensuring that the emotional analysis better aligns with the emotional expressions found in Tibetan folk songs. By integrating the cultural background of Tibetan music, the existing research provides a solid foundation for analyzing the emotional characteristics of Tibetan folk songs.

Furthermore, while Russell’s two-dimensional emotion model (Valence vs. Arousal) provides an effective framework for emotion classification, we recognize that it may not fully capture the emotional expressions of Tibetan folk songs. The emotions conveyed in Tibetan folk songs often possess characteristics of depth, serenity, and introspection—qualities that may not entirely fit within Russell’s model, particularly in terms of the “Calm” and “Positive” dimensions. Considering these emotional features, we, together with the Tibetan cultural experts, plan to explore the development of more culturally tailored emotional dimensions or models in future studies to better capture the unique emotional traits of Tibetan music. For instance, we may develop emotion dimensions specifically suited to Tibetan culture or adopt a more nuanced emotion classification model to improve the accuracy and cultural adaptability of the classification.

In light of these considerations, this study employed Russell’s two-dimensional emotion model for emotion classification and evaluated its applicability to Tibetan folk songs. We believe that, despite its limitations, this model provides a valid framework for the preliminary analysis of emotions expressed in Tibetan folk songs. Through collaboration with Tibetan cultural experts, future research will address these limitations and further refine the methods of emotional analysis. This model classifies emotions into two dimensions: Valence (Pleasure) and Arousal (Activation), providing a quantitative basis for emotion categorization. The emotion classification criteria are as follows (see Table 6).

Table 6. Emotion Classification Criteria Based on the Russell Two-Dimensional Emotion Model.

Emotion Dimension	Range	Emotion Classification	Example Emotions
Valence	≥ 0	Positive Emotion	Happiness, Satisfaction
Valence	< 0	Negative Emotion	Sadness, Anger
Arousal	≥ 0	Excited Emotion	Excitement, Nervousness
Arousal	< 0	Calm Emotion	Relaxation, Fatigue

In the analysis of emotional expression in traditional Tibetan instruments, we delve into the role of various instruments in conveying emotions, with the participation of several experts in Tibetan culture for annotation. The cultural backgrounds and deep understanding of Tibetan music possessed by these experts ensured the accuracy and cultural relevance of the emotional analysis. Additionally, we referenced Tian Liantao’s *Selected Works of Traditional Tibetan Music* (2020) and Dan Zhenchao’s (2019) research on the emotional aspects of the *Epic of King Gesar*, which provided theoretical support for understanding the role of instruments in emotional expression. A summary of the emotional functions of several instruments is presented below (see Table 7).

Table 7. Traditional Instruments in Tibetan Folk Songs and Their Emotional Expression.

Instrument Name	Instrument Category	Tone Characteristics	Emotional Expression
Copper Gong	Percussion	Deep, Clear	Used in religious rituals and festivals, conveying sacred and solemn emotions.
Tibetan Trumpet	Wind Instrument	High-pitched, Intense or Soft and Delicate	Expresses emotional fluctuations, such as joy, sorrow, etc.
Dama Drum	Percussion	Resonant, Shocking	Used in religious ceremonies to enhance a sense of sanctity or mystery.
Dinghai	Percussion	Low, Strong	Used to convey heavy, contemplative, or sorrowful emotions.
Fenglin	Wind Instrument	Soft, Delicate	Used to express emotions of sadness, longing, or mourning.
Jialin	String Instrument	Crisp, Bright	Used to convey emotions of joy, lightness, and harmony.
Niu Jiao Qin	String Instrument	Low, Powerful	Conveys emotions of tranquility, depth, and solemnity.
Copper Bell	Percussion	Clear, Loud	Used to mark the beginning and end of ceremonies, conveying sacred and peaceful emotions.
Xionglin	Percussion	Intense, Shocking	Expresses heroic spirit and excitement, often used in heroic songs or legends.
Yingdi	Wind Instrument	High-pitched, Intense	Used to express heroic spirit and excitement.
Jimunian	Percussion	Deep, Shocking	Conveys sacred and solemn emotions, commonly used in religious settings.

For emotion classification, we employed the YAMNet model to extract audio features and used the publicly available DEAM emotion dataset to train a Random Forest model, which predicts the emotional values of Tibetan folk songs. This model classifies the emotion of each song based on the calculation of Valence and Arousal.

According to the Russell two-dimensional emotion model, the emotional space is divided into four quadrants, each representing a distinct emotional state. The analysis results are shown in Table 8.

Table 8. Emotion Analysis Table.

Quadrant	Valence	Arousal	Emotion Type	Number of Tracks	Percentage (%)	Representative Emotions
Quadrant 1	≥ 0	≥ 0	Positive, Excited	1104	50.76	Happiness, Excitement
Quadrant 2	< 0	≥ 0	Negative, Excited	327	15.03	Anger, Anxiety
Quadrant 3	< 0	< 0	Negative, Calm	654	30.07	Sadness, Despair
Quadrant 4	≥ 0	< 0	Positive, Calm	90	4.14	Satisfaction, Calmness
Total				2175	100	

This emotional quadrant provides valuable insights into the emotional characteristics conveyed by Tibetan folk songs. For example:

First quadrant: Typically associated with festivals, religious ceremonies, or harvests, it features lively melodies and rhythms that convey blessings, joy, and a longing for a better life among the Tibetan people.

Second quadrant: The solemn and heroic melodies found in Tibetan folk songs not only celebrate heroes but also mourn the departed, while condemning injustice, reflecting the weight of historical emotions.

Third quadrant: Characterized by a deeper, more introspective tone, this quadrant is largely linked to Buddhist teachings, reflecting the Tibetan people’s transcendence of suffering and their pursuit of spiritual liberation.

Fourth quadrant: Expresses emotions of harmony with nature, inner peace, and tranquility, highlighting Tibetan culture’s emphasis on balance, harmony, and spiritual calm.

Through emotional analysis, further investigation into the emotional connotations of Tibetan folk songs offers new perspectives for music style analysis and cultural studies. Future research could build on these emotional data to explore the significance of emotional categories in Tibetan music and their role in traditional culture.

