Received: March 9, 2017

INASS

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Neural Network Based Indian Folk Dance Song Classification Using MFCC and LPC

Malay Bhatt[1]* **Tejas Patalia [2]**

[1]*Rai University, Ahmedabad, India*
[2]*V.V.P. Engineering College, Rajkot, India*
* Corresponding author's Email: malaybhatt202@yahoo.com

Abstract: A large number of folk dance videos are uploaded on the web or added as a situational song in the Bollywood movies. The classification of folk dance videos is essential for dance education, to preserve cultural heritage, and for music companies to provide better customer oriented service. India is a country having many regional languages and each region has its own popular folk dances. Four different Indian folk dances namely, '*Garba*', '*Lavani*', '*Ghoomar*' and '*Bhangra*' are considered. A Folk Dance Classification Framework is proposed which extracts audio signal from video, takes a fragment of 125 seconds from the beginning and further separates it into a set of small segments, calculates Mel-frequency Cepstral Coefficients (MFCC) and Linear Predictive Coding (LPC) coefficients, generates high dimensional feature vector, reduces dimensionality using Principle Component Analysis (PCA) and classifies segments using Scale Conjugate Gradient Neural Network. The performances of chosen classifiers, K-Nearest Neighbor, Naïve Bayes and Neural Networks, are compared. Class labels of all segments are clubbed together and based on majority voting class label is assigned to a folk dance song. System achieves more than 90% accuracy.

Keywords: Classification; Folk dance song; Mel-frequency Cepstral Coefficients; Neural network; Linear predictive coding

1. Introduction

The human generated tags are used to categorize and describe the living and non-living things of the vast universe. The tags used for dance are known as 'Dance Genre'. Dance song comprises of purposely selected sequences of human movement accompanied with vocal and background music. Folk music and folk dances constitute a significant part of the folk heritage around the world. The preservation of the folk music and choreographies and their dissemination to the younger generations is a very important issue since folk dance forms an important part of a country's or region's history and culture[1]. Many years ago, classification of dance genre was purely manual through conversation with public and experts' opinion. Automatic dance genre classification is essential because of resource digitization and plenty of new audio and video songs are added every year. Currently, most of the popular sites provide meta-data or text based annotation Automatic dance song analysis will be one of the services used by music distributors to catch the attention of music lovers and to achieve maximum profit. More than a decade, researchers are paying attention to audio signal analysis and classification.

The earliest civilizations discovered in the Indian subcontinent are those of Mohenjo Daro and Harappa in the Indus valley, and are dated about 6000 B.C. It seems that by that time dance had achieved a deemed measure of discipline and it is likely that it must have played some important role in the society, Two beautiful little statuette of bronze dancing girls were found at Mohenjo Daro in 1926 and 1931 respectively [2].

There are mainly two popular dances in India: Classical Dance and Folk Dance. Classical Dance was practiced in courts, temples and on special

occasions. Folk Dance in India is a term broadly used to describe all forms of folk and tribal dances in regions across India. Folk dance forms are practiced in groups in rural areas as an expression of their daily work and rituals. Folk and tribal dances are performed for every possible occasion, to celebrate the arrival of seasons, birth of a child, a wedding and festivals [3]. The dances burst with verve and vitality. Mostly, Men and women perform some dances exclusively, while in some performances they dance together. On most occasions, the dancers sing themselves, while being accompanied by artists on the instruments. Some of the popular folk dances that are performed across villages and cities are *'Bhangra', 'Garba', 'Lavani', 'Ghoomar', 'Kalbelia'* and *'Bihu'* [4]. An effect of various factors, along with comparative analysis, on automatic Indian music recognition, classification and retrieval is described in [5]. The impact of artist (singer) significantly affects the genre based classification of dance songs. The songs sung by same artist in both training and testing results in over optimistic accuracy. The Artist filters for genre based song classification are discussed in [6]. We have also collected dance songs of same genre of different artists to eliminate the effect of artist.

Lavani:

Lavani is a popular folk form of Maharashtra. Traditionally, the songs are sung by female artists, but male artists may occasionally sing Lavani. The dance format associated with Lavani is known as 'Tamasha'. This dance format contains the dancer ('Tamasha Bai'), the helping dancer - Maavshi, The Drummer - 'Dholki vaala' & The Flute Boy - 'BaasnuriVaala' [3].

Bhangra:

Bhangra is a form of dance-oriented folk music of Punjab. The present musical style is derived from non-traditional musical accompaniment to the riffs of Punjab called by the same name [3].

Garba:

Garba is a folk dance of Gujarat. It has been adopted all over the world. A garba song has lot of music in it. The traditional music is composed of drums, flute, etc. It is more in music than in vocal [3].

Ghoomar:

Ghoomar is a folk dance of Rajasthan. It has also been adopted widely in Bollywood movies. It has lot of music in it [3].

Dance and song sequences have been an integral component of films across the country. With the introduction of sound to cinema in the film 'Alam Ara' in 1931, choreographed dance sequences became ubiquitous in Hindi and other Indian films [7].

Often in movies, the actors don't sing the songs themselves that they dance too, but have another artist sing in the background. For an actor to sing in the song is unlikely but not rare. The dances in Bollywood can range from slow dancing, to a more upbeat hip hop style dance. The dancing itself is a fusion of Indian classical, Indian folk dance, Indian tribal, belly dancing, jazz, hip hop and everything else. Video song can be extracted from Indian movies or can be downloaded from popular site YouTube which gives large number of songs rapidly. It is difficult to identify the folk dance form for a normal user because it pertains to particular region. Folk dances have cultural differences too. Classification of a dance form can be done using audio and/or visual cues.

