Indian Folk Music Classification

Submitted By

Gor Krutarth Dhimantkumar

14MCEN07

तमसो मा ज्योतिर्गमय

NIRMA
UNIVERSITY

INSTITUTE OF TECHNOLOGY

NAAC ACCREDITED ‘A’ GRADE

DEPARTMENT OF INFORMATION TECHNOLOGY
INSTITUTE OF TECHNOLOGY
NIRMA UNIVERSITY
AHMEDABAD-382481
May 2017

Indian Folk music Classification

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

(Networking Technologies)

Submitted By

Gor Krutarth Dhimantkumar

(14MCEN07)

Guided By

Prof. Sapan H Mankad

DEPARTMENT OF INFORMATION TECHNOLOGY

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Certificate

This is to certify that the Major Project entitled **"Indian Folk Music Classification"** submitted by **Gor Krutarth Dhimantkumar (Roll No: 14MCEN07)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering (Networking Technologies) of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Major Project part-II, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Sapan H Mankad
Guide & Assistant Professor,
IT Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Gaurang Raval
Associate Professor,
Coordinator M.Tech - CSE (NT)
Institute of Technology,
Nirma University, Ahmedabad

Dr. Madhuri Bhavsar
Professor and Head,
IT Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Statement of Originality

I, **Gor Krutarth Dhimantkumar**, Roll. No. **14MCEN07**, give undertaking that the Major Project entitled "**Indian Folk Music Classification**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science and Engineering (Networking Technologies)** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Date:
Place:

Endorsed by

Prof. Sapan H Mankad

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Sapan H. Mankad**, Assistant Professor, Information Technology Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Madhuri Bhavsar**, Hon'ble Head of Information Technology Department, Institute of Technology, Nirma University, Ahmedabad for her kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Mrs. Geetanjali Barua**, Research Officer at Sanskar Kendra, Paldi, Ahmedabad for the information provided by her for my project work. Without that information I was unable to go ahead in my project work.

I would like to thank my dear friends, for their support, help, good suggestions where I stuck, and specially for the great time we have spent together.

I would also like to thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Krutarth Gor
14MCEN07

Abstract

The goal is to classify the folk songs of India based on regions. According to the region, basically India is divided into 4 parts for the music: North, East, West and South. But for each part there are also several sub regions which differs from each-others by culture, music, speech and more. Concentrating on folk music, to take folk songs of each and every region is better than to take generalized collection of folk songs of mainly divided regions. It can give better classification of the folk songs. For this task, no any dataset is available directly. Because of that we collected songs from different websites. We could collect the folk songs of five regions: Assamese, Marathi, Kashmiri, Kannada and Uttarakhandi. After collecting folk songs, we extract various features such as Mel Frequency Cepstral Coefficients (MFCC), RMS value (loudness feature) and Spectral Centroid from each folk song for the classification of folk songs. Then we apply different classification techniques Artificial Neural Network (ANN), K Nearest Neighbor (KNN), Random Forest and Support Vector Machine (SVM) of machine learning to classify the feature sets into different classes, where each class represents a region. Subsequently, we experimented for feature subset selection methods: SelectFromModel and Recursive Feature Elimination (RFE) to get which method is appropriate for out project to get best features from the dataset to perform better classification. We collect all the results and decide which model is better for the folk song classification and how different features and combination of all features affect the classification of folk songs. Then we decide ANN performs better for our purpose than KNN and random Forest classifiers. After deciding the classifiers, we proposed a method to compare the prediction results of two finalized datasets predicted with the use of ANN by using nested K fold Cross Validation for 5 folds with the prediction of feature sets of another classifiers SVM to check and compare which dataset performs better with respect to finally received prediction result.

Abbreviations

ANN	Artificial Neural Network
DCT	Discrete Cosine Transform
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
FPR	False Positive Rate
GMM	Gaussian Mixture Model
LPC	Linear Predictive Coefficient
MFCC	Mel-Frequency Cepstral Coefficient
MIR	Music Information Retrieval
MIREX	Music Information Retrieval Evaluation Exchange
PCD	Pitch Class Distribution
PCDD	Pitch Class Dyad Distribution
RFE	Recursive Feature Elimination
STFT	Short Time Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Folk songs are like heartbeats of the region. They are created in traditional culture of a region. Folk songs are easy to understand with the well-defined lyrics and are created with the available instrument resources for music. It is created on the basis of the human mood like sorrow, happiness also on the basis of change in a season etc. It shows that folk songs are rich in culture elements and dominant art of human history. Automatic genre based classification is widely studied in the area of music information retrieval field, but the regional based classification is hardly used. As the field of Music Information Retrieval became popular, a research on folk songs was involved in this field for the analysis of folk song melody and region based classification of folk songs. There are limited number of researches available which focus on classification of folk songs.

1.1 Folk music of India

There are 29 states and 7 Union Territories in India. They all have their own culture, music, life style and language. India is diversify country in music. There are number of regions in India popular for their regional music that is called folk music of that region. Folk songs are like heartbeats of the region. Folk songs are easy to understand with the well-defined lyrics and are created with the available instrument resources for music. It does not require any investment. The main thing is that each state has a special instrument which is used in that region only. It is created on the basis of the human mood like sorrow, happiness also on the basis of change in a season etc. Most of the people are used to sing folk songs on a daily basis for some incident or in a festival and enjoy that music. Some of the famous folk songs of India are: Dandiya and Garba from

Gujarat, Sambalpuri from Odisha, Lavni from Mahashtra, Bihugeet from Assam etc.
But it is not an easy task for the people who are not familiar with the folk songs of any region. For example, a Gujarati person may not able to identify Kashmiri song without help of the expert of the Kashmiri music. Even sometimes it happens that a person cannot identify a folk song of its own region just because of the lack of knowledge. Though there are different different varieties in Indian folk music, still we are unaware about this treasury. It is necessary that we can come to know at least our own state's folk music and folk songs. It can help to many music related areas and also very useful for the music lovers.