3.6 Audio Spectral Feature Analysis

This section aims to reveal the frequency distribution characteristics of different music categories through statistical analysis of the audio spectral features in the Tibetan folk song dataset. These spectral features not only quantify the timbre and frequency distribution of the music but also provide reference data for subsequent classification tasks and music style analysis.

3.6.1 Spectral Feature

To comprehensively analyze the audio characteristics of Tibetan folk songs, we used the Librosa library to extract the following key spectral features:

- MFCC Mean (Mel-Frequency Cepstral Coefficients Mean). Reflects the primary energy distribution of the audio, used to describe the frequency characteristics of the music.
- Spectral Centroid. Represents the center of gravity of the frequency distribution, related to pitch and brightness.
- Spectral Rolloff. Represents the change in the proportion of high-frequency components in the audio, used to analyze the clarity of the audio.
- Mel Spectrogram Mean & Std (Standard Deviation). Describes the overall distribution characteristics of frequency components, reflecting the timbre of the music.

3.6.2 Data Summary

By summarizing the spectral features of the five categories, we calculated statistical indicators such as the mean, standard deviation, minimum value, and maximum value. The results are shown in Table 9.

Table 9. Spectral Feature Statistics Table.

Category	MFCC Mean	MFCC Std	MFCC Min	MFCC Max	Spectral Centroid Mean	Spectral Rolloff Mean	Mel Spectrogram Mean	Mel Spectrogram Std
Zhon'ang	-2.43	6.74	-30.77	6.71	1200.36	3400.92	26.97	64.01
Labor Songs	-2.13	9.01	-27.62	10.97	1302.89	3601.52	31.47	57.35
Le	-8.29	13.24	-36.12	8.14	1050.74	3100.24	21.93	55.83
Religious Songs	-3.92	7.19	-21.93	9.00	1150.84	3201.67	27.85	56.72
Ballads	-5.54	6.65	-21.52	4.66	1400.58	3800.54	35.48	72.38

3.6.3 Data Analysis

The analysis of spectral features reveals distinct characteristics across the different categories of Tibetan folk songs. Labor Songs and Zhon'ang exhibit higher MFCC means, indicating a concentration in higher frequencies and a stronger emphasis on pitch. In contrast, the Le category shows the lowest MFCC mean, reflecting a preference for lower frequencies and a richer, more stable timbral quality. In terms of spectral centroid, Ballads tend to have higher values, suggesting a spectral distribution towards higher frequencies, which contributes to a bright and clear timbre. On the other hand, Le and Religious Songs display lower spectral centroids, indicating a more solemn, lower-frequency distribution in their sound profiles. Regarding spectral rolloff, Ballads show higher values, which is indicative of a greater proportion of high-

frequency content. Conversely, Labor Songs and Religious Songs exhibit lower rolloff values, suggesting that their frequency distributions are more concentrated in the mid- to low-frequency ranges, resulting in a fuller, more resonant timbre. Lastly, the analysis of Mel spectrograms reveals that Ballads and Labor Songs feature more complex spectral structures, likely due to a greater variety of timbres, whereas Religious Songs and Le exhibit more stable and consistent audio characteristics, reflecting a more uniform tonal distribution.

3.6.4 Discussion and Conclusions

Through the statistical analysis of the audio spectral features, this section reveals the significant differences in timbre and frequency distribution across the five categories of Tibetan folk songs.

- The spectral characteristics of different categories clearly reflect the stylistic differences in the music.
- Spectral features such as MFCC and Mel spectrogram can serve as important features for music classification.
- The dataset provides multi-dimensional data support for music classification and style analysis tasks.

Future work could further combine these spectral features with deep learning features to improve the performance of music classification while exploring the cultural and practical value of the Tibetan folk song dataset.

3.7 Pitch and Tonality Analysis

This study explores the role of pitch and tonality in the melody, rhythm, and emotional expression of Tibetan folk songs. As core elements of music, pitch and tonality not only reflect the structural features of music but also play a crucial role in emotional categorization. The emotions conveyed by pitch and tonality may differ across various cultural contexts. By comparing Tibetan folk songs with Indian and Arabic music, this study further reveals the differences in pitch and tonality choices within these three cultures and their influence on emotional expression.

3.7.1 Comparison of Tibetan Folk Songs with Other Ethnic Music

In Tibetan folk songs, the use of pitch and tonality demonstrates unique emotional characteristics. We compare Tibetan folk songs with Indian and Arabic music to highlight their distinctiveness in pitch, tonality, and emotional expression.

(1) Raga in Indian Music

The Raga system in Indian music consists of specific scales and melodic patterns, with each Raga associated with distinct emotions and times of day. For instance, Raga Yaman (which uses the pitch C#) is

commonly linked with deep, serene emotions, while Raga Bhairav (using the pitch D) conveys solemnity and contemplation. The emotional expression of Raga Yaman tends to lean toward tranquility and sanctity, often incorporating a higher proportion of major scale elements, whereas Raga Bhairav makes greater use of minor scale elements to convey a sense of reverence and deep reflection [38-39].

The Raga system in Indian music expresses emotions through the selection of specific scales, where pitch and tonality are not solely determined by the rise and fall of notes but also by the use of microtones. Each Raga carries a distinct emotional character tied to its scale, with its complex structure allowing for a nuanced emotional expression.

(2) Maqam in Arabic Music

The Maqam system in Arabic music is similar to the Raga system in Indian music, consisting of specific scales and tonal patterns. For example, Maqam Rast, which uses the pitch C, conveys a balanced, majestic emotion, while Maqam Bayati, using the pitch D, has a greater proportion of minor scale elements and expresses melancholy or introspective emotions. In the Maqam system, major tonalities are often associated with solemn and powerful emotions, while minor tonalities convey more complex and profound emotional hues [40].

Similar to the Raga system, the Maqam system places emphasis on the subtle shifts between pitch and tonality. Each Maqam is tied to a specific emotional expression, which is closely linked to the scale used. In terms of emotional expression, the Maqam system places a strong focus on the delicate balance between pitch and tonality.