Folk dance classification based on visual cues involves large amount of video processing like key-frame extraction, shot boundary detection, Moving object (Dancer) detection, feature based posture identification, recognition and classification. Visual cues based classification is purely affected by how the actual dance song is choreographed and how it is picturized by cameramen. Many choreographers and cameramen prefer to take different kinds of shots like Top View, Side View, Long Shot, Close Up or Circular motion. It is very difficult to recognize the actual dance move under such different camera situations. When words are important at that time only Close Up shot is considered so it is not possible to detect the posture. While folk dance classification based on audio cues involves less processing of data, fast recognition with good accuracy. Figure 1 shows (a) Ghoomar and (b) Garba shot.

Indian movies consist of 2 main events: Non-Song and Song. Audio part can be easily separated from the extracted video song. Each dance form has its own non-vocal and vocal part. Non-vocal part comprises of instruments which are widely used in that specific state/region. Vocal part also differs from region to region.

Classification based on the vocal part is possible if dance forms belong to different region having its own Natural Language. Natural Language Processing gives a necessary clue regarding the region and possible dance form but has the limitation that it cannot recognize different folk dances of the same region. Non-vocal segment distinguishes different folk dances of the same region as proportionate use of instruments varies significantly.

Figure.1 Indian folk dance: (a) Rajasthani Ghoomar and (b) Gujarati Garba

Table 1. Popular Folk Dances in Bollywood Movies

Movie Name	Song Title	Song Type
Bajirao Mastani(2015)	Pinga	Lavani
Aiyaa	Sava Dollar	Lavani
Agneepath (2012)	ChikniChameli	Lavani
Ferrari ki Sawaari (2012)	Mala Jau De	Lavani
Inkaar (1978)	MungdaMungda	Lavani
Singham Returns (2014)	Aata Majhi Satakli	Lavani
Sailaab (1990)	Humko Aajkal Hai Intezaar	Lavani
Lagaan (2001)	RadhaKaise Na Jale	Garba
Kai poChe (2013)	Shubharambh	Garba
Jai Santoshi Maa (1975)	Jai Santoshi Maa	Garba
Hum Dil De Chuke Sanam (1999)	Dholi Taro Dhol Baje	Garba
Ramleela (2013)	Nagada sang dhol	Garba
AA Ab Laut Chalein (1999)	Yehi hai Pyaar	Garba
Jab we met (2007)	Nagadanagada	Bhangra
Rang De Basanti (2006)	Rang De Basanti	Bhangra
Love Aaj kal (2009)	Ahunahun	Bhangra
Mujhse Shaadi karogi (2004)	AajaSoniye	Bhangra
Band Baaja Baaraat (2010)	Ainvayi Ainvayi	Bhangra
Paheli (2005)	Laaga Re Jal Laaga	Ghoomar

Figure 2 provides a flow chart of a content based retrieval system. An audio query is the audio file that is given as an input to the system. The features of the input audio are calculated. A query of the extracted features is then generated and is compared with all the other features of the audio files present in the database. Based on the similarity measures the system retrieves the required audio files from the data base and presents it in the form of the result.

The remainder of this paper is organized as: Section 2 covers literature review, Section 3 gives a detailed view of the proposed system Folk Dance Song Classification, Section 4 discusses dataset and experimental setup, Section 5 describes results and comparison among chosen classifiers and Section 6 concludes our work.

2. Literature survey

Kapsoras et. al. considered classification of Greek folk dances of Western Macedonia region as sub activity of human action recognition and it comes under computer vision. Creation and Preservation of folk dance database is beneficial for research and cultural heritage and also connects the youth to their roots. Risks and precautions associated with each folk dance are also stored altogether that makes it easy for new learners and thus useful in education. Challenges of Greek dance are: most of the folk dances are group performance, long skirt that hides legs etc. are highlighted. They considered various video songs of 2 popular folk dances, namely *Lotzia* and *Capetan Loukas.* Extended Independent Subspace Algorithm and Space-Time Intereset Points (STIPs) are used for feature extraction. K-means clustering is incorporated on extracted features for codebook generation. Histogram over the codebook is constructed for each sequence using vector quantization. Support Vector machine is used with chi-square kernel on the histogram for classification. Highest accuracy 89 % is achieved with STIPs[8].

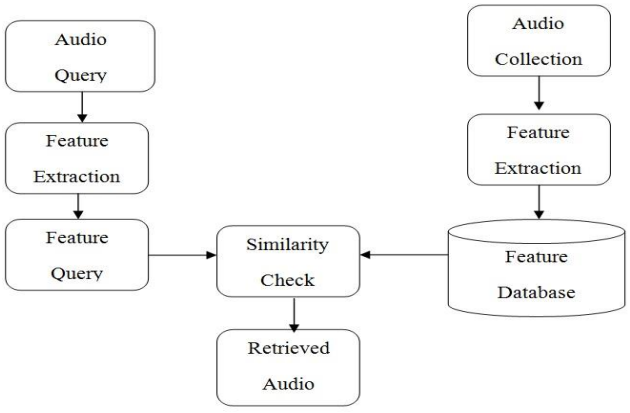

Figure.2 Content Based Audio Retrieval System

Samanta et al. in [9] build up a method based on visual features for classification of Indian Classic Dances, namely Kathak, Bharatnatyam and Odissi . The authors propose a pose descriptor which is a combination of Gradient and Optical Flow. The descriptor of 168 dimension is calculated in a hierarchical manner to represent each frame of a sequence is based on histogram of oriented optical flow, The pose basis is learned using an on-line dictionary learning technique and each video is represented sparsely as a dance descriptor by pooling pose descriptor of all the frames. Finally, dance videos are classified using support vector machine (SVM) with intersection kernel with an average accuracy of 83.33% for Bharatnatyam, 90% for Kathak and 86.6% for Odissi

