Evaluation of the Combination of Features for Classifying Traditional Songs

May Thu Myint
University of Computer Studies (Hpa-an)
mthumyint@gmail.com

Phyu Phyu Khaing
Myanmar Institute of Information Technology, Mandalay

Abstract

The classification of music is a relatively research area and there are interesting areas for future explorations, including; in the feature extraction stage, and the use of classifiers that is a main idea of the active research. Although Myanmar's music has many similarities with other music styles in the district but the ethnic music styles differ depend on their cultural musical instruments. In this paper, the problem of music classification and highly similar of cultural music style of Myanmar's ethnic music is examined. The experiments are conducted by using the combination of timbre features and combining the nine major features. For this work, in the use of classification methods, Sparse Representation Classifier and k-Nearest Neighbours classifier are commonly used which is to compare the classification results. Moreover, it shows that MFCC (FC4, FC5, FC6) feature combination gives 79% of best classification result with the use of SRC classifier. When all feature combinations are used, the SRC provide the best classification accuracy of 81.64% for Shan ethnic songs than other ethnic songs.

Keywords: Sparse Representation Classifier, *k*- Nearest Neighbors, Timbre feature, Mel frequency Cepstral Coefficients

1. Introduction

With the rapid growth of the Internet and music multimedia technology, the amount of music data can be stored and it brings new changes and severe challenges. The development of automated music classification technology [2] plays a key role in the indexing and searching of music and helping to make it easier to manage a different type of music. Each country may have different types of folk music. From town to town, village by village this can be changed. Myanmar ethnic music similarly includes a variety of folk traditions. Different musical instruments are used in different style of ethnic music. Different local customs, local dialects, and living conditions have a great influence on the formation of folk songs' melodies.

According to geographic factors and named their study "Music Geography", Chinese ethnomusicologists have developed the division of folk songs of Chinese based on the characteristics [5]. In addition, research on classification of Myanmar ethnic songs will helpful in understanding the musical structure of ethnic songs, the way it is automatically analyzed these songs by using classification methods. However, the temporal structure of the melody is a key feature of folk songs.

In this paper, we proposed a performance evaluation based on sparse representation classifier which is to identify each ethnic class label. The main objectives are to improve the performance evaluation of SRC and to provide the best classification results, by combining the features. We first used various signal features for these purposes. Many audio classification problems involve data with high dimensional and noise. In [12], the author proposed with the SRC, the theoretical step to finding sparse representation is fast if the sparsest solutions are found. The system finds the optimal description of the music parts from the feature set in respect to more similar function defined in Sparse Representation Classifier (SRC) method.

2. Literature Review

The task of classifying folk music from different countries on the basis of monophonic melodies using hidden Markov models. Irish, German and Austrian folk music collections are used as datasets in various symbolic formats described in [10]. They tested and compared different representations and HMM structures. In this experiment, by using 6-state left-right HMM with the interval representation, the classification performances reached 75%, 77% and 66% for 2-way classifications and for 3-way classification reaches 63%. Therefore, some researchers usually separate the types of folk songs according to regional area. In current approaches, for music regional classification are similar to those for music genre classification [4, 7, 11], although there are many variations between folk songs and genre songs. In [8], the core goal is to explore the possibilities for using CNN in the retrieval of music information and to collect knowledge from the different musical patterns. Features such as statistical spectral patterns, rhythm and pitch derived from audio clips are less precise and produce less accurate models. GTZAN,

that consists of 10 genres for each of 100 audio clips, was used as the dataset for the study. The system got 84% of classified accuracy and eventually higher. The features extracted using CNN is good to get the more reliable result by comparing with MFCC.