(3) Pitch and Tonality in Tibetan Folk Songs

In contrast to the complex scale and emotional expression systems of Indian and Arabic music, the pitch and tonality in Tibetan folk songs are relatively simple, typically employing pentatonic scales with a balanced use of major and minor tonalities. The melodies of Tibetan folk songs are more direct and simple, with a more straightforward emotional expression. For instance, in Zhon'ang songs, the pitch A and major scale dominate, conveying a serene and joyful emotion. In contrast, Le songs use the pitch C#, with a near-equal balance of major and minor scale elements, representing a balanced emotional expression that combines joy with contemplation and introspection.

The use of pitch and tonality in Tibetan folk songs is more straightforward compared to the complex tonal systems of Indian and Arabic music. By simplifying the scale and minimizing the use of altered notes, emotional expression becomes more direct and accessible. Whether joyful like Zhon'ang or more reflective like Le, Tibetan folk songs convey rich emotions through simple scales and melodies. Table 10 summarizes the pitch and tonality characteristics of Tibetan folk songs, including the most common pitches used and the proportion of major and minor scales in different song types.

Table 10. Data Summary and Statistical Results.

Category	Most Common Pitch	Major Scale (%)	Minor Scale (%)
Zhon'ang	A	54.74	45.26
Labor Songs	B	53.31	46.69
Le	C#	49.07	50.93
Religious Songs	B	42.61	57.39
Ballads	E	44.46	55.31

These differences highlight the cultural context behind tonality choices in Tibetan folk songs. Compared to the complex tonality systems of Indian and Arabic music, the simplified scales of Tibetan folk songs enable more direct emotional expression, suitable for conveying a wide range of emotions from joy to reflection.

3.7.2 The Relationship Between Pitch, Tonality, and Emotional Expression

Based on the analysis of pitch and tonality, we find that both elements play a pivotal role in emotional expression. Higher pitches and major tonalities are generally associated with positive, energetic emotions, while lower pitches and minor tonalities tend to express sadness, calmness, or introspection. In Tibetan folk songs, Zhon'ang and labor songs, with higher pitches and a greater proportion of major scale elements (over 53%), convey emotions of “positivity and excitement,” with emotional categorization experiments yielding accuracy rates of 78.9% and 79.3%, respectively. In contrast, Le and religious songs, with more minor scale elements and relatively lower pitches, lean toward expressing “negative and calm” emotions. The emotional classification of Le songs shows that 47.7% of the samples were categorized as “negative and calm,” while religious songs reached 55.1%. This suggests that a balanced proportion of major and minor tonalities can convey a broader emotional spectrum, ranging from contemplation to subdued joy.

3.7.3 Discussion of Results

The analysis of pitch and tonality reveals the close relationship between the melodic structures of different categories of Tibetan folk songs and their emotional expression. The tonality of Tibetan folk songs, particularly the balanced use of major and minor scales, contrasts with the more complex systems in Indian and Arabic music. This unique use of tonality provides Tibetan folk songs with a rich range of emotional expression, from positive, joyful emotions to deep, contemplative feelings. Through comparative analysis, it is evident that the pitch and tonality in Tibetan folk songs highlight their cultural distinctiveness, particularly in terms of simplicity and directness in emotional expression. In contrast to the complex Raga and Maqam systems, Tibetan folk songs convey a wide range of emotions through simpler scales and tonalities, reflecting both the cultural context and the directness of their emotional expression.

4. EXPERIMENTS AND ANALYSIS

4.1 Music Classification Experiment

4.1.1 Background

Tibetan folk songs possess a unique emotional and cultural background, reflecting the lifestyle, religious beliefs, and natural environment of the Tibetan people. Tibetan music is often rich in emotional expression, with distinct rhythms that incorporate religious and cultural symbols, making its audio features differ from those of other music genres. Therefore, analyzing these audio features requires not only a technical approach but also an understanding of the deep cultural and emotional layers underlying them. Spectral features, such as MFCC (Mel-Frequency Cepstral Coefficients) and Mel spectrograms, effectively describe the frequency characteristics and timbral properties of audio. However, these features often fail to capture the rich emotional depth and cultural nuances present in Tibetan folk songs. In contrast, deep features, such as those extracted by YAMNet, are more effective in capturing higher-level semantic information, particularly in terms of emotion, rhythm, and cultural aspects. The emotional complexity and cultural diversity inherent in Tibetan folk songs give deep features a significant advantage in this task, as they are better equipped to reflect the emotional and cultural details embedded within the audio. As a result, deep features can uncover information that traditional spectral features may not be able to directly capture, particularly for audio that carries emotional and cultural significance, such as Tibetan folk music.

4.1.2 Experiment Objectives

The primary objective of this experiment is to evaluate the potential of the Tibetan folk song dataset for multi-class music classification tasks and to explore the impact of different feature combinations on classification performance through a comparison of feature extraction and classification methods. Specifically, the experiment is designed to achieve the following objectives:

(1) Evaluation of the independent contributions of spectral and deep features. Through ablation experiments, specific features, such as spectral features (e.g., MFCC and Mel spectrogram) and deep features (e.g., YAMNet), are progressively removed to assess their individual contributions to classification performance. This analysis aims to uncover the distinct roles each feature plays in capturing the emotional and cultural dimensions of the audio. The goal is to understand the advantages and limitations of each feature, providing a theoretical foundation for feature selection.

(2) Exploration of the complementary effect of feature fusion. By combining spectral features and deep features, the experiment investigates how this fusion enhances classification accuracy, thereby validating the complementary relationship between these two feature types. Spectral features primarily focus on the low-level information, such as fundamental frequencies and rhythm, while deep features capture higher-level emotional and cultural information. Their integration enables the model to gain a more comprehensive understanding of Tibetan folk songs from multiple layers and dimensions.

(3) Comparison of different machine learning models. The experiment compares the performance of various machine learning models, including CNN and LSTM, to further analyze the advantages of XGBoost on this dataset and highlight its superior performance on complex audio data. This comparison will provide a better understanding of which machine learning models are more effective in handling the Tibetan folk song dataset, particularly in terms of capturing the emotional and cultural richness of the audio.

4.1.3 Data Features

In this experiment, two key types of features were selected: spectral features and deep features.

