

# A study on the classification of Chinese folk instruments based on deep learning

**Chuhan Zhou**

Jiaxing Senior High British Colombia Offshore school, Jiaxing, Zhejiang Province, China, 314000

zhoidi1114@163.com

**Abstract.** Due to the Internet's tremendous expansion in recent years, digital music has seen explosive growth. It is becoming increasingly important to provide fast and accurate music retrieval for users, and music information retrieval has gradually become a research hotspot. Among them, the classification of musical instruments is one of the hot research directions. As an important part of the world's musical instruments, the Chinese folk musical instrument tradition has great research value, but there is less research in this direction. Traditional music classification methods can be roughly divided into two steps: manual feature extraction and traditional machine learning. But traditional music classification has several shortcomings, as follows. Firstly, manual extraction of music features has a high error rate and it is very difficult to ensure the accuracy of the features, and the labor cost is high; secondly, traditional machine learning still has problems with multiple classifications and training on large scale data. Therefore, in this paper, based on previous research, the original data features are extracted from the Mel Frequency Cepstrum Coefficient (MFCC), which is a feature that can represent musical timbre, and deep learning is proposed to be used to classify traditional Chinese folk instruments. This paper finds that the results of identifying and classifying Chinese folk musical instruments using different classical classification methods are compared when the training set and test set data are the same. According to the experimental results, the use of KNN has the best performance among these classification methods.

**Keywords:** Deep learning, Mel frequency cepstral coefficient, Chinese folk instrument classification, K-Nearest Neighbor.

## 1. Introduction

With the spread of the internet and electronic devices, most people choose to enjoy music on a variety of music software or web pages. This has led to new discussions on how to retrieve the music that users want from the vast number of music libraries accurately and quickly. As a result, more and more researchers are entering the field of music information retrieval. So far, music information retrieval [1] includes the following areas: identification and classification of musical genres; identification and classification of composers; identification and classification of musical instruments; identification and classification of singers; identification and classification of emotions; separation of vocals; and automatic music generation. Researchers are increasingly

interested in research on the identification and classification of musical instruments. In 1999, Marques J extracted linear predictive coefficients (LPC) and Mel frequency cepstral coefficients (MFCC) features of different instruments with Gaussian mixture models (GMM) and support vector machines (SVM) to classify music for nine instruments and obtained a 70% recognition rate [2]. In 2004, Krishna A G also investigated two acoustic features, LPC and MFCC, and two types of classification algorithms, K-nearest neighbour algorithm and Gaussian mixture model, to classify and recognize 14 musical instruments, achieving a maximum average classification of 90% [3]. Although some researchers have made some achievements in this area, the vast majority of research on the classification of musical instruments has focused on Western instruments. There has been very little research into Chinese folk instruments.

In this paper, we decided to collect music clips of five different Chinese folk instruments from the Internet and extract musical features of the five instruments, among which, short-time features of the instruments were extracted. Next, different classifiers were used to classify them. Finally, the accuracy rates obtained with the different classifiers are compared to determine which classifier works best.

Deep learning has been widely used in the field of image processing, but rarely in the field of audio and music information retrieval. Chinese folk instruments are also an important part of the world's musical instruments, so it is interesting to study this area. That is why this paper wants to use deep learning to study the classification of Chinese folk musical instruments.

## **2. Fragments of five Chinese folk instruments**

Chinese folk instruments are unique to China and have a long historical background. It can be classified according to the form in which they are played into four main categories: wind instruments, plucked instruments, percussion instruments, and stringed instruments. In addition, different folk instruments produce distinctive tones or only single tones, which will be easier to identify and more accurate.

The flute, erhu, guzheng, drums, and pipa have been chosen for this paper because they comprise four forms of playing and because these instruments are more popular and familiar to the public. The following is a detailed description of the five instruments, including their history, pitch, tones, and features of their sound.

The flute is a wind instrument. It is mostly found in the south. The pitch and tone depend on the length of the flute. The thicker and longer flutes have a lower pitch.

An Erhu is a stringed instrument. This instrument has been around since the Tang Dynasty and has a long history. In the Ming and Qing dynasties, it became one of the most popular musical instruments and was often used in folk operas and ensembles.