Kaynak et. al. developed Audio-Visual speech recognition using lip geometric features which includes horizontal, vertical lip aperture and first - order derivative of the lip corner. They classified 20 hours of audio-video speech database of isolated utterances from 22 nonnative English speakers using Hidden Markov Model. This kind of system have fixed camera position and mostly static background for speech recognition[10].Extraction of dialogue and action scene from movies using audio cues is investigated in [11].

Song et.al. in [12] have proposed a system for classifying the audio information into four different classes: speech, non-speech, music and non-pure speech. For the purpose of feature extraction, 8000 Hz is used as a sampling frequency and the frame size used is 32 milliseconds which is equal to 256 samples per frame (0.032*8000=256). As well as, 25% overlapping is considered in each of the two adjacent frames. The features like low short time energy ratio (LSTER) and high zero crossing rate ratio (HZCRR), Bandwidth and Noise Rate were extracted from the audio and 'Decision Tree' based classification is carried out.Reasonable classification accuracy is achieved on 12000 audio clips in 'wav' format having total duration 200 minutes. The database contains 1236 music clips, 5935 pure speech clips, 1793 silence clips and 3036 non-pure speech clips. Highest accuracy achieved is 95% for silence and lowest accuracy achieved is 88.8% for non-pure speech.

Silla, Jr., Carlos N., Celso, A., A Kaestner, and Alessandro L. Koerich in [13] have worked for an automatic genre classification system using ensemble classifiers, and is based on extraction of multiple feature vectors from a single music piece. The 3 segments are extracted as: first 30-seconds from beginning, one from middle and one from the end part. Features from the 30 second segments are extracted with 1153 audio samples per second.

Zhang, Tong, and CC Jay Kuo in [14] have developed a hierarchical system for audio classification and retrieval. The system describes that generally the audio features are divided into 2 categories: Perceptual and Physical. The perceptual features are subjective and are related to human perception. For example pitch, timbre, loudness, etc. Physical features refer to the features that can be mathematically calculated using the equations. For example the spectrum, ZCR, etc. Their system consists of three stages. The First stage is known as coarse-level audio classification and segmentation in which samples are classified and segmented into speech, music, environmental sound and silence. Environmental sounds are further classified in the second stage using the Hidden Markov Model. Third stage implements query by example audio retrieval.

Tzanetakis et.al.in [15] focused on music genere classification of audio signals. Audio classification hierarchy is proposed in which at level 0 audio signal is considered, at level 1 music and speech are separated, at level 2 music is further classified into classical, country, disco, hiphop, jazz, rock, blues, reggae, pop and metal while speech is separated into male, female and environment(spots etc.). Classical and Jazz music is further classified at level 4. Software used by them is a part of MARSYAS which is available under the GNU Public License. They achieved 61% accuracy for 10 musical genres and timbral texture, rhythmic content and pitch content are proposed as freatures. Similarly, Nopthaisong et.al. performed Thai music classification based on audio cues in [16].

Chu et. al. contributed for environmental sound recognition using time -frequency audio features. Environmental sound recognition is essential for understanding of scene or context surrounding the audio sensor. They proposed Matching Pursuit(MP) algorithm to obtain effective time-frequency features. Matching Pursuit generates flexible, intuitive and physically interpretable set of features using a dictionary of atoms. MP- based feature is adopted to supplement the MFCC features to achieve higher recognition accuracy. They considered 14 different environmental sound and used k-Nearest Neighbour classifier and Gaussian Mixture Model classifier. Average accuracy achieved for 14 classes is 83.9% and for 7 classes accuracy is higher than 90 %. Highest accuracy achieved with kNN (K=32) is 77.3% [17].

Palecek, K., and Chaloupka, J. focused in [18] on the audio-visual speech recognition under environmental noise. The White and the babble

noise are added in the audio signal. Discrete Cosine Transform and Active Appearance Model(AAM) based features are identified and extracted from video signals. These features are then enhanced through Hierarchical Linear Discriminant Analysis and normalized using z-score approach for all speakers. MFCC based audio features are clubbed with visual features using middle fusion approach. Only Visual cues based speech recognition is affected by many factors like high computational cost, speaker's appearance, lighting condition, camera's relative position, moving or stationary camera etc. They also mentioned that speech recognition rate is very low for very small vocabularies based on visual cues only. Audio visual speech database having recordings of 35 speakers in Czech language is used. Word recordings of 30 speakers are used for training and word recordings of remaining 5 speakers are used for testing. Digital web Camera is kept in front of speakers. They achieved 79.2 % recognition rate by combining AAM based features with Audio features. They carried out 3 experiments: 14 -states whole word Hidden Markow Model for audio only and 14-state whole word Hidden Markow Model for visual only experiments while 2-stream 14 state whole word HMM for combined audio-visual speech recognition.