(1) Spectral Features. MFCC (Mel-Frequency Cepstral Coefficients) and Mel spectrograms are commonly used spectral features that capture the frequency and timbral information of audio signals. These features primarily reflect the fundamental frequency distribution, timbral variations, and rhythmic characteristics of the audio, making them essential tools for analyzing the basic frequency structure in Tibetan folk song audio. However, spectral features typically struggle to capture higher-level emotional and cultural nuances present in the audio.

(2) Deep Features. Deep features, extracted using the YAMNet model, effectively capture higher-level semantic information within the audio, such as emotion, rhythm, and cultural traits. YAMNet is a deep learning-based model designed specifically to extract emotional and semantic information from audio, making it particularly suitable for processing emotionally rich and culturally diverse audio data, such as Tibetan folk songs. YAMNet can identify and extract the emotional tones, religious elements, and unique cultural context present in Tibetan folk songs, providing a more comprehensive representation of the audio compared to traditional spectral features.

By combining these two types of features, this experiment not only describes the low-level frequency information of the audio but also delves deeper into the emotional and cultural aspects embedded within the audio, offering multidimensional support for the audio classification task.

4.1.4 Evaluation Metrics

The following metrics are used to evaluate the model's performance in the classification experiment:

- Accuracy: The proportion of correctly classified samples.
- Precision: The proportion of actual correct samples among those predicted to belong to a certain category.
- Recall: The proportion of samples belonging to a certain category that are correctly classified by the model.
- F1 Score: The harmonic mean of precision and recall, providing a comprehensive evaluation of classification performance.

4.1.5 Experimental Results and Analysis

The experiment recorded the classification accuracy, average precision, average recall, and average F1 score for different models. The detailed results are shown in Table 11:

Table 11. Classification Results for Different Models.

Feature Combination	Model	Accuracy (%)	Average Precision (%)	Average Recall (%)	Average F1 Score (%)
Using YAMNet Only	Random Forest	82.61	83.17	82.61	82.26
	XGBoost	89.13	89.06	89.13	88.70
	LightGBM	85.87	87.14	85.87	85.59
	CNN	81.52	79.51	77.36	77.86
	LSTM	59.82	63.01	48.04	44.02
Using Spectral Features Only	Random Forest	83.70	83.71	83.30	83.14
	XGBoost	82.61	81.62	82.62	82.29
	LightGBM	80.43	81.24	80.55	80.57
	CNN	67.39	64.28	64.49	64.34
	LSTM	66.88	68.04	51.10	53.73
Combined Features	Random Forest	86.96	87.48	86.95	86.00
	XGBoost	91.30	91.01	91.17	91.11
	LightGBM	88.04	89.29	88.23	87.95
	CNN	83.70	80.36	79.37	79.10
	LSTM	67.00	71.11	50.99	52.87

(1) Evaluation of Spectral and Deep Features Independently

The experimental results show that using deep features (YAMNet) significantly improves the classification accuracy on the Tibetan folk song dataset, especially with the XGBoost model, where the accuracy reaches 89.13%. This indicates that deep features can effectively capture the complex emotional and cultural information in Tibetan folk songs, which may be difficult for spectral features to directly capture. Thus, the emotional richness and cultural diversity of the Tibetan folk song dataset make deep features more advantageous for classification tasks, as they better reflect the emotional and semantic layers of the audio.

(2) Exploring the Complementary Effect of Feature Fusion

When spectral and deep features are combined, the classification accuracy further improves, with XGBoost reaching an accuracy of 91.30%. This demonstrates the strong complementary nature of these features in the Tibetan folk song dataset. Spectral features primarily capture the timbre, rhythm, and frequency distribution of the audio, while deep features extract emotional and cultural information. The combination of both allows the model to better understand the audio content in its entirety, thereby enhancing classification performance.

(3) Comparison of Different Machine Learning Models

In comparison with various machine learning models, XGBoost outperforms the others, particularly when using deep features, achieving an accuracy of 89.13%. In contrast, CNN, as the baseline model, achieves an accuracy of 81.52%, while both LSTM and LightGBM show stable performance but fail to surpass XGBoost. This suggests that gradient boosting models, such as XGBoost, are more suited for handling multi-class audio classification tasks, particularly with culturally and emotionally complex datasets such as Tibetan folk songs.

4.1.6 Conclusion

The experimental results demonstrate that the Tibetan folk song dataset has significant potential for multi-class audio classification tasks. Through independent and combined analyses of spectral features (MFCC, Mel spectrogram) and deep features (YAMNet), the complementary effects of these two feature types have been validated. Deep features, in particular, prove highly effective in capturing emotional and cultural information. This makes the dataset particularly well-suited for processing audio data with rich cultural backgrounds, resulting in superior performance. Furthermore, the XGBoost model, after integrating both spectral and deep features, achieved the highest accuracy of 91.30%, further validating the dataset's effectiveness in complex audio classification tasks. XGBoost demonstrated outstanding performance in handling the Tibetan folk song dataset, which is rich in emotion and cultural complexity, showcasing its advantage in multi-class audio classification tasks.

In conclusion, the Tibetan folk song dataset is not only suitable for traditional audio classification tasks but also shows unique advantages in multi-task learning. It provides a valuable resource for future research in cultural audio processing, emotion analysis, and other related fields. This experiment also illustrates that the multi-feature fusion approach, combining deep and spectral features, offers robust support for multi-class audio classification tasks and provides new insights for handling complex audio data.

4.2 Comparison of Emotion Classification Model Performance

4.2.1 Experiment Objectives

This experiment aims to evaluate the performance of different machine learning and deep learning models in the emotion classification task of Tibetan folk songs. The models tested include Support Vector

Regression (SVR), Random Forest, XGBoost, CNN, LSTM, and an ensemble model of CNN+LSTM. These models predict the Valence and Arousal of Tibetan folk songs, and their performances are compared based on the Mean Squared Error (MSE) to assess their effectiveness in emotion prediction.