Guzheng: A plucked instrument with a crisp sound and a wide range of tones, is one of the most important musical instruments in China and is very malleable in its musical style.

The drum is a percussion instrument with many different types. From ancient times to the present day, drums have been widely used, both in temples and on the battlefield, and are of great historical importance. They are characterised by their bass sound, which is deep and mellow.

The pipa is one of the outstanding plucked instruments, with a brighter and more penetrating tone than other instruments.

These are the statements of characteristics of these 5 instruments, and the reason that the author chose 2 instruments in the same way to be played is that the author wished they also had different results in classification. At the same time, using all types of ways to play can maintain the diversity of classify and get a better experimental result.

## **3. Methodology and experimental process**

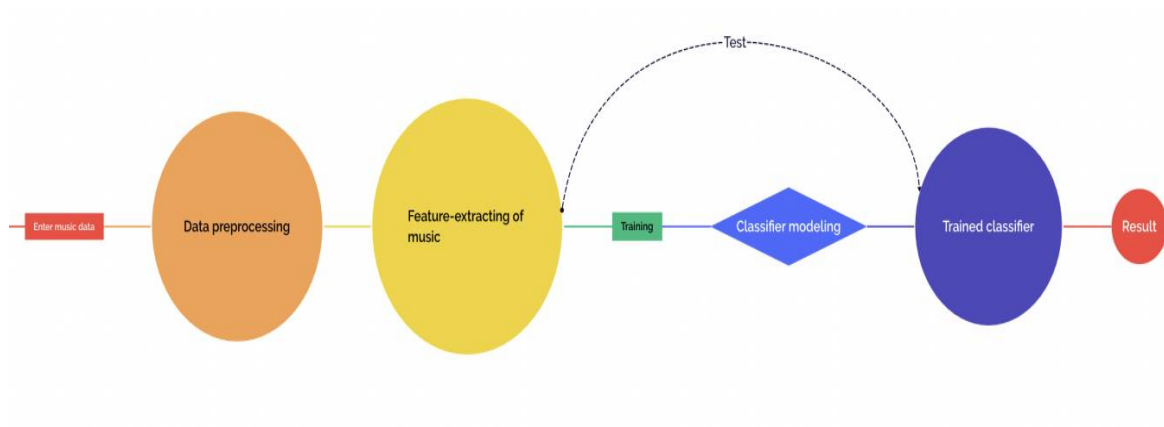
As with the process of music genre classification [4], the general classification process for Chinese folk instruments is as follows: feature extraction is performed first before achieving classification recognition.

(1) Preprocessing. As the music files from different data sources have different sampling frequencies and storage formats, they need to be preprocessed by noise removal and other work. Therefore, preprocessing is a very basic and important part that is related to the progress of subsequent experiments.

(2) Feature extraction. This is a crucial step in finding the key parameters that better fit the characteristics of the music signal. Feature extraction determines the classification effect to a large extent.

(3) Training the classifier. The training of the classifier is the key module that determines the final performance of the classification. The classifier is trained by constructing a training sample set to find the appropriate parameters to obtain a trained classifier.

(4) Testing of classification results. New music samples with predictions are put into the classifier for classification and prediction, and the classification accuracy is finally counted.



**Figure 1.** Flow chart for instrument identification.

### 3.1. Data set creation

The dataset used in this experiment was created by the author, based on the fact that Chinese folk instruments are divided into four types of play: blowing, pulling, playing, and hitting, so this experiment looked for music clips from five Chinese folk instruments on the internet: the flute, erhu, guzheng, drums, and pipa.