Chen et al highlighted mixed type audio classification using support vector machine. They took 1 hour of movie database and segmented into 5 second segments. These five second segments are further windows into 1 second and classified into five different classes: Pure Speech, Pure Music, Environment Sound, Speech with music and music with environment sound. They took Variance of zero crossing rate, Silence Ratio, Harmonic Ratio and Sub-band energy as features. They also considered four different representation for each feature: min, max, (min +max)/2 and mean. It is also observed that variance of zero crossing rate is non-linearly separable and hence support vector machine (one against all approach) is selected for classification with gaussain kernel. They experimented with different values of error penalty (C) and 'sigma'. They found that SVM outperforms as compared to kNN, NN and Naive Baysian approaches. Because of the broad spectrum of environment sound (opening of door, sound of footsteps, sounds of nature and sounds of animals) clubbed with music highest accuracy achieved is 78.649%[19].

Li, T., and Tzanetakis, G. focused on factors like Timbral Texture, Rhythmic Content Features and Pitch Content features for classification of audio signals. For pairwise comparison, LIBSVM is used by authors for multi-class classification and Multi category Proximal Support Vector Machines are used for multi-objective functions for classification. Best classification result is obtained by Linear Discriminant Analysis on full feature set having 30 features which includes MFCC, FFT, Pitch and Beat[20].

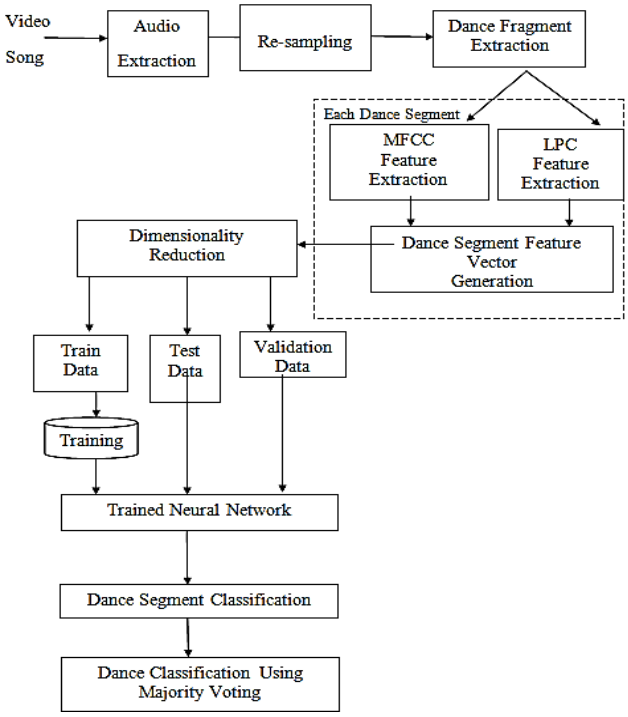

Figure.3 Proposed Framework for Folk Dance Classification

3. Proposed framework

Folk dance classification framework is depicted in Fig.3. It consists of mainly 9 phases mentioned above.

3.1. Audio extraction

Plenty of folk dance videos are available on various sites on the internet. Folk dance videos for '*Garba*', *'Lavani'* ,*'Bhangara'* and '*Ghoomar*' are downloaded and audio part is separated from the video part. 40 different dance videos are collected for each class. Total 160 videos are collected from various web sites.

3.2. Resampling

It is observed that audio files obtained from the previous phase having different sampling rate. It is essential for further processing that all audio files must have same frame rate. All extracted audio files

are resampled to 24Kbps frame rate (24000 bits per second).

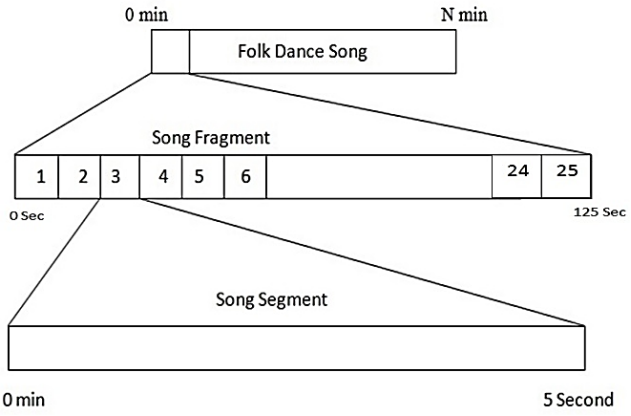

Figure.4 Dance Fragment Extraction and Dance Segments Generation

3.3. Dance fragment extraction

It is not necessary to analyze the whole audio file for the classification of the folk dance. It is found that normally duration for dance is between 5-8 minutes. We have taken initial 125 seconds of each dance for processing. Extraction process is described in Fig.4.

Let's assume that fragmented audio files each having size of 125 seconds are numbered from D1,D2,.....,DN. In our case N is 160.

Each Segmented audio file is further decomposed into a portion of 5 seconds. Let's consider D(i, j) denotes jth portion of ith dance file.

Total number of portions in each dance file = 125/5 = 25.

$$Totalportion = \sum_{i=1}^{N}\sum_{j=1}^{M} D(i,j) = 4000 \quad (1)$$

In our case, N=160 and M=25.

The base of any automatic analysis system is the feature vectors. Large number of time and frequency domain features have been originated for automatic music recognition.