Basically Indian folk songs are categorized into 4 broad categories: Eastern India, North India, Western India and Southern India. But it will not help to study precisely about the classification of folk music, because in India, each and every state has different culture. So rather focus on broad categories, we will focus on each and every region's folk songs. It will make easy to classify and also to differentiate the folk music among neighborhood regions. As per the varieties in the folk song, each region have mostly same kind of varieties of folk songs. It shows that there is fixed kind of music pattern in accordance of rhythm.

1.2 Folk Music Classification

Any song is basically a wave formatted data which contains combinations of different frequencies of male/female voice, instruments etc. On the basis of these information, we can classify songs into different categories. W have applied machine learning techniques to classify the Indian Folk music. Machine learning[1] is used to train the example data or past experience to program computers so that computers can solve the given problem on the basis of the result collected from the example data. machine learning can be used for many application areas like patter recognition, Image Processing, Signal Processing, Statistical Analysis, etc. Machine learning tasks are mainly divided into three categories:

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

[1] https://mitpress.mit.edu/books/introduction-machine-learning

There are some another categories of machine learning based on the tasks when someone considers the desired output of the machine learning system. these categories are[2]:

- Classification: Can help to visualize and demonstrate data more accurately
- Regression: Build model to predict continuous data
- Clustering: Used on unlabeled data to find natural grouping and patterns

I have used machine learning with python. As python provides many libraries like numpy, scipy, scikitlearn for complex mathematical calculation support like signal processing, prediction work and many more. Even there are many frameworks and scientific tools which provides collection of libraries for the particular field. OpenCV, Anaconda, Keras are examples of this kind of libraries which makes development quite easy.

1.3 Motivation

Audio signal contains different kind of frequencies containing information using which we can identify some required result. There are regional folk songs in India. By processing these songs, we can make it useful for multiple purpose like:

- Distinguish different kind of music played and songs are sung in a same region
- It becomes useful for unrelated people and they can identify the category of Indian folk music
- Students learning music will be able to identify type of a song and can utilize it further
- Identify musical instruments used in each region
- For Music Industry

1.4 Problem Statement

To know 'The Given unknown song as input is related to which region's folk song ??' For this a classification technique has been used. After collecting the song from various websites[2], features were extracted from all the songs to train them by which we can

[2] https://www.mathworks.com/solutions/machine-learning.html

classify the unknown songs according to the region it matches. But only feature selection is not enough. There can be many features which are less useful or completely useless. these features reduce the accuracy. To overcome this we have applied feature selection techniques as a pre-processing step. Through which only important features were selected from the whole dataset. It also reduces the process time of the model. This full process is shown in the diagram 1.1.

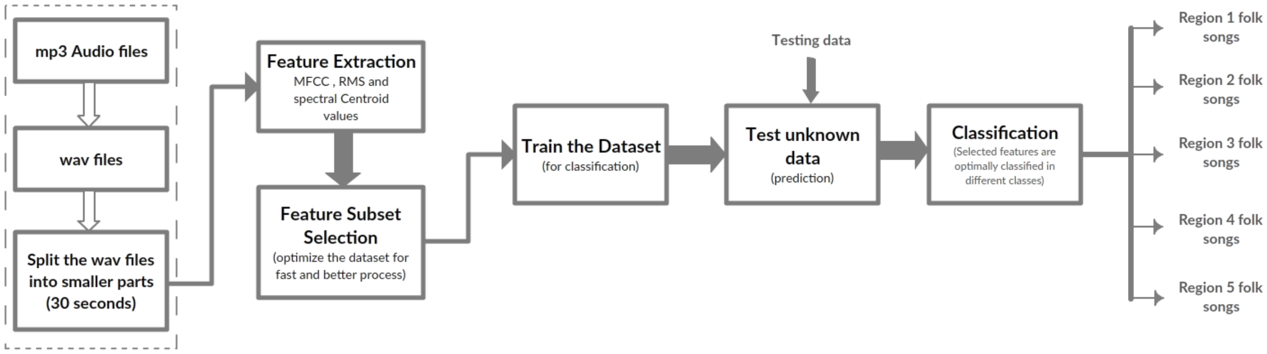


Figure 1.1: Steps for Classification

Chapter 2

Literature Survey

2.1 Techniques

Music Information Retrieval is a vast area conducting many kind of research work and development of applications for audio processing, singer identification and lots more. Most of the research work done in MIR field are related to western music. As there are different countries with their own culture and music, it is necessary to work on those music also. So that we can identify some information like, similarities between different country's music, instruments used, Genre etc.

There are different applications available for music identification. Some of them are: SHAZAM, Cuidado, audentify!, Themefinder and lots more.

	Input		Matching					Features									
Name	Audio	Symbolic	Audio	Symbolic	Exact	Approximate	Polyphonic	Audio Fingerprints	Pitch	Note Duration	Timbre	Rhythm	Contour	Intervals	Other	Indexing	Collection Size (Records)
audentify!	●		●			●	●	●								Inverted files	15000
C-Brahms		●		●	●	●	●		●	●		●		●		none	278
CubyHum	●			●		●								●		LET	510
Cuidado	●		●			●	●				●	●			●	not described	works for > 100, 000
GUIDO/ MIR		●		●		●			●	●		●		●	●	Tree of transition matrices	150
Meldex/ Greenstone	●	●		●		●							●	●		none	9354
Musipedia	●	●		●		●							●			Vantage objects	> 30, 000
notify! Whistle	●	●		●		●			●			●				Inverted files	2000
Orpheus		●		●		●	●		●	●		●		●		Vantage objects	476000
Probabilistic "Name That Song"		●		●		●								●	●	Clustering	100
PROMS		●		●	●	●			●			●					12000
Cornell's "QBH"	●			●		●							●			none	183
Shazam	●		●		●		●	●								Fingerprints are indexed	> 2.5 million
SOMeJB	●		●			●	●								●	Tree	359
Sound-Compass	●			●		●			●			●				Yes	11132
Super MBox	●			●		●			●			●				Hierarchical Filtering	12000
Themefinder		●		●	●				●				●	●		none	35000

Table 2.1: Applications in the field of MIR [1]

By seeing all these work and applications, very less work is done for the folk songs till now. But the countries like India has vast area of folk music and folk songs. I found research papers of some work done in the era of folk songs for the India, China, Greece and, Korea[3] Japan.