3. Dataset and Experimental Setup

The audio collection consists of music by different singers, different categories of ethnic songs (Kayin, Kachin, Mon, Yakhine, Shan). All ethnic songs are performed with their cultural musical instruments and some other information such as sound produced by cultural people. The dataset collection includes 500 songs from the popular ethnic music songs in which there are 100 audio recordings of each ethnic group respectively. These songs are from Myanmar Radio and Television (MRTV) station. Each song lasts about 3 to 5 minutes. In all experiments, the whole songs are used for all evaluations. In all experiments, there are three main components, pre-processing, feature extraction and classification. The input audio signal (wav file) is resampled at 44100 Hz with 16 bits per sample. After the input audio is converted from stereo to mono and divided into 100 Ms frames, these frames of audio samples are taken as 50% overlapping of the successive frames. The major features of audio sample are extracted from the overlapped frame. The system was implemented by MATLAB programming language.

4. Research Methods

In music classification, many different types of audio feature extraction methods and different classifiers have been proposed on the tasks of traditional songs classification.

4.1. Feature Extraction

This section provides about the different features that have been extracted from the audio samples. In this system, features can be divided into time domain and frequency domain features. All of the features are 114 features in which 1-57 features are mean values and 58-114 features are standard deviation values of nine features (MFCC-mean, MFCC-std, MFCC delta-mean, MFCC delta-std, zcr, centroid, skew, kurtosis, bandwidth).

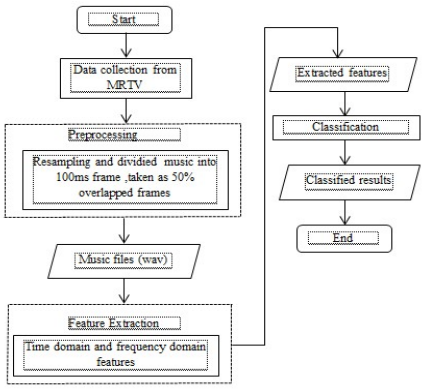

Figure 1. System Flow Chart

4.1.1. Feature Description

Feature extraction is the second step in most music classification systems. This section describes the core features which is described in table 1.

Table 1. Feature descriptions

No.	Feature Name	Descriptions
1.	ZCR	Mean (ZCR)
2.	ZCR-STD	Std(ZCR)
3.	CENTROID	Mean (Centroid)
4.	CENTROID-STD	Std(Centroid)
5.	SKEWNESS	Mean(Skewness)
6.	SKEWNESS-STD	Std (Skewness)
7.	KURTOSIS	Mean(Kurtosis)
8.	KURTOSIS-STD	Std(Kurtosis)
9.	BANDWIDTH	Mean (Bandwidth)
11.	BANDWIDTH-STD	Std (Bandwidth)
12.	FC1	Mean (MFCC-MEAN)
13.	FC2	Std (MFCC-MEAN)
14.	FC3	Mean (MFCC-STD)
15.	FC4	Std (MFCC-STD)
16.	FC5	Mean (MFCC-DELTA MEAN)
17.	FC6	Std (MFCC-DELTA MEAN)
18.	FC7	Mean (MFCC-DELTA STD)
19.	FC8	Std (MFCC-DELTA STD)

- **Zero Crossing Rate (ZCR)**: ZCR can be used as a statistical measure of spectral characteristics in music and speech recognition by analyzing the changes of ZCR over time. ZCR is also possible to differentiate between unvoiced and voiced speech components. The audio signal is divided into smaller frames, and zero-crossings number are determined in each frame. The mean and standard deviation of the ZCR across all frames are chosen as representative features.

$$Z_n = \frac{1}{2N} \sum_{m=n-N+1}^{N} | \, sgn[x(m)] - sgn[x(m-1)] \, | \quad (1)$$

where Z_n is the ZCR, sgn[x(n)] = 1 when x(n) > 0, sgn[x(n)] = -1, when x(n) < 0, and N is the number of samples in one window and m is the window size in this short-time function.