3.4. MFCC feature extraction

MFCC are used widely for speech recognition, singer identification etc. Audio Song classificaiton based on listeners' taste is using MFCC is explained with 82% accuracy in[1]. Wei, Zhang, et al. in [21] have developed a system for an audio classification using a sampling rate of 22.020 KHz with MFCC and other features. The accuracy of the system was recorded more than 77% .

Each extracted segment of the song fragment is taken as input for MFCC feature extraction. Number of samples in each portion are 5x24000=120000. We have selected frame size of 15(15x24 samples) millisecond and frame shift is considered as 10 (10x24 samples)millisecond for analysis. In total 13 MFCC features are extracted. Pre-emphasis for MFCC is taken as 0.97.

Let's denote the feature matrix obtain for each segment of the dance fragment as PxK. Here, P represents total number of overlapping frames in one portion and K is MFCC coefficients. Here K is 13 and value of P is 499.

MFCC feature extraction involves following steps:

1.**Tw**=15 ms
2. **Ts**=10 ms
3.No_of_Frames= round**(1E-3** * **Tw** * **fs**)
4.No_of _Frame_shift= round(**1E-3 * Ts** * **fs**)
5..Preemphasis:
speech(t)=1 - α x speech(t-1);
Here, α = 0.97
6. For each frame
1. Compute Discrete Fourier Transform (DFT)
2. Take log of Amplitude of DFT
3. Perform Mel Scaling and Smoothing using triangular filterbank having 26 filters equally spaced between lower and upper frequency.
Mel_freq=1127 x log(1+freq/700)
4. Perform Discrete Cosine Transform (DCT)
5. Take initial 13(N) cepstral coefficients.
6. Perfrom Cepstar Lifting
Cep_lifter= 1+0.5 * **L*** sin(**pi** * [0:N-1]/**L**)

Here, L is sine lifter parameter with value =22;
N is cepstral coefficients

3.5. Linear predictive coding (LPC)

Linear predictive coding(LPC) is defined as a digital method for encoding an analog signal in which a particular value is predicted by a linear function of the past values of the signal. It was first proposed as a method for encoding human speech by the United States Department of Defense in federal standard 1015, published in 1984[22].

At a particular time, *t*, the speech sample *s(t)* is represented as a linear sum of the *p* previous samples.

LPC coefficients are computed using the below given equation

$$s(t) = 1 - a(2)s(t-1) \ldots - a(p+1)s(t-p) \quad (2)$$

Here, p is the order of the polynomial and A=[1 a(2) a(3) .... a(p+1)].

The most important aspect of LPC is the linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. Normally, speech is sampled at 8000 samples/second with 8 bits used to represent each sample. This provides a rate of 64000 bits/second. Linear predictive coding reduces this to 2400 bits/second. At this reduced rate the speech has a distinctive synthetic sound and there is a noticeable loss of quality. However, the speech is still audible and it can still be easily understood. Since there is information loss in linear predictive coding, it is a lossy form of compression. In Proposed framework, 13 coefficients are extracted with value of P =12(12th order polynomial).

3.6. Dance segment feature vector generation

Vector is generated by combination of 13 MFCC coefficients with 13 LPC coefficients. Each song segment of 5 second comprises of 499x26 size vector. Pictorial view of process is in Fig.5.

Each dance fragment (part of Song) is a combination of 25 segments each of size 499x26. This way size of dance fragment is 25x499x26 which is equivalent to 12475x26. Principal Component Analysis is a widely used method for dimensionality reduction [23][24]. Principal components are achieved by a linear transformation to a new set of features which are uncorrelated, ordered, arranged as per importance and also retains as much possible variance. Singular Value Decomposition method is used for extraction of principle components. Dimensionality reduction process converts 12475x26 vector into 2000x26 vector is depicted in Fig.6.

3.7. Dimensionality reduction

Figure 6 illustrates the dimensionality reduction for a song fragment.

3.8. Dance song segment classification

Classification accuracy is compared using 3 different approaches described here.

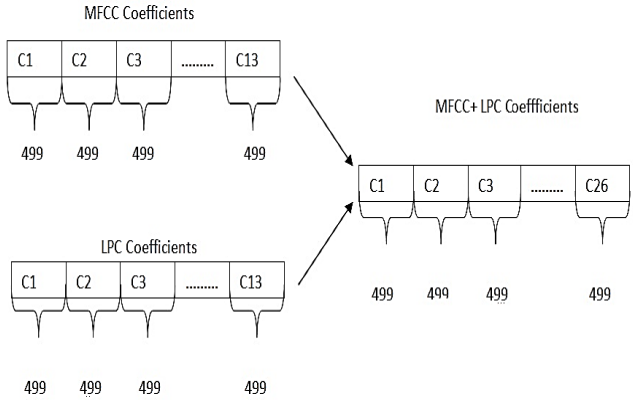

Figure.5 Vector generation for 5 Second Segment

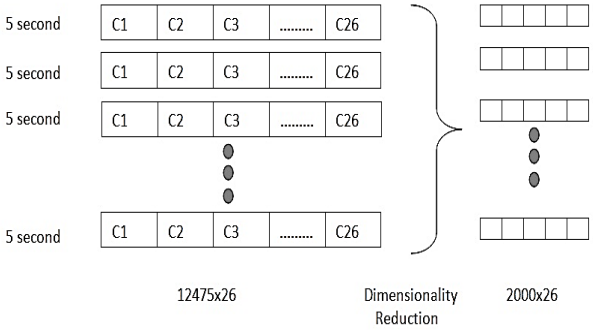

Figure.6 Dimensionality Reduction for a Song Fragment

3.8.1. K-Nearest neighbour(kNN) classification [25]

The k-nearest neighbor method is originated in 1950s. It belongs to lazy learners' category as it does less work when training tuple is presented and does more work at the time of classification. It is a supervised learning approach in which a new query is classified based on the majority of its k nearest neighbors. Commonly used distance measures are Euclidean distance, Malhobis distance and Manhattan distance. 10 -fold cross validation is used for performance comparison. Value of K is considered as 1. Higher value of K does not improve the performance.