2.2 Work done in the area of Music Information Retrieval

The research work done in India in the arrea of Music Information Retrieval is related to improve melodic similarity in Indian art music i.e. **North Indian Music** and **Carnatic Music.** They have represented melody of an audio signal by the pitch of the predominant melodic source. For predominant pitch estimation in Carnatic music, they have used the method proposed by Salamon and Gomez.[4] This method performed favourably in MIREX 2011 on a variety of music genres, including IAM, and has been used in several other studies for a similar task. Some work is done on the Carnatic music[5][6]. Another work done in India in the field of MIR is for raga. Table 2.2 and 2.3 show the details regarding the work of the music recognition system for raga.[4]

Paper	**Synopsis**	**Results**
2004 Parag Chordia et al. [7]	Using datset of 2 ragas of different length with the use of tone and spectral profiles features	Accuracy of 100% using HMM
2007 Parag Chordia et al. [8]	datset of small raga performances is used contained 72 minutes of data of 17 ragas using pitch detection, onset detection, PCD features. MVN, KNN , FFNN methods are used.	94 % in 10-fold cross validation and accuracy of 82% usinf PCDs
2009 Parag Chordia et al. [9]	A dataset of 43 performances on 30 ragas performed by 22 performers with the use of features pitch-detection, onset-detection, PCD amd PCDDs	Accuracy of 92.4% using bayesian classifier Method

Table 2.2: Research work on Classification of ragas
[10]

Paper	Synopsis	Results
2008 Parag Chordia et al.[11]	Dataset of 897 classical music audio tracks of North Indian music and of 14 different instruments with the features like timbre and PCDs.	With ccuracy of 32.69% for artist recognition and 90.30% for instruments using MFCC and GMM
2009 Surendra Shetty et al.[12]	Note transcription, Arohana and Avarohana	Used samples of 50 different ragas and tested for 20 different ragas with 3-5 songs of each raga totally 90 songs

Table 2.3: Work done on Retrieval system in India
[10]

Apart from these research work, some work for recognition systems[10] has also been done in India. For this systems, datasets are taken Arohi and Avarohi sequences[8], audio of different ragas, samples of audios sung by different singers and more. All these systems are processed with the features like swaras, pakad, annotations, onnset detection, pitch class Distributions (PCDs), Vocal pitch etc using HMM, SVM, MFCC, ANN, KNN and some more methods.

2.2.1 Work done in the area of Classification and Identification of Folk songs

In the area of MIR, most of the work done is for genre based audio classification and this work is mostly done for the western music. In accordance of these work, very less work is done for the other types of songs though many of the types of song from them are very popular. For the Folk songs, very less research is done though folk songs are very useful material for generating new music patterns. Countries like China, Greece, Austria, research for classification of folk songs has been done successfully. Work done in the area of folk song in some countries is shown in the table 2.4.
From the Table 2.4, it can be seen that two kind of datasets were used for folk song classification: Audio and Symbolic. Different kind of features were extracted for the classification from the audio file for the Audio based approach. For the classification of

Author	type of the audio clip format	size of Dataset	Synopsis	Results
Yi Liu, JiePing Xu, Lei Wei, Yun Tian[13]	WAV or MP3	495 (328 for training and 167 for testing)	**feature extraction:** 12 basic acoustic features, **classification:** SVM by combining it with 4 feature selection techniques	47.40% (before post processing), 75.2% (after post processing)
Suisin Khoo, Zhihong Man, Zhenwei Cao[14]	Symbolic notations (16 bit monophonic audio wav file)	312 (281 for training, 31 for testing)	**Feature extraction:** used MFDMap to encode symbolic features of music **classification:** Using one of the ANN technique, R-ELM	72.1% (for symbolic approach), 49% (for audio approach)
Suisin Khoo, Zhihong Man, Zhenwei Cao, Jinchuan Zheng[15]	Symbolic notations (16 bit monophonic audio wav file)	106 German and 104 Austrian Folk songs (168 for training, 42 for testing)	**Feature extraction:** MFDMap to encode symbolic features of music, **classification:** Using one of the ANN technique, FIR-ELM	31.43% (for interval), 78.57% (for combination with different features), 83.33% (for combination of interval, duration and duration ratio)
Nikoletta Bassiou, Constantine Kotropoulos, and Anastasios Papazoglou-Chalikias[16]	Monophonic wav audio file	98 songs from Pontus and 94 from Asia-Minor	28-MFCC used for each frame of the size 30ms for feature extraction and for the classification has been done by resorting CCA and Deep CCA	91% with the auditory spectrotemporal modulation
Chai, Wei, and Barry Vercoe[17]	Monophonic audio files encoded in kern or EsAC	187 - Irish, 200-German, 104-Austrian folk songs	focused on the melody of each folk song and used different HMMs for classification	75% (2-way HMM), 77% (3-way HMM), 66% (6-state left right HMM) using the interval representation

Table 2.4: Research work on Classification of folk music

the folk songs, two ways were used till now: Genre based and Region Based. From the paper[10], though genre based approach is mostly used for the audio classification, they have shown that region based approach is better for folk song classification rather to use genre based approach. From these survey we have decided to use Region based Approach for the classification of Indian Folk songs with different feature extraction techniques and classifiers used in machine learning.