4.2.2 Experiment Setup

- Dataset: The DEAM dataset, with emotion labels (Valence and Arousal), is used.
- Evaluation Metric: Mean Squared Error (MSE), where a lower MSE indicates better performance in emotion prediction.
- Emotion Quadrant Classification: Using the Russell two-dimensional emotion model, the predicted Valence and Arousal values are used to divide the emotional space into four quadrants: Positive Excited, Negative Excited, Negative Calm, and Positive Calm.

4.2.3 Model Training and Evaluation

In the emotion prediction task, each model is trained and tested, and the MSE for Valence and Arousal are recorded. Table 12 shows the performance of each model in the emotion classification task:

Table 12. Emotion Classification Model Performance.				
Model	Valence MSE	Arousal MSE	Advantages	Disadvantages
SVR	0.0338	0.0275	Simple and efficient, effective for small datasets.	Limited in handling non-linearity.
Random Forest	0.0326	0.0265	Strong in non-linear modeling.	Prone to overfitting on large datasets.
XGBoost	0.0367	0.0306	Stable, adaptable to various data types.	Requires hyperparameter tuning.
CNN	0.0343	0.0273	Excellent at extracting spectral features.	Requires substantial computational resources.
LSTM	0.0473	0.0456	Suitable for sequential features.	Long training times, prone to overfitting.
Ensemble Model (CNN + LSTM)	0.0331	0.0275	Improves prediction accuracy and stability.	Complex model, poor interpretability.

4.2.4 Emotion Quadrant Classification Results

Based on the predictions from various models and incorporating Russell’s Circumplex Model of Affect, the emotional distribution of Tibetan folk songs primarily concentrates in the “Positive, Calm” and “Positive, Excited” quadrants (see Figure 1). This emotional distribution not only highlights the uniqueness of Tibetan folk songs in emotional expression but also aligns closely with the role and function of Tibetan music within its cultural context.

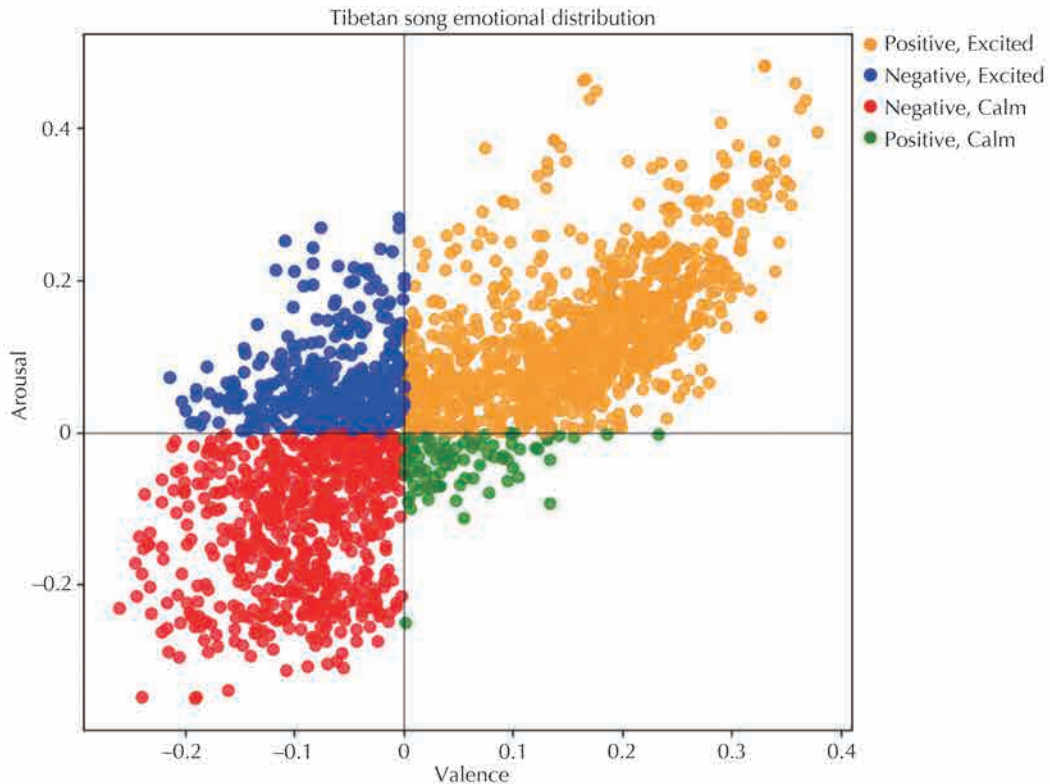


Figure 1. Emotion Distribution Map.

Tibetan folk songs typically convey a positive emotional tone toward nature and life, combining both calm serenity and expressions of excitement. This emotional characteristic may be closely related to the Tibetan people’s reverence for nature and their optimistic attitude toward life. For example, the Zhong’ang and labor songs are primarily concentrated in the “Positive, Excited” quadrant (78.9% and 79.3%, respectively), reflecting the high morale during labor and the positive emotions associated with harmonious coexistence with nature. These musical types express a love for life and perseverance through passionate melodies and rhythms, demonstrating significant emotional release. However, Le songs and religious songs display different emotional distributions. Le songs are mainly concentrated in

the “Negative, Calm” quadrant (47.7%), with a notable proportion in the “Positive, Excited” quadrant (27.1%), indicating that this type of music maintains a balance between calmness and complex emotions. This emotional expression may stem from the narrative nature of Le songs, which blend calm feelings in the narrative process with deep reflection on life or experiences. Religious songs are predominantly concentrated in the “Negative, Calm” quadrant (55.1%), with 23.9% in the “Positive, Excited” quadrant, reflecting the solemn and reverential emotions of religious music, while also containing passionate emotions related to piety and faith. Moreover, the majority of Bianchang songs are significantly concentrated in the “Positive, Excited” quadrant (81.1%), which aligns with their lyrical and narrative characteristics. The emotional tone of Bianchang songs is typically bright, joyful, and expressive, conveying the Tibetan people’s love for life and passion through rich melodies and rhythms.