The Euclidean distance between two tuples, say, X1= <x11,x12,x13,.....,x1n> and X2=<x21,x22,x23,.....x2n> is defined as follows.

$$\text{dist(x1, x2)}=\sqrt{\sum_{i=1}^{N}(x1i - x2i)^2} \quad (3)$$

3.8.2. Naive baysian method [26]

The Naive Bayes algorithm calculates discriminant function for each possible n classes. It assumes that each feature of an audio clip is drawn independently from a normal distribution and classifies according to the Bayes optimal decision rule.

3.8.3. Neural network method

Neural Network is very popular in Machine Learning. Automatic Speech recognition based on MFCC, LPC and preceptual linear prediction (PLP) altogether as features using multilayer feed forward Neural Network with back-propagation is discussed in [27].

Scaled conjugate gradient (SCG) algorithm [28] is a fast supervised learning algorithm based on conjugate directions. The performance is far better against that of standard back propagation algorithm, conjugate gradient algorithm with linear search. SCG does not include user dependent parameters and neglects linear search in each iteration in order to determine an appropriate step size. Back propagation is used to calculate derivatives of performance with respect to the weight and bias variables.

Training terminates when one of the following conditions are satisfied[29]

1) The maximum number of epochs is achieved
2) The maximum amount of time is exceeded
3) Performance is minimized to the goal
4) The performance gradient falls below min_grad
5) Validation performance has increased more than max_fail times since the last time it is decreased

3.9. Folk Dance Classification

Song Classification is very important because each song segment which is of 5 second may contain pure vocal (male or female) or may contain extensive use of musical instrument which is not dominating in the remaining segments. Here, Majority voting is considered in the final decision for folk dance song classification.

Table 2. Neural Network Parameters used in MATLAB

Parameter	Value
Epochs	1000
Goal	0
Time	Inf
Min_grad	1e-6
Max_fail	10
Sigma	5.0e-5
Lamda	5.0e-7
Hidden layer Neurons	50
Input Neurons	26 (13MFCC+13 LPC)
Output Class	4
Performance	Mean Squared Error

Algorithm for Folk Dance Classification

```
For each song used in Testing
    For each song segment i
      Extract class_label and store into
                    calculated_segement_label[i]
       if ( calculated_segement_label[i] == '1')
                    then  C1_class=C1_class+1
      elseif ( calculated_segement_label[i] == '2')
                    then  C2_class=C2_class+1
      elseif ( calculated_segement_label[i] == '3')
                    then  C3_class=C3_class+1
       elseif ( calculated_segement_label[i] == '4')
                    then  C4_class=C4_class+1
       end
    end
Out_class=max(C1class, C2class, C3class, C4class)
end
```

4. Experimental Setup

A novel folk dance database is constructed for four Indian Folk dances widely popular in Indian movies. Folk dance songs are extracted from movies and albums which are publicly available in Cds/DVDs/MP3s. A collection of 40 songs is prepared for each folk dance. Proposed approach is evaluated using 160 popular songs. To maintain uniformity, all the audio songs extracted from video are resampled at a rate of 24Kbps. We have used MATLAB R2013a for experiment on a Laptop with i7 -5500U processor with 2.40 GHz & 8GB RAM.

5. Experimental Results

Folk Dance Recognition System is divided into three main parts: Training Phase , Validation and Testing Phase. Folk Dance Model(FDM) is generated and 80% of data is used for training and 20% of remaining data is used for testing. The reliable measure of classifier accuracy can be obtained by Holdout, random subsampling, k-fold cross validation or the bootstrap techniques. To calculate the accuracy of our system, we have used 5-fold random cross validation method. Accuracy of the proposed system is computed using the following eq. (4). Fig.7 and Fig. 8 show confusion

Table 3. Folk Song Database

Id	Folk Dance Name	State	No. of Songs
D1	Garba	Gujarat	40
D2	Lavani	Maharastra	40
D3	Bhangra	Punjab	40
D3	Ghoomar	Rajasthan	40

matrices for song segments and for whole song respectively.

$$Accuracy\ (\%)\ =\ (DIC\ /\ TD\ )\ x\ 100 \quad (4)$$

Here, DIC depicts correctly identified Dance and TD indicates Total Dance used for Testing.

ROC stands for Receiver Operating Characteristic. ROC curves come from signal detection theory shows the trade-off between the true positive rate and the false positive rate for a given model. The area under the ROC curve is a measure of the accuracy of the model. The closer the ROC curve of a model is to the diagonal line, the less accurate the model [30]. A ROC curve for song segment (Not for the whole song) is plotted in Fig. 9. It is clear from Testing ROC curve that specified model is highly accurate as the curve moves steeply up from zero.

Accuracy obtained via different classification algorithms is mentioned in Table 4.