- **Mel Frequency Cepstral Coefficients:** Mel Frequency Cepstral Coefficients (MFCC) is a representation of the spectrum of an audio signal and

takes into account the nonlinear human perception of pitch. This is one of the most common features used to recognize speech and musical signals as well. A recent study [3] confirmed that the MFCCs is suitable for music description. The Mel scale approximates this relationship as shown in the following conversion between frequencies in Hz (f) and Mels (m);

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

$$\text{MFCC}_i = \sum_{k=0}^{N-1} X_k \cos[i(k+\frac{1}{2})\frac{\pi}{2}] \; ; \; i=0,1,\ldots,M\text{-}1$$

where M is the number of desired cepstral coefficients, N is the number of filters, and X_k is the log power output of the k^{th} filter. Each frame of signals in time domain is represented by 13 feature vectors.

- **Spectral Centroid:** centroid of the spectrum of the signal and is calculated as the weighted mean of the frequencies in the sound. Sharpness is related to the high frequency content of the spectrum, because the higher values of the centroid is corresponded to spectra skewed in high frequency range. Many types of music involve percussive sounds which increase the spectral mean higher by including high-frequency noise. As a result, music has a higher spectral centroid than speech [8]. The spectral centroid for a frame is computed as follows;

$$SC = \frac{\sum_k k \cdot X[k]}{\sum_k |X[k]|} \qquad (3)$$

where k is an index corresponding to a frequency within the overall measured spectrum, and X[k] is the power of the signal at the corresponding frequency band.

- **Skewness**: a measure of the asymmetry of the distribution of the probability of a truly random variable relative to its mean. The value of the skewness can be positive or negative or uncertain. For a single peak it is symmetrical or one-sided. If the curve is positive, the data is positively skewed or skewed to the right. This means that the right tail of the distribution is longer than the left tail. If the curve is negative, the data is negative or left to right. In other words, the left tail gets longer. If the curve is zero, the data is completely symmetric.

$$skewness = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{ns^3} \qquad (4)$$

where $\bar{x}$ is the mean value of the sample and s is the standard deviation of sample and n is the number of samples.

- **Kurtosis:** a measure of thickness or weight of the tail distribution for the random variable. If the number of data in the tail is greater, kurtosis is as positive, and if the number of data in the tails is less, in normal distribution, the kurtosis is as negative. The distribution can be divided into three types according to the value of kurtosis- The distribution of kurtosis is equal to 3 which is a normal distribution of 3. If kurtosis is less than 3, the tail of the split is shorter and thinner than the normal split.The peaks are smaller and wider than the normal distribution. If the kurtosis is greater than 3, the split tail is longer and wider than the normal split. The peaks are higher and faster.

$$kurtosis = \frac{\sum_{i}^{n}(x_i - \bar{x})^4}{ns^4} - 3 \qquad (5)$$

where $\bar{x}$ is the mean value of the sample and s is the standard deviation of sample and n is the number of samples.

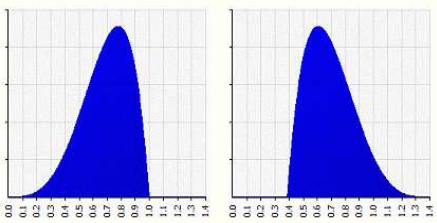

Figure 2. Positive skew Vs Negative skew

- **Bandwidth:** The bandwidth is used to indicate the frequency range between the lowest and highest frequencies that are reached when a certain level of signal strength is reached. An important feature of bandwidth is that any band of a given bandwidth can carry the same amount of information where that band is located in the frequency spectrum. Speech has typically accumulated 90% of its power at frequencies less than 4 kHz but music can propagate beyond the maximum ear's response at 20 kHz. In general, in music waveforms, most of the signal power is concentrated at lower frequencies.