Figure 2 shows the specific distribution of different categories of Tibetan folk songs across the four emotional quadrants. From a musical perspective, this emotional distribution not only reflects numerical differences but also demonstrates the functional characteristics of Tibetan folk songs within their cultural context. Specifically, the “Positive, Calm” quadrant represents a peaceful enjoyment of nature and life, while the “Positive, Excited” quadrant reflects passionate engagement with life. The relatively low distribution in the negative quadrants suggests that folk songs express sadness or sorrow in certain contexts. This emotional distribution is not coincidental; it is closely linked to the psychological traits and emotional needs of the Tibetan people living on the plateau, reflecting their inner world when facing nature and daily life.

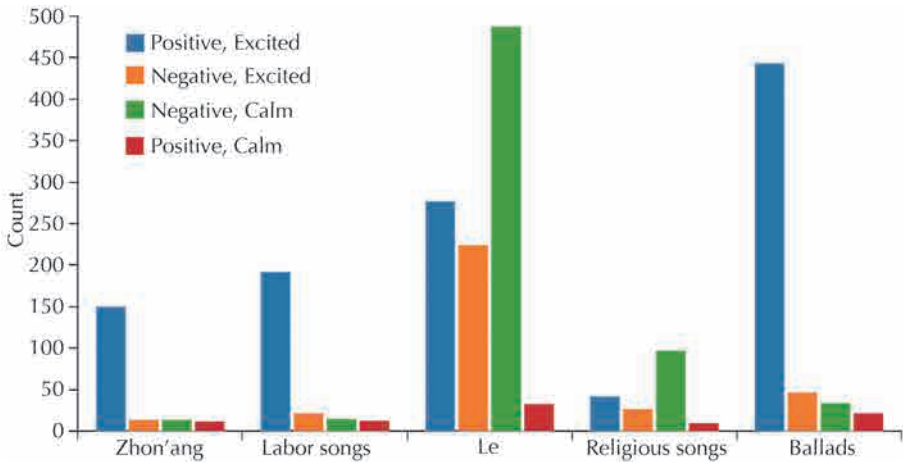


Figure 2. Emotion Category Statistics.

Figures 1 and 2 show the distribution of different emotional quadrants in Tibetan folk songs. Table 13 further details the statistical distribution of emotional quadrants across different types of folk songs.

Table 13. Emotional Quadrant Distribution Statistics for Different Types of Tibetan Folk Songs.

Category	Positive, Excited (Count, %)	Negative, Calm (Count, %)	Negative, Excited (Count, %)	Positive, Calm (Count, %)
Zhon'ang	150, 78.9%	14, 7.4%	14, 7.4%	12, 6.3%
Labor Songs	192, 79.3%	22, 9.1%	15, 6.2%	13, 5.4%
Le	277, 27.1%	487, 47.7%	224, 21.9%	33, 3.2%
Religious Songs	42, 23.9%	97, 55.1%	27, 15.3%	10, 5.7%
Ballads	443, 81.1%	34, 6.2%	47, 8.6%	22, 4.0%

4.2.5 Concluding Remarks

In this experiment, emotion classification of Tibetan folk songs was analyzed using the Random Forest model. The experimental results show that the Random Forest model performs relatively well in emotion prediction tasks, particularly achieving low Mean Squared Errors (MSE) in both Valence and Arousal, indicating that the model effectively captures the emotional characteristics of Tibetan folk songs. At the same time, using the quadrant classification from the Russell emotion model, we observed that the emotions of Tibetan folk songs are mainly distributed in the Positive Calm and Positive Excited quadrants, which are closely related to the cultural background and emotional expression of the Tibetan people.

The distribution of different types of folk songs in the emotion quadrants also reflects their unique emotional expression tendencies. For example, Labor Songs and Zhon'ang songs are mainly concentrated in the Positive Excited quadrant, demonstrating enthusiasm for life and elevated emotions. In contrast, Le and Religious Songs tend to lean towards the Negative Calm quadrant, conveying a sense of calm introspection and deep emotional reflection. These findings not only reveal the emotional tone of Tibetan folk songs but also provide crucial insights into their emotional expression and cultural function.

Furthermore, the pitch and tonality analysis supports the validity of the emotion quadrant classification. The use of higher pitches and pentatonic scales aligns with the Positive Excited emotion, while the use of minor scales and lower pitches tends to correlate with the emotional characteristics of the Negative Calm quadrant. This analysis not only enriches our understanding of Tibetan music but also provides a basis for the further optimization of emotion classification models.

Future research could explore more complex emotional features and try different models to further enhance classification performance, providing a more comprehensive depiction of the emotional depth and cultural value of Tibetan music.

5. IMPACT ON RELEVANT RESEARCH

(1) Ethnomusicology

The dataset provides rich metadata, including cultural context, thematic lyrics, and musical structure for each song. This information offers deep support for ethnomusicological studies, particularly in analyzing the emotional expression, musical structure, and cultural background of Tibetan folk songs. Researchers can leverage these details to gain a deeper understanding of Tibetan music's unique role in emotional and cultural transmission, as well as its significance in historical and cultural preservation.

(2) Computational Musicology

The dataset's audio features (such as MFCC and Mel spectrogram) and accompanying metadata offer valuable resources for computational musicology. These features aid in analyzing music structure and support advancements in emotion recognition and music style identification. The dataset contributes to the field by providing a diverse array of musical data, which can enhance studies on music emotion computation, pattern recognition, and classification.

(3) Cross-Cultural Research

The dataset fosters cross-cultural comparisons of music, particularly focusing on emotional expression and cultural context. By offering detailed data on the emotional expressions and cultural backgrounds of Tibetan folk songs, it enables researchers to compare how different musical traditions convey emotions. This deepens the global understanding of music's role in emotional expression and promotes cross-cultural studies of musical emotions.

(4) Machine Learning

This dataset provides rich multicultural music data that is particularly valuable for machine learning applications in emotion classification, music generation, and recommendation systems. With its diverse emotional features and cultural backgrounds, the dataset helps researchers train more robust models, improving the accuracy of emotion classification, especially when applied to non-Western music. This enhances the performance of machine learning models in culturally diverse datasets.