Chapter 3

Implementation

From all the regions' folk music, we have taken folk songs of five regions for this project. For this work, we have taken our dataset of .wav file to process the audio file. And after that different feature extraction techniques were applied to extract features from the audio file for the classification of the folk songs. Following the feature extraction step, the classification methods were applied on the feature set to decide the best classification approach for this project. All these steps are explained in details in following sections in this chapter.

3.1 Tools and Technology

- Programming Laguage : Python (Version 2.7)
- Library/platform : Anaconda , Open data science platform
- IDE : Spyder
- Music Signal Pre-processing : Praat and Audacity

3.2 Dataset

Here we have collected five regions' folk songs for the process. Total Number of folk songs collected by 299. All the songs are taken from the different websites cited in [2, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]. To get the results of the classification, dataset is divided into two parts: Training and Testing with the ratio of 80:20. Folk songs per region is shown in the Table 3.1.

Folk song	No of Songs	Training set	testing set
Assamese	134	108	26
Uttarakhandi (Pahadi)	29	23	6
Kashmiri	39	32	7
Kannada	63	51	12
Marathi	34	28	6

Table 3.1: Description of Dataset

Normally the minimum length of the song is 2 to 3 minutes which is very large to generate accurate feature sets. Better features can be generated from the linear audio file[28]. So each file is divided into 30 seconds slots before applying feature extraction techniques to the audio files.

3.3 Feature Extraction

Here we have used three different feature extraction techniques for better classification of folk songs. These techniques are:

- Mel Frequency Cepstral Coefficients (MFCC)
- Root Mean Square (RMS)
- Spectral Centroid

3.3.1 Mel Frequency Cepstral Coefficient (MFCC)

Steps for feature extraction:

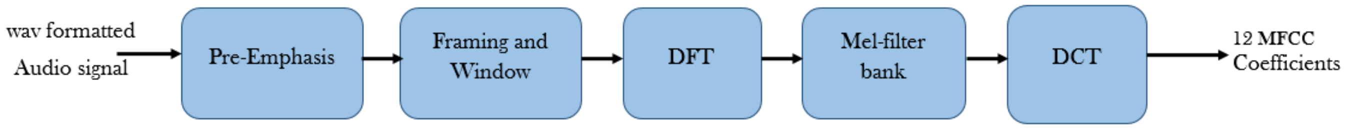


Figure 3.1: MFCC feature Extraction

Pre-Emphasis

Pre-Emphasis stage is convenient in many ways:[1]

- Balance the frequency spectrum as higher frequency has lower magnitude and lower frequency has higher magnitude.

[1] http://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

- Ignore numerical problems during Fourier Transform
- Possibly improve **Signal-to-noise (SNR)** ratio.

The pre-emphasis filter can be applied to the signal x using the first order filter as follows:

$$y(t) = x(t) - \alpha x(t-1) \tag{3.1}$$

where the default values of filter coefficient α are 0.95 or 0.97

Framing and Window

After the pre-emphasis stage, I have split the audio signal into short-time frames. Here I have split each .wav file into 30 seconds of portion. The reason for this step is, frequencies of a signal change over time also in a single whole file. To neglect this, we can consider that frequencies in a signal are static over a small time period. So that, by applying Fourier Transform over this short time period frames, a better approximation of the frequency contours of the audio signal can be got.
Here, i have taken $frame\_size = 0.025$ and 10s stride i.e. 15ms overlap.

After the splitting a song into frames, we have applied Hamming Window to each frame. Hamming window can be formulated as:

$$w[n] = 0.54 - 0.46\cos(\frac{2\pi n}{N-1}) \tag{3.2}$$

where, $0 \leq n \leq N-1$, N is the window length.

Now we can calculate N-point FFT on every frame to calculate frequency spectrum which is known as Short Time Fourier Transform (STFT). Normally the value of N is 256 or 512. I have taken $N = 512$. On the basis of this, power spectrum is calcueted as:

$$P = \frac{|FFT(x_i)|^2}{N} \tag{3.3}$$

Where x_i is the i^{th} frame of signal

Filter Banks

Filter Banks are calculated by applying triangular filters. We have taken 40 filters here i.e. $nfilt = 40$. These filter banks are applied on a mel scale to power spectrum for extracting frequency bands.

Mel scale can be calculated using following formula:

$$m = 2595log(1 + \frac{f}{700}) \tag{3.4}$$

and this mel value can be remodeled back to the hertz(f) using,

$$f = 700(10^{m/2595} - 1) \tag{3.5}$$

- The first filter bank starts at the first point, reach its peak at the second point then return to the zero at the third point
- Second filter bank will start from second point, reach its maximum at third point and at its fourth point, it will be zero

This can be formulated as below:

$$H_m(k) = \begin{cases} 0, & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)}, & f(m-1) \leq k < f(m) \\ 1, & k = f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)}, & f(m) < k \leq f(m+1) \\ 0, & k > f(m-1) \end{cases}$$

Mel-Frequency Cepstral Coefficients (MFCCs)

In the previous section, we saw a computation of Filter banks coefficients which are highly correlated with each other. In some of the Machine learning algorithms, it becomes problematic. Therefore, we can use **Discrete Cosine Transform (DCT)** here to decollate the filter bank coefficients. From that we can get compressed representation of filter banks.

From the all generated cepstral coefficients, only $2 - 13$ are kept and remaining are discarded. the reason is these discarded coefficients represent fast changes in filter bank coefficients and it makes complication for computation for audio signal.