4.2. Classification Methods

This section provides a brief overview of the two classifiers adopted in this study. To examine the classification performance of SRC and KNN, the dataset was divided into three parts by using 3-fold cross-validation. In all experiments, each part is used as a test set in turn, and the remaining parts is combined into a training set. We compare the SRC method with KNN. Accuracy is a performance metric used in our research. The three cross validations were repeated 10 times and take the average results for a fair comparison, the obtained three partition files are unified in all approaches, since it had a significant effect on the accuracy of the partition file obtained by cross-validation.

4.2.1. Sparse Representation Classifier (SRC)

In classification of the audio signal, the system needs to compare its features with known features contained in the feature set. The SRC classification algorithm is as follows: if there are sufficient training examples in each category, the test data is considered as a linear composition of the training data set belonging to the same category. For large data collections, the SRC classification is appropriate because system optimization is important. The SRC first encodes the query sample as a linear combination of several atoms from a predefined dictionary. It also determines the label by evaluating which of class results in the smallest reconstruction error. The SRC can accurately classify test sample outside the overlapped area, but in this region the accuracy is close to the random guess. The proposed method is called the Meta-sample-based SR classification (MSRC) in [1].

Table 2. Algorithm of Sparse Representation Classifier

Input: matrix of training samples $A=[A_1, A_2,\ldots,A_k] \in R^{m\times n}$ for k classes; testing sample $y \in R^m$

Step1: Normalize the columns of A to have unit L_0- norm.

Step 2: Extract the meta- samples of every class using NMF.

Step 3: Solve the optimization problem defined in

$$J(x,\lambda) = \min_x\{\| W_x - y \| + \lambda \| x \|_1\}$$

Step 4: Compute the residuals $r_i(y) = \| y - W\delta_i(x) \|_2$

Output: Identify = $\arg\min_i r_i(y)$

Consider a training gene expression dataset represented by an $m\times n$ matrix A with m genes and n samples. The n-dimensional vector r_q, i.e., the q^{th} row of A and the m-dimensional vector c_1 , i.e., the l^{th} column of A. Arranging the n^i samples of the i^{-th} class as a matrix $A_i = [c_{i,1}, c_{i,2},\ldots,c_{i,ni}]$, with each sample being a column. Given that the training samples of the i^{th} class are sufficient, any (testing) sample $y \in \mathbb{R}^m$ in the same class will approximately lie in the linear span of the training samples associated with class i. Suppose that the samples with the same class are conjoint, i.e., $A = [A_1,A_2,\ldots,A_k]$, then the linear representation of y as $y= Ax_0$. Finding the solution to solve SR problem, $J(x,\lambda) = \min_x\{\| Ax - y \|_2+\lambda\| x \|_1\}$ is considered which allows for certain degree of noise is minimized and is reduced to solving an l_1 -regularized least square problem. The positive parameter λ is a scalar regularization that balances the reconstruction error and sparsity. Then, they classify based on these approximations by assigning it to the class that minimizes the residual between y and $\hat{y}$. For this algorithm, it is used as trainSet: matrix, each column is a training sample, trainClass: column vector, the class labels for training samples, testSet: matrix, the test samples, testClass: column vector, the class labels of the test/unknown set, testClassPredicted: column vectors, the predicted class labels of testing samples, lamda: scalar, the parameter to optimization algorithm l_1, the default is 0.1, sparsity function: the sparsity of the sparse coefficient matrix and each sample has to be normalized to unit l_2 norm and we implemented the Sparse Representation (SR) toolbox in Matlab version, 1.5.

4.2.2. *k*- Nearest Neighbors

The first machine learning technique is *k* nearest neighbor (*k*-NN) [8], as is popular known by its simple use. *k*-NN is designed to be non- linear and can detect direct or indirect scattered information. The basic calculation of *k*-NN is to measure the distance between two songs. More precisely, for a specific feature vector in the target set, select the closest vectors in the training set. The target feature vector is the label of most of the neighbor's representations. KNN is the most common classifier that is, the training data is stored so that the classification for newly unclassified data is compared with the training data by taking the data of the most common training.