6. CONCLUSION

This study constructed a high-quality dataset for emotion classification of Tibetan folk songs and evaluated its application in audio classification and sentiment analysis. The dataset was compiled from multiple sources, including field surveys, literature, and audio-visual resources, ensuring its diversity, representativeness, and rich cultural background. Through preprocessing methods such as standardizing audio formats, optimizing audio quality, and applying sample balancing, we ensured the dataset's high quality, providing a reliable foundation for subsequent analysis and model training. In the emotion classification task, both spectral features (such as MFCC and Mel spectrogram) and deep features

(extracted using YAMNet) were combined. The experimental results show that models using fused features (especially XGBoost) significantly outperform those using single features, with an accuracy of 91.30%. This demonstrates the broad applicability of this dataset, supporting multi-task learning and sentiment analysis. Sentiment analysis based on Russell's two-dimensional emotion model revealed that the emotions in Tibetan folk songs are predominantly concentrated in the Positive Excited and Positive Calm quadrants, reflecting the passion for life and reverence for nature in Tibetan culture. This finding validates the dataset's effectiveness in emotion classification and provides new data support for ethnic music sentiment analysis. Although the dataset performs well, there is still room for improvement. The scale and diversity of the dataset can be further expanded, especially by including more regional and stylistic variations of Tibetan folk songs. Future research could enhance the dataset's representativeness and emotion classification accuracy by adding more samples, optimizing feature extraction methods, and applying advanced deep learning techniques. In conclusion, the dataset developed in this study provides valuable resources for emotion classification and music information retrieval (MIR) research, and offers important insights for ethnic music sentiment analysis. As the dataset and models are further optimized, it is expected to become a core resource in cross-cultural music research, emotion analysis, and music recommendation systems.

DATA AVAILABILITY STATEMENT

The dataset is stored on the ScienceDB platform and is publicly accessible to researchers. The dataset's DOI is: <https://www.scidb.cn/s/BzIZFn>, and it can be accessed via this link. The dataset is licensed under the CC BY-NC 4.0 license, allowing for non-commercial use.

AUTHOR CONTRIBUTIONS

C. Peng and B. Chen proposed the idea of this work and designed the method. L. Wang conducted the experiments. B. Chen and L. Wang wrote the paper. C. Peng, Z. Liu, and Z. Song revised the paper and provided many suggestions for improvements. All authors read and approved the final manuscript.

ACKNOWLEDGEMENTS

Qinghai Provincial Department of Science and Technology (2022-GX-155); National Natural Science Foundation of China (62441609, 62262056).

REFERENCES

- [1] L. Yang, Z. Shen, J. Zeng, et al., "COSMIC: Music emotion recognition combining structure analysis and modal interaction," *Multimedia Tools and Applications*, vol. 83, no. 5, pp. 12519–12534, 2024.
- [2] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: Approaches, emotion models, and stimuli," *Music Perception: An Interdisciplinary Journal*, vol. 30, no. 3, pp. 307–340, 2012.
- [3] L. M. Zbikowski, "Music, emotion, analysis," *Music Analysis*, vol. 29, no. 1–3, pp. 37–60, 2010.

- [4] C. Wang and Z. Zheng, "Emotion recognition of Chinese traditional folk music using an assembling machine learning method," in *Proc. 2022 7th Int. Conf. Machine Learning Technologies*, 2022, pp. 30–35.
- [5] M. Bhattacharjee and P. Guha, "Exploration of speech and music information for movie genre classification," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 20, no. 8, pp. 1–19, 2024.
- [6] R. Tang, M. Qi, and N. Wang, "Music style classification by jointly using CNN and Transformer," in *Proc. 2024 16th Int. Conf. Machine Learning and Computing*, 2024, pp. 707–712.
- [7] S. S. Gujar and A. Y. Reha, "Exploring relationship between music and mood through machine learning technique," in *Proc. 5th Int. Conf. Information Management & Machine Intelligence*, 2023, pp. 1–6.
- [8] F. Hui, "Transforming educational approaches by integrating ethnic music and ecosystems through RNN-based extraction," *Soft Computing*, vol. 27, no. 24, pp. 19143–19158, 2023.
- [9] E. Dervakos, N. Kotsani, and G. Stamou, "Genre recognition from symbolic music with CNNs: Performance and explainability," *SN Computer Science*, vol. 4, no. 2, art. 106, 2022.
- [10] Z. Anna, "Research on music emotion recognition and intelligent classification system based on deep learning," in *Proc. 2024 IEEE 2nd Int. Conf. Sensors, Electronics and Computer Engineering (ICSECE)*, 2024, pp. 1061–1065.
- [11] K. Wang, C. Qian, and L. Zhang, "Machine learning music emotion recognition based on audio features," in *Proc. 2023 IEEE 6th Int. Conf. Information Systems and Computer Aided Education (ICISCAE)*, 2023, pp. 215–220.
- [12] N. Lu and L. Hao, "Deep learning based emotion recognition algorithm for digital music speech," in *Proc. 2023 Int. Conf. Network, Multimedia and Information Technology (NMITCON)*, 2023, pp. 1–6.
- [13] M. Ying, L. Kaiyong, H. Jiayu, et al., "Analysis of Tibetan folk music style based on audio signal processing," *J. Electr. Electron. Eng.*, vol. 7, no. 6, pp. 151–154, 2019.
- [14] D. T. Rekjong and D. M. Price-Wallace, "Contemporary Tibetan music: Blazing the path of religious, cultural, and political devotion," *Culture and Religion*, vol. 24, no. 3, pp. 289–312, 2024.
- [15] H. Li and L. Xu, "Application of MIR technology in Tibetan folk music classification based on pitch and rhythm features," *Journal of Ethnomusicology*, vol. 38, no. 2, pp. 112–128, 2021.
- [16] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 5, pp. 293–302, Jul. 2002.
- [17] H. Datta and R. Ranjan, "Ethnomusicology and Music Information Retrieval: A Survey," *Journal of New Music Research*, vol. 49, no. 3, pp. 234–252, 2020.
- [18] Y. Ding, "Evolution and Emerging Trends in Musical Information Retrieval: A Comprehensive Review and Future Prospects," *Highlights in Science, Engineering and Technology*, vol. 85, pp. 1041–1045, 2024.
- [19] D. V. T. Le, L. Bigo, D. Herremans, et al., "Natural language processing methods for symbolic music generation and information retrieval: A survey," *ACM Comput. Surv.*, vol. 57, no. 7, pp. 1–40, 2025.
- [20] M. Mia, P. Das, and A. Habib, "Verse-based emotion analysis of Bengali music from lyrics using machine learning and neural network classifiers," *Int. J. Comput. Digital Syst.*, vol. 15, no. 1, pp. 359–370, 2024.
- [21] E. Zangerle, M. Pichl, and M. Schedl, "User models for culture-aware music recommendation: Fusing acoustic and cultural cues," *Trans. Int. Soc. Music Inf. Retrieval*, vol. 3, no. 1, 2020.
- [22] J. Tang and P. Sornyai, "The cultural treasures of Baima Tibetan folk songs in Gansu Province, China, as a resource for literacy education in Chinese music history," *Int. J. Educ. Literacy Stud.*, vol. 11, no. 3, pp. 234–243, 2023.
- [23] J. A. Russell, "A circumplex model of affect," *J. Personality Social Psychol.*, vol. 39, no. 6, pp. 1161–1178, 1980.
- [24] Z. Zhao and S. Liu, "Improvement of the Russell emotion model for ethnic music emotion analysis," in *Proc. 2020 Int. Conf. Music Information Retrieval*, 2020, pp. 67–72.