As $2-13$ MFCC values are considered here, the number of cepstral are $num\_ceps = 12$.

DCT function is applied here to calculate MFCC values using following function.

$$mfcc = dct(filter\_banks, type = 2, axis = 1, norm =' ortho')[:, 1 : (num\_ceps + 1)] \tag{3.6}$$

3.3.2 Root Mean Square (RMS) Value

RMS is a single value feature which is used to measure of the power value of the signal. Sometimes it is used to measure loudness feature of the audio signal[**?**] for the classification problem. RMS value can be found using following equation:

$$RMS_r = \sqrt{\frac{1}{N}\sum_{i=0}^{N-1} |s_r(i)|^2} \tag{3.7}$$

where,

N: Spectral length, s: Audio Signal

s_r: r^{th} audio frame of a signal using N samples

3.3.3 Spectral Centroid

Spectral Centroid is a single valued feature which is a measured value of center of mass of the power spectrum[2]. Also we can say that it is the measure of the spectral brightness of audio signal[14]. It indicates where the center of mass in the signal is. One of the sue of the spectral centroid is as the predictor of the brightness of a audio[3]. Through this value we can classify the songs into different categories. Spectral centroid can be calculated using following equation:

$$SC_r = \frac{\sum_{n=0}^{N-1} M_r(n) \times n}{\sum_{n=0}^{N-1} M_r(n)} \tag{3.8}$$

where,

N: Spectral length n: frequency bin

M_r(n): Magnitude of the Fourier Transform at frame r

[2] http://jaudio.sourceforge.net/jaudio10/features/spectralcentroid.html

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

3.4 Feature Subset Selection

Suppose we have large dataset which contains 1000 columns of the features, that will take lots of time to process. It may be possible that there can be a subset from the dataset by using that we can get better accuracy than with the use of whole dataset. Feature selection is basically a process to reduce the dimensional of the dataset in the way of getting a better (or nearer) accuracy value by selecting a proper subset. There are different techniques applied for feature subset selection. From those techniques we have applied following two techniques: Recursive Feature Elimination (RFE) and Selectfrom-Model method. these two methods are provided in the feature_Selection class of sklearn library[4].

- **Recursive Feature Elimination (RFE)**

  RFE is used to select a subset of features recursively as smaller as possible. For this process, firstly, the estimator is trained using training set and weights are assigned to each feature. Then those features will be selected whose absolute weights are smallest [5]. This set of features is reduced from the current set of features. The process is recursively going on till it finds the smallest subset of features. A function to find set of recursive features is shown below:

$$RFE(estimator, n\_features\_to\_select = None, step = 1, verbose = 0) \tag{3.9}$$

Where,

estimator: a classifier(generally Support Vector Classifier is used)

n_features_to_select: number of features to apply for the selection, id None is given, half of the features are selected

step: It is a floating value. If values is between 0.0 and 1.0, the step will take the percentage of that value of features to remove at each iteration. If value is greater or equals to 1, the step will take that number of features to remove at each iteration.

[4] http://scikit-learn.org/stable/modules/feature_selection.html

[5] http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

- **SelectfromModel Method**

  SelectFromModel is a meta-transformer that can be used along with any estimator that has a 'coef_' or 'feature_importances_' attribute after fitting. In this method, unimportant features are removed that is decided on the basis of threshold value. There are different possibilities to set threshold. Generally $1.25 * mean$ value is used as a threshold value. Also a median of feature importance is also taken to set threshold. A function to find best subset using this method is shown below:

  $$SelectFromModel(estimator, threshold = None, prefit = False) \tag{3.10}$$

  Where,

  estimator: a classifier(generally Support Vector Classifier is used)
  threshold: it can be a string or float value. Default value is None. If median, the threshold value is median of feature importance.
  prefit: it is Boolean value. Default is None. Used to decide Whether a prefit. model is expected to be passed into the constructor directly or not.

3.5 Proposed Method

For the one more step, we have applied nested K-fold cross validation for the classification with the use of ANN as it performs better for our dataset.