5. Results and Discussion

In all experiments, the proposed classification system is evaluated by using feature combination.

5.1. Evaluation of Combined Two Features

According to the table 3, the combination of MFCC (FC3, FC4) features are used on all of five ethnic songs in which the best classification results of 75.33% accuracy is obtained from SRC classifier than the result of 66.00% from KNN classifier. In this evaluation, the performance of SRC is achieved the best classification accuracy for all two combination of MFCC features. But, the combinations of (FC3, FC8) features give the lowest classification accuracy of 63.33% from SRC classifier and also 45.00% is achieved from kNN classifier for all ethnic songs. In these two feature combinations table, all feature combinations provide the best classification accuracy using SRC classifier than kNN classifier.

Table 3. Classification accuracy of combined two features

Whole Songs (All ethnic classes)		
Feature Combinations of MFCC	**SRC**	**KNN**
FC3, FC4	75.33%	66.00%
FC3, FC5	73.33%	45.00%
FC1, FC6	72.00%	59.66%
FC5, FC2	71.33%	49.67%
FC1, FC7	70.67%	57.67%
FC1, FC8	70.33%	59.66%
FC3, FC6	70.33%	58.00%

FC5, FC4	70.33%	61.00%
FC3, FC2	70.00%	43.66%
FC5, FC6	69.67%	64.67%
FC1, FC3	69.00%	54.00%
FC5, FC8	69.00%	58.67%
FC1, FC2	68.67%	65.00%
FC7, FC6	68.00%	54.67%
FC7, FC8	67.66%	53.67%
FC7, FC2	67.66%	58.67%
FC3, FC7	66.67%	57.67%
FC1, FC5	65.00%	53.33%
FC1, FC4	65.00%	57.00%
FC7, FC4	65.00%	53.67%
FC3, FC8	63.33%	45.00%

5.2. Evaluation of Combined Three Features

According to table 4, the only of all feature combinations of MFCCs are used in which the better accuracy of 79.01% is obtained from SRC than the classification accuracy of kNN classifier. The combinations of MFCC (FC5, FC4, FC6) and MFCC (FC2, FC4 FC6) are the best features in this evaluation. But, the performance of SRC is decreased to the lowest accuracy of 64.00% by using MFCC (FC3, FC4, FC6) features. With the exception of this features combination (FC3, FC4, FC6), SRC classifier can provide the best results over kNN. According to the feature combinations (FC3, FC2, FC8), the accuracy of kNN is significantly dropped to 46%. In summary, the evaluation of three feature combinations, SRC give the better accuracy than kNN.

Table 4. Classification accuracy of combined three features

Whole Songs (All ethnic classes)		
Feature Combinations of MFCC	**SRC**	**KNN**
FC5, FC4, FC6	79.01%	72.33%
FC2, FC4, FC6	78.61%	52.66%
FC5, FC4, FC8	78.00%	67.00%
FC5, FC6, FC8	77.00%	57.00%
FC7, FC6, FC8	77.00%	53.00%
FC3, FC4, FC8	76.67%	68.67%
FC7, FC4, FC6	76.67%	66.66%
FC7, FC2, FC4	75.67%	61.66%
FC3, FC5, FC7	74.67%	56.67%
FC3, FC2, FC4	74.33%	53.33%
FC2, FC4, FC8	73.84%	57.00%
FC1, FC2, FC4	73.33%	66.33%
FC5, FC2, FC4	73.33%	59.00%
FC3, FC2, FC8	73.00%	46.00%
FC5, FC2, FC6	73.00%	53.67%
FC7, FC4, FC8	72.33%	66.66%
FC1, FC2, FC6	71.67%	64.67%
FC5, FC2, FC8	71.67%	48.00%
FC3, FC2, FC6	71.66%	49.67%
FC1, FC4, FC6	71.00%	52.67%
FC1, FC6, FC8	70.44%	62.33%
FC7, FC2, FC8	69.66%	52.67%
FC1, FC3, FC5	68.67%	60.66%
FC3, FC6, FC8	67.33%	46.67%
FC1, FC2, FC8	67.00%	65.00%
FC1, FC4, FC8	67.00%	57.00%
FC7, FC2, FC6	67.00%	59.33%
FC1, FC3, FC7	66.33%	55.00%
FC1, FC5, FC7	64.47%	54.33%
FC3, FC4, FC6	64.00%	64.36%