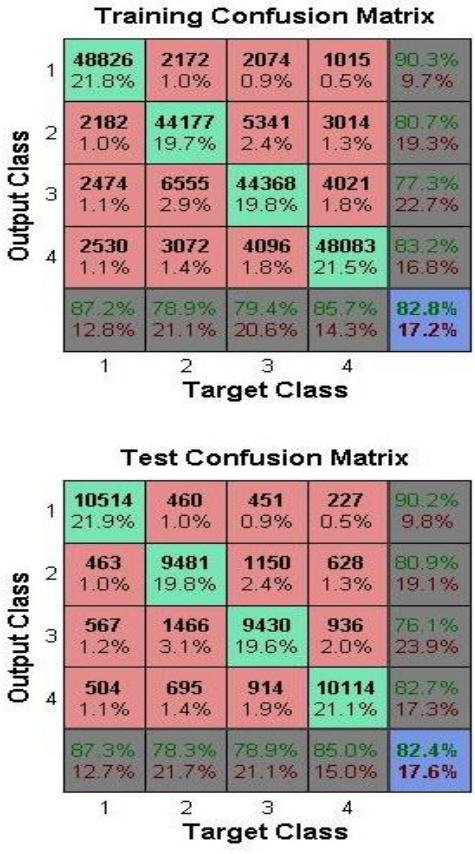

Figure.7 Folk Dance Song Segment Confusion Matrices for Training, and Testing

Table 4. Classification Accuracy Comparison

K-Nearest Neighbour	Neural Network	Naive Bayesian
85%	90.8%	60%

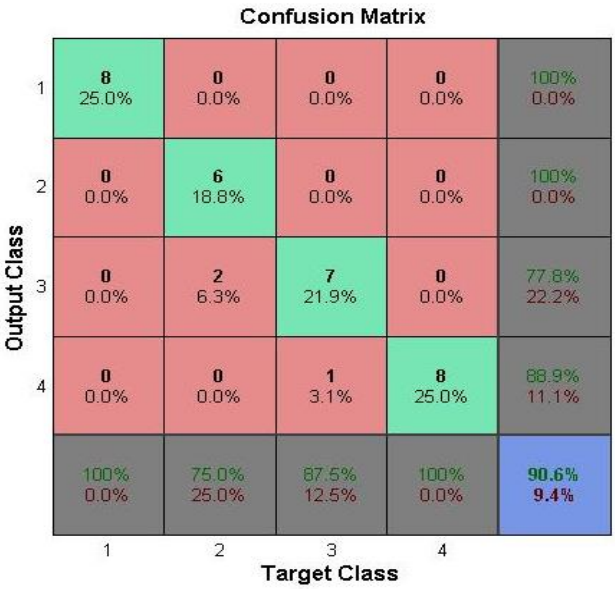

Figure.8 Whole Folk Dance Song Test Data

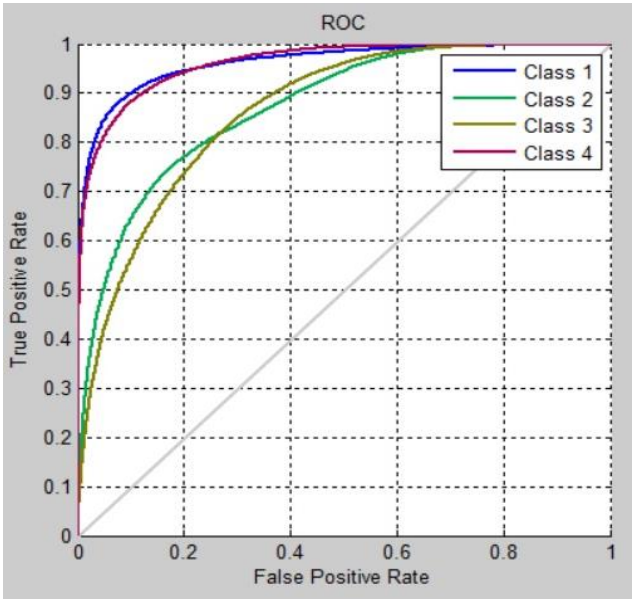

Figure.9 ROC Curve for Song Segment

6. Conclusion

Indian folk dance plays a vital role in human life as well as widely used in Bollywood Movies. '*Garba*', '*Lavni*', '*Ghoomar*' and '*Bhangra*' are four different folk dances considered for classification. For each class 40 different songs are considered which gives a total of 160 songs. MFCC and LPC based features are extracted from audio and classification accuracy is measured using and K-Nearest Neighbour, Neural Network and Naive Bayesian approaches. More than 90% classification accuracy is achieved from Neural Network. Similar to folk dance songs present in Bollywood movies, the other song genres like Bhajan, Kawaali; events

like War, Violence, Wedding; the voices of animals; chirping of birds can be classified. This can be considered as a step towards 'Content Based Bollywood Movie Mining'.

References

[1] R. Sharma, Y. S. Murthy, S.G. Koolagudi, "Audio Songs Classification Based on Music Patterns", In: *Proc. of Second International Conf. On Computer and Communication Technologies*, pp. 157-166, 2015.

[2] R. Massey. *India's dances: Their History, Technique, and Repertoire*, Abhinav Publications, p. 26, 2004.