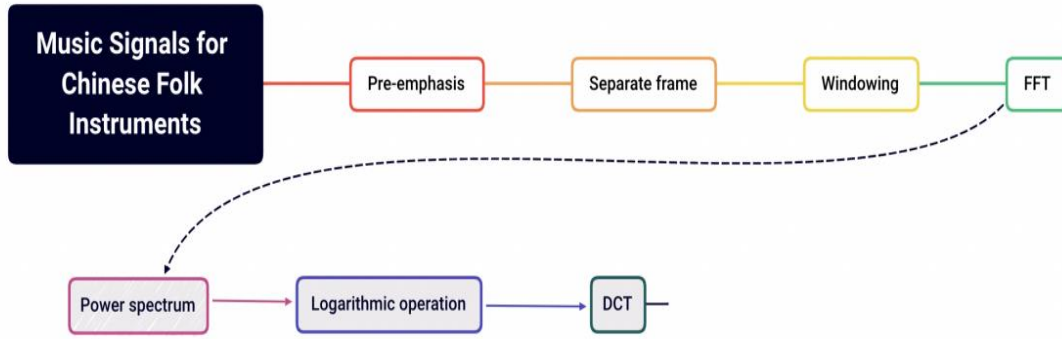
There are several important parameters in digital audio, including the number of bits sampled, the sampling frequency, and the bit rate. The music that was downloaded from the internet was sampled for the convenience of the tests that followed. The majority of audio on the internet is used in CD-like formats, and the most commonly used sampling rate for audio is 44.1 kHz, or 44,100 samples per second. This experiment uses the standard sampling frequency of 22.05 KHz and Nuendo software for audio processing. The three standard frequencies for sampling audio signals are 44.1 KHz, 22.05 KHz, and 11.025 KHz. Also, these music clips were preprocessed to remove useless information such as noise.

### 3.2. Feature extraction

As the Mel frequency cepstrum can represent the feature of timbre in music, the author intends to extract the Mel frequency cepstrum in Chinese folk instruments as a feature extraction. The Mel Inverse Spectral Coefficient [5] is a non-linear Mel scale based on sound frequencies. Compared to LPCC, it has better lupinness and is more in line with the auditory characteristics of the human ear. Secondly, the MFCC parameters are mainly expressed in the static characteristics of the music signal, while the dynamic characteristics of music can be described by static difference. The two complement each other to improve system performance. Its relationship with frequency can be expressed by the following equation [6]:

$$\text{Mel} = 2595 \log_{10} \left( 1 + \frac{f}{700} \right)$$

The process of extracting Mel frequency cepstrum coefficients for Chinese folk instruments is shown as follows:



**Figure 2.** The process of extracting Mel frequency cepstrum coefficients for Chinese folk instruments.

### 3.3. Classification method

In this experiment, four traditional classifiers—KNN, Softmax, SVM, and CART—were chosen to be used for the experiment. The KNN algorithm [7] is a very common traditional algorithm that can be used for classification, regression, dimensionality reduction, matrix decomposition, clustering, outlier detection, and so on. We need to find the K most adjacent data points to it in a fixed data set space, and then use the majority voting method to decide the prediction category. It is also known as a lazy learning algorithm, with no loss function and no associated parameters of its own.

For KNN-related formulae, assume that the data set has a feature sequence of  $\{a_1, a_2, a_3, \dots, a_n\}$ , P is the predicted data and A is the relative individual data.

Euclidean distance:

$$D(P, A) = \sqrt{(p_1 - a_1)^2 + (p_2 - a_2)^2 + \dots + (p_n - a_n)^2} = \sqrt{\sum_{i=1}^n (p_i - a_i)^2}$$

This is the most commonly used distance formula, other distance formulas include Manhattan distance, Minkowski distance, etc.

The Mel frequency cepstrum coefficients extracted from one of the music clips are fed directly into KNN to obtain a recognition correctness matrix. A recognition rate of 86% can be achieved.

Taking the same music sample and feeding it directly into the rest of the different traditional classification methods, it can be found that the decision tree has the lowest recognition rate of 75%. 85% is achieved by both Softmax and SVM.

## 4. Analysis of results

The following accuracy algorithm is used in this thesis:

$$\text{Accuracy} = \frac{\text{Correct identification of the number of music samples}}{\text{Total number of music samples tested}} \times 100\%$$

$$\text{Average Accuracy} = \frac{\text{Add up the accuracy for each instrument}}{\text{Total number of musical instruments}} \times 100\%$$

In this experiment, five Chinese folk instruments were trained and tested with different traditional classifiers, and the following table shows the average accuracy rates.