5.3. Evaluation of Combined Four Features

In this evaluation, SRC classifier give the better accuracies of 76.17% and 75.54% by combining MFCC (FC7, FC4, FC5, FC6) features and (FC3, FC1, FC4, FC8) features. With the exception of the combination (FC5, FC1, FC3, FC7) features, SRC classifier provide the best results than kNN classifier. According to the table 5, kNN classifier give the lower accuracy of 59.33% by combining MFCC (FC6, FC5, FC7, FC8) features.

5.4. Evaluation of Combined Five Features

In table (6), the combination of (FC5, FC7, FC1, FC3, FC4) features, SRC is obtained the accuracy of 77.08% which is better than the accuracy of kNN. In these combinations of five features, SRC give the better accuracy than the accuracy of kNN classifier. Obviously, the classification accuracy of kNN is decreased to 54.00% when using the combination of MFCC (FC1, FC3, FC5, FC7, FC6) features.

Table 5. Classification accuracy of combined four features

Whole Songs (All ethnic classes)		
Feature Combinations of MFCC	**SRC**	**KNN**
FC5, FC7, FC4, FC6	76.17%	67.66%
FC1, FC3, FC4,FC8	75.54%	60.66%
FC5, FC7, FC4, FC8	74.40%	62.66%
FC5,FC7, FC2, FC4	72.79%	65.66%
FC1, FC3,FC, FC8	72.08%	58.66%
FC1, FC3, FC4, FC6	71.94%	60.00%
FC1, FC3, FC6, FC8	71.58%	64.00%
FC1, FC3, FC2, FC6	70.95%	64.00%
FC1, FC3, FC2, FC4	70.13%	66.00%
FC5, FC7, FC2, FC8	69.85%	57.00%
FC5, FC7, FC2, FC6	68.45%	59.33%
FC5, FC7, FC6, FC8	64.81%	59.33%
FC1, FC3, FC5, FC7	61.38%	55.66%

Table 6. Classification accuracy of combined five features

Whole Songs (All ethnic classes)		
Feature Combinations of MFCC	**SRC**	**KNN**
FC1, FC3, FC5,FC7, FC2	77.08%	68.66%
FC1, FC3, FC5, FC7, FC6	75.44%	54.00%
FC1, FC3, FC5, FC7, FC8	72.87%	60.00%
FC1, FC3, FC5, FC7, FC8	67.95%	61.00%

5.5. Evaluation of the Best Combined Features

According to all evaluations, SRC obtained the best classification results than the results of kNN classifier by using the combination of MFCC (FC4, FC6, FC5) and MFCC (FC4, FC8, FC5) features. Then, the results of SRC is decreased to 75.44% in the use of combination of five features. But, in all evaluations, SRC classifier provide the better accuracy than the accuracy of kNN. In summary, the combination of MFCC (FC4, FC5, FC6) feature is the best MFCC combination features in all evaluation of SRC classifier.

5.6.Evaluation of Combined All Features

Table 7 and fig.3 show the classification accuracy of all features (114) which is the combination of time domain features and frequency domain features. According to each ethnic class, SRC classifier provide the highest classification result for all Shan ethnic songs than other ethnic songs. By using SRC classifier, Kayin and Rakhine ethnic songs achieve the better accuracy than the use of kNN classifier. But the classification accuracies of Kachin and mon ethnic songs are significantly decreased by using SRC than the kNN classifier. In summary, the only use the MFCC features give the better classification accuracy instead of using all features (114).