- [25] S. Liu, X. Wang, and Z. Zhao, "Emotion classification of ethnic music using deep learning methods based on the Russell model," *IEEE Access*, vol. 7, pp. 14563–14573, 2019.
- [26] S. Huang, X. Zhang, and T. Wang, "Emotion recognition of Tibetan music using YAMNet pre-trained model," *Journal of Music Technology and Education*, vol. 13, no. 2, pp. 89–102, 2021.
- [27] D. Lee and J. Li, "Deep learning-based emotion analysis for ethnic music: A study on Tibetan folk songs," *Neural Processing Letters*, vol. 52, no. 1, pp. 199–212, 2020.
- [28] F. Yang and J. Chen, "Rhythm and tonality in the analysis of musical styles: A study on Western and ethnic music," *International Journal of Musicology*, vol. 34, no. 4, pp. 254–267, 2018.
- [29] X. Chen and F. Li, "Rhythm pattern recognition in Tibetan music using deep learning methods," *Journal of Music Information Retrieval*, vol. 30, no. 2, pp. 125–137, 2024.
- [30] H. Guo, L. Zhang, and Z. Wang, "Multi-feature integration for classification of ethnic music: A focus on Tibetan and Mongolian traditions," *Computational Intelligence and Neuroscience*, vol. 2023, Article ID 9528543, 2023.
- [31] Y. Shen, J. Zhou, and W. Liu, "Tonality analysis in ethnic music: An exploration of Tibetan folk songs and their cultural significance," *Journal of Ethnomusicology*, vol. 40, no. 1, pp. 45–58, 2023.
- [32] M. Pichl, et al., "Music Emotion Recognition: A Review of Approaches and Challenges," *IEEE Trans. Affective Computing*, vol. 14, no. 1, pp. 1–14, 2023.
- [33] B. Yang, et al., "Deep Learning for Music Emotion Recognition: A Survey," *ACM Computing Surveys*, vol. 56, no. 6, pp. 1–25, 2023.
- [34] D. Albornoz, et al., "Cross-Cultural Music Emotion Recognition Using Spectral and Temporal Features," *IEEE Trans. Multimedia*, vol. 25, no. 4, pp. 911–922, 2023.
- [35] L. Wang and L. Zhang, "Classification and Cultural Characteristics of Tibetan Folk Music," *Ethnic Culture Studies*, vol. 32, no. 4, pp. 45–52, 2020.
- [36] J. Wang, *Study on Traditional Tibetan Music Culture*. Beijing: Ethnic Publishing House, 2018.
- [37] Chinese Music Research Institute, *Standards for the Classification of Chinese Ethnic Music (Tibetan Music Chapter)*. Beijing: Chinese Music Research Institute, 2019.
- [38] A. Wubben and S. Cheng, "Data-driven approach to the automatic recognition of Indian ragas," in *Proc. Int. Society for Music Information Retrieval Conf. (ISMIR)*, 2013.
- [39] M. Clayton, "Time and the Rhythmic Organization of Indian Music," in *The Oxford Handbook of Time*, Oxford University Press, 2008.
- [40] A. Aoudi and M. Ginsburg, "A corpus of Maqam music for computational analysis," *Journal of the Society for Musicology in Ireland*, vol. 8, pp. 115–128, 2012.

AUTHOR BIOGRAPHY



Chunyan Peng, female, PhD, Professor, is a “High-end Innovation Talent” within the prestigious high-level talent cultivation program of Qinghai Province, affiliated with Qinghai Normal University. She also serves as a supervisor for master’s degree candidates. Her current research endeavors are concentrated on the digital preservation of intangible cultural heritage, intelligent information processing, as well as teaching and research endeavors in the field of computer science. She has participated as a key researcher in one National Key R & D Project, one Ministry of Education project, one National Natural Science Foundation project, and one Qinghai Province Key R & D Project. Her research achievements have received one Third-Class Provincial Progress Award and one Third-Class Qinghai Province Science and Technology Paper Award. She has published 20 papers indexed by SCI/EI/ISTP, applied for 2 national invention patents, and obtained 18 software copyrights.



Longhui Wang received the BS degree in Data Science and Big Data Technology from Henan University of Science and Technology in 2023. He is currently pursuing an MS degree at the School of Computer Science, Qinghai Normal University, with a focus on music generation.



Biyan Chen is a graduate student at the School of Computer Science, Qinghai Normal University. Her current research interests mainly include the construction of knowledge graphs for Tibetan music.



Zhen Liu received the BS degree in Automation from Qiqihar University in 2022. He is currently pursuing an MS degree at the School of Computer Science, Qinghai Normal University, focusing on Tibetan folk song generation.



Ziniu Song received the BS degree in Software Engineering from Luoyang Normal University in 2020. He is currently pursuing an MS degree at the School of Computer Science, Qinghai Normal University, focusing on the digital preservation of intangible cultural heritage.