[3] https://en.wikipedia.org/wiki/Indian_folk_music

[4] D. Hoiberg, *Students' Britannica India*, Vol. 2, Popular Prakashans, p.392, 2000

[5] T. C. Nagavi, and N.U. Bhajantri, "Overview of automatic Indian music information recognition, classification and retrieval systems", In: *Proc. of the International Conf. on Recent Trends in Information Systems*, pp. 111-116, 2011.

[6] A. Flexer, "A closer look on artist filters for musical genre classification", *World*, Vol.19, No.122, p.16-7, 2007.

[7] S. Shreshthova, *Between cinema and performance: Globalizing Bollywood Dance*, ProQuest. p.372, 2008.

[8] I. Kapsouras, S. Karanikolos, N. Nikolaidis, & A. Tefas, "Folk dance recognition using a bag of words approach and ISA/STIP features", *In: Proc. of the 6th Balkan Conf. in Informatics*, pp. 71-74, 2013.

[9] S. Samanta, P. Purkait, & B.Chanda, "Indian Classical Dance classification by learning dance pose bases". In: Proc. of the *IEEE Workshop on Applications of Computer Vision*, pp. 265-270, 2012.

[10] M.N. Kaynak, Q.Zhi, A.D. Cheok, K. Sengupta, Z. Jian, and K.C.Chung, "Analysis of lip geometric features for audio-visual speech recognition", *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, Vol..34, No.4, pp.564-570, 2004.

[11] L. Chen, S.J. Rizvi, and M.T. Ozsu, "Incorporating audio cues into dialog and action scene extraction", In: *Proc. of Electronic Imaging*, pp. 252-263, 2003

[12] Y. Song, W.H. Wang, and F.J. Guo, "Feature extraction and classification for audio information in news video." In: *Proc. of IEEE International Conf. on. Wavelet Analysis and Pattern Recognition*, pp. 43-46, 2009.

[13] C.N. Silla Jr, C. A Kaestner, and A. L. Koerich. "Automatic music genre classification using ensemble of classifiers", In: *Proc. of IEEE International Conf. on Systems, Man & Cybernetics*, pp. 1687-1692, 2007.

[14] T. Zhang, and C.C. Kuo, "Hierarchical classification of audio data for archiving and retrieving", In: *Proc. of IEEE International Conf. on. Acoustics, Speech, and Signal Processing*, pp. 3001-3004, 1999.

[15] G. Tzanetakis, and P. Cook, "Musical genre classification of audio signals", *IEEE transactions on Speech and Audio Processing* Vol.10, No.5, pp. 293-302, 2002.

[16] C. Nopthaisong, and M.M. Hasan, "Automatic music classification and retreival: Experiments with Thai music collection", In: *Proc. of IEEE International Conf on Information and Communication Technology*, pp. 76-81, 2007.

[17] S. Chu, S. Narayanan, and C.C. J. Kuo, "Environmental sound recognition with time–frequency audio features." *IEEE Transactions on Audio, Speech, and Language Processing*, Vol.17, No.6, pp. 1142-1158, 2009.

[18] K. Palecek and J. Chaloupka, "Audio-visual speech recognition in noisy audio environments", In: *Proc. of 36th IEEE Telecommunications and Signal Processing*, pp. 484-487, 2013.

[19] L. Chen, S. Gunduz, and M.T. Ozsu, "Mixed type audio classification with support vector machine", In: *Proc. of IEEE International Conf. on Multimedia and Expo*, pp. 781-784, 2006.

[20] T. Li, and G. Tzanetakis, "Factors in automatic musical genre classification of audio signals", In: *Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 143-146, 2003.

[21] W. Zhang, Q. Zhao, Y. Liu, and M. Pang, "Co-training Approach for Label-Minimized Audio Classification", In: *Proc. of IEEE International Conf. on Measuring Technology and Mechatronics Automation*, pp. 860-863, 2010.

[22] M.N. Raja, P.R. Jangid, and S.M. Gulhane, "Linear Predictive Coding", *International Journal of Engineering Sciences & Technology*, pp.373-379, 2015.

[23] M. J. Zaki, W. Meira Jr, and W. Meira, *Data mining and analysis: fundamental concepts and algorithms*, Cambridge University press, pp.187-201, 2014

[24] V. Panagiotou and N. Mitianoudis, "PCA summarization for audio song identification using Gaussian mixture models", In: *Proc. of*

18th international conf. on digital signal processing, pp.1-6, 2013.
[25] H. Maniya, M. Hasan, and K.P. Patel, "Comparative study of naive bayes classifier and KNN for Tuberculosis", In: *Proc. of International Conf. on Web Services Computing*, pp.22-6, 2011.
[26] D. W. Aha, "A study of instance based algorithms for supervised learning tasks: Mathematical, empirical, and psychological evaluations", Doctoral Dissertation, University of California, 1990.
[27] N. Dave, "Feature extraction methods LPC, PLP and MFCC in speech recognition", *International Journal for Advance Research in Engineering and Technology*, Vol.1, No.6, pp. 1-4, 2013.
[28] M.F.Moller, "A scaled conjugate gradient algorithm for fast supervised learning", *Neural Networks* ,Vol.6, No. 4, pp.525-533,1993.
[29] http://in.mathworks.com/help/nnet/ref/trainscg.html
[30] G. M. Williams, D.A. Ramirez, M.M. Hayat, and A.S. Huntington, "Instantaneous receiver operating characteristic performance of multi-gain-stage APD photoreceivers", *IEEE Journal of the Electronic Devices Society* ,Vol.1, No.6, pp. 145–153, 2013.