1. Apply ANN for the classification

2. Use K fold Cross validation for main dataset for 5 fold

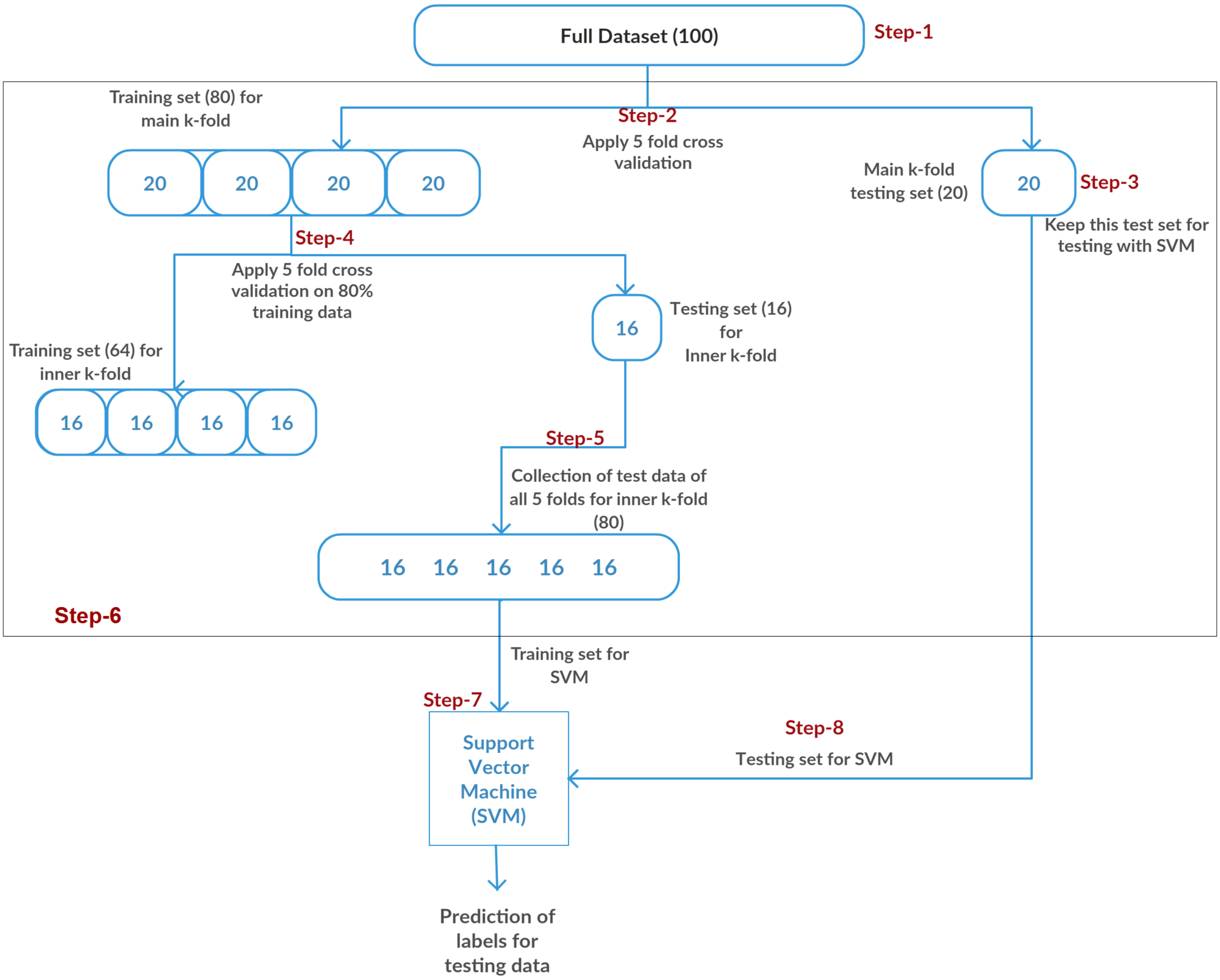
   a. 80% as Training

   b. 20% as Testing

3. Keep 20% i.e. testing portion of outer K fold cross validation without considering accuracy. Only consider prediction of labels for that and store those values.

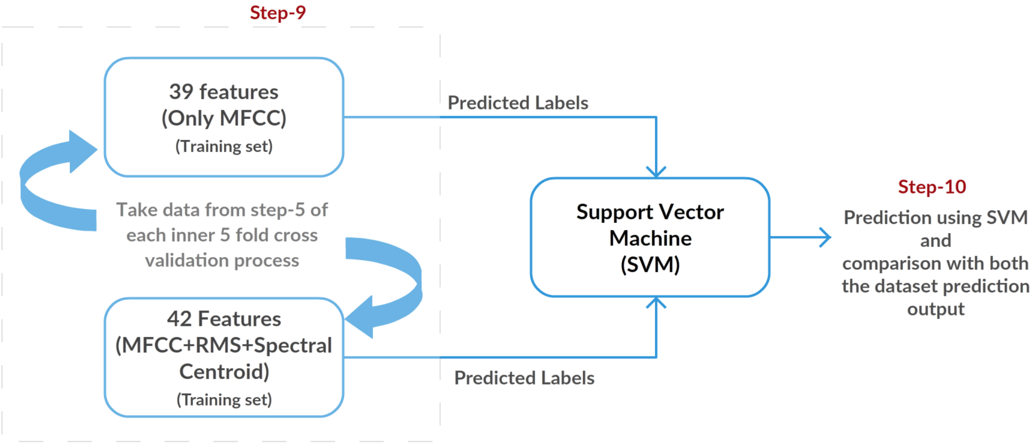
4. Use inner k fold cross validation for the 80% training dataset for main K fold cross validation

5. Keep testing data results (5th fold 20%) of inner k-fold cross validation for each outer k-fold cross validation



Full Dataset (100)
Step-1
Training set (80) for main k-fold
Step-2
Apply 5 fold cross validation
20
20
20
20
Main k-fold testing set (20)
20
Step-3
Keep this test set for testing with SVM
Step-4
Apply 5 fold cross validation on 80% training data
16
Testing set (16) for Inner k-fold
Training set (64) for inner k-fold
16
16
16
16
Step-5
Collection of test data of all 5 folds for inner k-fold (80)
16 16 16 16 16
Step-6
Training set for SVM
Step-7
Support Vector Machine (SVM)
Step-8
Testing set for SVM
Prediction of labels for testing data

(a) Steps 1 to 8



Step-9
39 features (Only MFCC) (Training set)
Take data from step-5 of each inner 5 fold cross validation process
42 Features (MFCC+RMS+Spectral Centroid) (Training set)
Predicted Labels
Predicted Labels
Support Vector Machine (SVM)
Step-10
Prediction using SVM and comparison with both the dataset prediction output

(b) Steps 9 and 10

Figure 3.2: Steps for Proposed Method

6. Repeat the process till the completion of the 5 fold cross validation process of outer k fold cross validation.

7. Apply collected testing data of inner k fold cross validation as input for Support Vector Machine(SVM) (80% of total data)

8. Use not considered 20% testing data of outer k fold cross validation as a test data for SVM

9. Apply this process for both the case of dataset: MFCC and MFCC+RMS+Centroid

10. Compare the class prediction of both the dataset with the final prediction results got from the SVM.

3.6 Classifier Methods

Here the data is available and labeled properly, we have used supervised algorithms of machine learning. We have tried Artificial Neural Network (ANN), K nearest Neighbor(KNN), and Random Forest methods to check which method gives better accuracy for this project. Mainly we have applied ANN for the classification, but reason to use other classifiers is ANN consumes more time than the other two KNN and random forests. So if we can get nearer accuracy results with anyone of these two classifiers to which we get accuracy by using ANN, it will be better to use that classifier. by testing this, we could analyze that random forest method is not giving better outcome for this work. On other hand KNN was working better. So we have finally experimented with both ANN and KNN separately as a classifier.