Table 7. Classification accuracy of combined all features (114)

Classification Accuracy (%)		
Ethnic classes	**SRC**	**KNN**
Shan songs	81.64%	58.33%
Yakhine songs	75.00%	70.00%
Kayin songs	71.66%	58.33%
Mon songs	66.68%	60.67%
Kachin songs	65.00%	58.33%
All	72.00%	61.13%

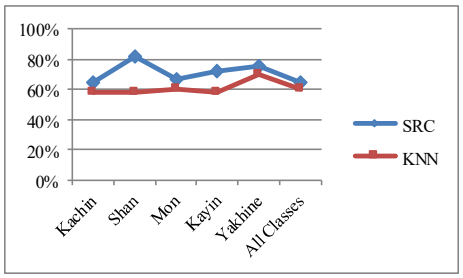

Figure 3. Chart of combinations of all features

6. Conclusion

In Myanmar, the language and culture differ based on the geographical area, hence the Myanmar ethnic group's music also varying based on the geographical area. In this work, the main objective is to achieve the better classification accuracy by using the conventional auditory feature analysis and to get the best outcome by calculating all the results based on SRC. Then, the system was developed among five ethnic music classes; all of the songs have cultural styles that are played with their respective traditional instruments. The obtained results have clearly shown that SRC is the best classifier for the classification of Myanmar's ethnic music when compared to *k*NN. In conclusion, the system is developed to analyse the influence of SRC upon the timbre features and also the future develop could also be performed by using another classification methods and other feature extraction methods can be used to get the better features.

References

[1] C.H. Zheng, L.i Zhang, T.Y. Ng and C. K.Shiu, " Metasample based Sparse Representation for Tumor Classification", Poletechnic University, Hong Kong, Chinal, 2011.

[2] Fu ZY, Lu GJ, Ting KM, Zhang DS , "A survey of audio-based music classification and annotation", IEEE Trans Multimedia 13(1):303–319, 2011.

[3] G.Tzanetakis,G. Essl, and P. Cook. "Automatic Musical Genre Classification of Audio Signals", In Proceedings of the International Symposium on Music Information Retrieval (ISMIR), Paris, France, 2002.

[4] Huang YF, Lin SM, Wu HY, Li YS, "Music genre classification based on local feature selection using a self-adaptive harmony search algorithm", Data Knowl Eng 92:60–76, 2014.

[5] Huang YF, Lin SM, Wu HY, Li YS, "Music genre classification based on local feature selection using a self-adaptive harmony search algorithm", Data Knowl Eng 92:60–76, 2014.

[6] Kereliuk C, Sturm BL, Larsen J, "Deep learning and music adversaries", IEEE Trans Multimedia 17(11):2059–2071, 2015.

[7] Nanni L, Costa YMG, Lucio DR, Silla CN Jr, Brahnam S , "Combining visual and acoustic features for audio classification tasks", Pattern Recogn Lett 88:49–56, 2017.

[8] S.Jothilakshmi and N.Kathiresan," Automatic Music Genre Classification for Indian Music , Department of information technology, International Conference on Software and Computer Applications (ICSCA), 2012.

[9] Tom LH Li, Antoni B Chan, and A Chun, "Automatic musical pattern feature extraction using convolutional neural network", In Proc. Int. Conf. Data Mining and Applications, 2010.

[10] W. Chai and B. Vercoe, " Folk Music Classification Using Hidden Markov Models", USA, 2009.
[11] Wu MJ, Jang JSR, “Combining acoustic and multilevel visual features for music genre classification”, ACM Trans Multimed Comput Commun Appl 12(1):1–17, 2015.
[12] Y.Yang , J. Wright, Y.Ma , and S. Sastry, “Feature selection in face recognition: A sparse representation perspective,” UC Berkeley Tech Report UCB/EECS-2007-99,2007.