**Table 1** Average correct results of traditional classifiers for the identification and classification of five Chinese folk musical instruments.

Method	Average accuracy rate
KNN	91.7%
Softmax	88.4%
SVM	86.2%
CART	76.8%

Table 1 demonstrates the performances of five traditional Chinese instruments using several traditional classifiers. According to the testing findings, KNN is the most accurate in identifying and categorizing the five Chinese classical musical instruments, with an accuracy rate of 91.7 percent. However, decision trees performed the worst for the recognition and classification of musical instruments, with a low of 76.8%. The remaining two classification methods, Softmax and SVM, obtained similar accuracies of 88.4% and 86.2%, respectively.

## 5. Conclusion

People's neural network research is becoming more mature and in-depth. Thanks to the growth of digital music, music information retrieval has emerged as a significant research area. Numerous initiatives have been made to speed up and improve the retrieval process. In this research direction of music information retrieval, musical instrument classification is also one of the important areas, especially nowadays when there is less research in this direction of Chinese folk musical instrument classification, yet it is also very meaningful. In this thesis, the author tries to make a connection between music information retrieval and deep learning.

In this thesis, for this experiment, the author first gives a rough overview of the experimental process. Since the process of music genre classification is essentially the same as the process of instrument classification, this flowchart is available for both. Secondly, the author introduces the dataset used for this experiment as my own built music fragments about five different kinds of Chinese folk instruments, including guzheng, flute, erhu, and drums, which include four different playing styles. Furthermore, regarding the extraction of the original musical features, here the MFCC features are extracted. When compared among different traditional classifiers, KNN has the highest average accuracy of 91.7% for the classification of five Chinese folk musical instruments.

Although this paper compares the traditional methods and finds the optimal one among them, there are several areas that need to be improved due to my limited ability. The data set collected is small and the music clips are all monophonic, which is not complex enough. This thesis does not take the approach of combining two algorithms or constructing a new classification algorithm, which is not innovative enough, but only compares traditional algorithms. This is also a direction that can be researched and improved upon in the future.

## Acknowledgement

First of all, I would like to thank my father, who got me interested in studying music. I would also like to thank my mother for standing behind me and supporting me in the decisions I want to make. I would also like to thank my friends for believing in me and supporting me when I was facing difficulties. This has given me the motivation to move forward.

Secondly, I would like to thank my supervisors for helping me throughout the writing process of my dissertation and making it possible for me to complete this paper.

## References

- [1] HE Li, YUAN Bin. Classification of Music Genres Using Long and Short-term Memory Network[J]. China Academic Journal Electronic Publishing House, 2019(11): 190-191.
- [2] WANG Fang. Deep learning-based classification of musical genres and traditional Chinese musical instruments[D]. Wanfang Data, 2016:4-5.
- [3] KrishnaA G SreenivasT V Music instrumentrecognition: fromisolatednotes to solo phrases[C]. IEEEInternationalConferenceon Acoustics. 2004, 4:265-268.
- [4] SONG Yang, WANG Hailong,Liu Lin,Pei Dongmei.A review of research on music genre classification in the field of music information retrieval[J].Journal of Mongolia Normal University (Chinese version of Natural Sciences),2022,51(04):418-425.
- [5] Lu Huan. Music genre classification based on convolutional neural network[J]. Electronic Measurement Technology,2019,42(21):149-152.
- [6] HUANG Xun. Research on the classification method of opera based on deep learning[D].South China University of Technology,2019: 150.
- [7] Long, L. Jing. Extraction and study of timbre eigenvalues of western musical instruments[D]. China Academic Journal Electronic Publishing House, 2011:33.
- [8] Wang Bingcong. Research and implementation of a content-based automatic music genre classification system [D]. Wangfang Data,2018: 9-11.
- [9] Meng R. J. Research on music genre recognition based on improved AlexNet[D]. China Academic Journal Electronic Publishing House, 2020: 13-16.
- [10] Ye Hongliang,Zhu Wanning,Hong Lei. A music style conversion method with vocals based on CQT and Meier spectrum[J]. China Academic Journal Electronic Publishing House, 2021: 328-329.