Chapter 4

Results of Experiments

4.1 Results

We have completed various experiments related to classification of Indian folk songs. For that we have taken different features, combination of all these features to see the effect of it on the classification accuracy. Also we have

4.1.1 Results using ANN for 12 MFCC features

In this part, we have shown the effect of sample rate on the classification. For that we have taken a dataset of 12 value MFCC features for each song for two different sample rates: 16 kHz and 44.1 kHz. This test is mentioned in the Chinese Folk song classification research paper[13]. Parameters we have set for ANN model is described below:

- Number of epochs: 55
- Number of hidden layers: 2 (except input and output layers)
- Length of one feature set:
    - MFCC: 12 values
- Size of dataset: (80-20)% ratio for training and testing set
    - Training set: 242 songs
    - Testing set: 57 songs

From the Table 4.2, it is shown that dataset of 16 kHz performs better than 44.1 kHz dataset. On the basis of that we took 16 kHz songs' dataset for further processing.

K-Folds	Dataset with different Sample rates	
	44.1 kHz	16kHz
3 Fold	0.7368	0.7602
5 Fold	0.7439	0.7509
10 Fold	0.7298	0.7526

Table 4.1: Comparison of 12-MFCC features for different sample rates using ANN

K-Folds	Dataset with different Sample rates	
	44.1 kHz	16kHz
3	0.7544	0.7016
5	0.7193	0.6667
7	0.7368	0.6842
9	0.7368	0.6667

Table 4.2: Comparison of 12-MFCC features for different sample rates using KNN

4.1.2 Results using ANN for 39 MFCC features

Here we have selected MFCC(13), delta MFCC(13) and delta-delta MFCC(13) features (total 39 features). for classification.

Below the list of parameters we have set for our ANN model:

– Number of epochs: 55

– Number of hidden layers: 2 (except input and output layers)

– Length of one feature set:

- MFCC: 39 values
- MFCC + RMS = 39 + 2
- MFCC + RMS + centroid: 39 + 2 + 1

– Size of dataset: (80-20)% ratio for training and testing set

- **Results of full dataset**

From Table 4.3 and 4.4, it is clearly seen that combination of different features is affecting to the classification. And with the 5 folds, we achieved better and consisting classification testing accuracy results. From the results of Table 4.4, it can be seen the effect of combination of features on classification. Considering

Dataset	Accuracy for K Number of Folds		
	3 Folds	5 Folds	10 Folds
MFCC features	0.7544	0.7368	0.7719
MFCC + RMS features	0.7193	0.7719	0.7544
MFCC + spectral centroid features	0.7193	0.7544	0.7544
MFCC + RMS + centroid features	0.7894	0.7719	0.7894

Table 4.3: Testing Accuracy results for full dataset

Dataset	Accuracy for K Number of Folds		
	3 Folds	5 Folds	10 Folds
MFCC features	0.7485	0.7158	0.7491
MFCC + RMS features	0.7135	0.7333	0.7421
MFCC + spectral centroid features	0.7602	0.7368	0.7474
MFCC + RMS + centroid features	0.7251	0.7544	0.7421

Table 4.4: Testing K fold score for full dataset

only MFCC, RMS and spectral centroid features, we achieved nearly 4% higher classification accuracy. It shows the importance of other two features loudness and Spectral Centroid for the classification of Indian Folk songs.

- **Results after feature subset selection with the use of SelectFromModel method**

  I have tried this method for two different threshold values: threshold = (1.25 * mean) which is standard value for the threshold taken default threshold i.e mean of all values.

Case 1:

Dataset	Accuracy for K Number of Folds		
	3 Folds	5 Folds	10 Folds
MFCC features	0.7017	0.6491	0.7192
MFCC + RMS features	0.6315	0.7543	0.6315
MFCC + spectral centroid features	0.6842	0.7018	0.7368
MFCC + RMS + centroid features	0.7017	0.6140	0.7543

Table 4.5: Accuracy results for threshold = $1.25 * mean$

Dataset	**Accuracy for K Number of Folds**		
	3 Folds	5 Folds	10 Folds
MFCC features	0.7018	0.6912	0.7053
MFCC + RMS features	0.6842	0.7018	0.7000
MFCC + spectral centroid features	0.6667	0.6982	0.7105
MFCC + RMS + centroid features	0.7251	0.6947	0.7105

Table 4.6: Cross Validation Score for threshold = $1.25 * mean$

Case 2:

Dataset	**Accuracy for K Number of Folds**		
	3 Folds	5 Folds	10 Folds
MFCC features	0.6491	0.6842	0.7017
MFCC + RMS features	0.7192	0.7368	0.7192
MFCC + spectral centroid features	0.7192	0.6492	0.7719
MFCC + RMS + centroid features	0.7076	0.7368	0.6491

Table 4.7: Accuracy results for default threshold

Dataset	**Accuracy for K Number of Folds**		
	3 Folds	5 Folds	10 Folds
MFCC features	0.7018	0.6947	0.7018
MFCC + RMS features	0.7251	0.7228	0.7228
MFCC + spectral centroid features	0.7135	0.6982	0.7175
MFCC + RMS + centroid features	0.7193	0.7123	0.7105

Table 4.8: Cross Validation Score results for default threshold

From both the cases for feature selection method SelectFromModel, by taking default threshold value for this dataset performs better and gives nearer classification accuracy. From Table 4.8, accuracy is lesser than original dataset just of 2.1% for MFCC dataset and 1.7% for combination of all features.

4.2 Results using K nearest Neighbor Classifier

We have examined for different values of K to select best value of K for the classification. Table 4.7 shows the accuracy results for different values of K.
With the use of K-nearest Neighbor Classifier, with the 3 neighbors, we got better accuracy result. But is is very much lesser than the accuracy we got with the use of ANN as a classifier.

Dataset	**Number of Neighbors**			
	3	5	7	9
MFCC features	0.7017	0.6842	0.6842	0.6842
MFCC + RMS features	0.7017	0.6842	0.6842	0.6842
MFCC + Spectral centroid features	0.7017	0.6842	0.6842	0.6842
MFCC + RMS + centroid features	0.7017	0.6842	0.6842	0.6842

Table 4.9: Accuracy results for different number of K neighbors

4.3 Effect of Loudness on the each regions' folk songs

For overall dataset, it may be quite unfeasible to check the effect of the loudness. It is better to check for each region separately that in which region loudness feature affects more. Table 4.10 and 4.11 shows the comparison results for this case.

Dataset	**Accuracy for K Number of Folds**		
	3 Folds	5 Folds	10 Folds
Assamese	0.9175	0.9245	0.9373
Kannada	0.9615	0.9358	0.9375
Kashmiri	0.7883	0.8036	0.7433
Marathi	0.8995	0.9000	0.8800
Uttarakhandi	0.7179	0.7178	0.6083

Table 4.10: Accuracy results for MFCC dataset of each region

Dataset	**Accuracy for K Number of Folds**		
	3 Folds	5 Folds	10 Folds
Assamese	0.9172	0.9241	0.9171
Kannada	0.8788	0.8648	0.8821
Kashmiri	0.7916	0.8355	0.8350
Marathi	0.8995	0.9000	0.8800
Uttarakhandi	0.6667	0.7714	0.7417

Table 4.11: Accuracy results for MFCC + RMS dataset of each region

From the Table 4.10 and 4.11, it is clearly seen that 10 fold cross validation is better for this classification. And it can be conclude from both the table that loudness feature is important for the Kashmiri and Uttrakhandi folk songs. The difference of accuracy fro kashmiri folk songs is 9.17% (for 10 fold) and for Uttarakhandi folk songs the different is 13.34% (for 10 fold). It shows the effect of loudness in the Kashmiri folk songs.

4.3.1 Results of proposed Method

By following the process mentioned in section 3.6, we could retrieve the prediction results shown in table 4.12. From all the possible 5 fold cross validation process, the highest accuracy we achieved here is 0.817 in the 4^{th} possibility of outer 5 fold cross validation for 39 and 42 values datasets.

Prediction Results for proposed work

For 4th combination of outer 5 fold validation (0.7333 Accuracy)				
	original labels	predicted labels for 39 feature set	predicted labels for 42 feature set	Prediction of SVM
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
5	0	0	0	0
6	0	0	0	0
7	0	0	0	0
8	0	0	0	0
9	0	0	0	0
10	0	0	0	2
11	0	0	0	2
12	0	0	0	0
13	0	0	0	0
14	0	0	0	0
15	0	0	1	1
16	0	0	0	0
17	0	3	0	0
18	0	0	1	2
19	0	0	0	0
20	0	3	0	0
21	0	0	0	0

22	0	0	0	0
23	1	0	1	1
24	1	0	1	0
25	1	0	1	0
26	1	3	1	1
27	1	0	1	1
28	1	0	1	0
29	1	0	1	1
30	1	1	1	1
31	1	1	1	0
32	1	1	1	0
33	1	1	1	0
34	2	1	2	2
35	2	1	0	0
36	2	1	2	0
37	2	1	0	0
38	2	1	2	2
39	3	1	3	2
40	3	1	3	3
41	3	1	3	3
42	3	2	3	3
43	3	4	3	3
44	3	2	3	2
45	3	2	3	0
46	3	2	3	3
47	4	0	2	2
48	4	2	0	2
49	0	3	1	2
50	0	3	0	0
51	0	3	0	0
52	0	3	0	0

53	0	3	3	2
54	0	3	0	0
55	1	1	1	0
56	1	2	1	1
57	2	0	2	2
58	4	4	1	1
59	4	0	4	2
60	4	0	0	0

Table 4.12: Result of Proposed Method

Confusion Matrix for proposed work

Confusion matrix for 39 feature set and 42 feature set is shown in Table 4.13 and 4.14 respectively.

Actual	**Predicted Labels**					**Original number of labels**
	0	**1**	**2**	**3**	**4**	
0	25	1	1	0	1	28
1	0	10	2	0	0	12
2	1	0	7	1	0	9
3	1	0	0	6	0	7
4	0	1	2	0	1	4
Predicted number of Labels	27	12	12	7	2	60

Table 4.13: Confusion Matrix for 39 MFCC features dataset

Recall for 42 dimentiona dataset	0.7222
Precision value for 42 dimention dataet	0.7399

Table 4.14: Precision Recall for 39 dimension dataset

Actual	**Predicted Labels**					**Original Number of Labels**
	0	**1**	**2**	**3**	**4**	
0	18	6	2	1	0	27
1	1	11	1	4	0	17
2	0	0	3	1	0	4
3	0	0	0	5	0	5
4	2	1	3	1	0	7
Predicted Number of Labels	21	18	9	12	0	60

Table 4.15: Confusion Matrix for 42 MFCC+RMS+Spectral Centroid features dataset

Recall for 42 dimentiona dataset	0.6127
Precision value for 42 dimention dataet	0.4436

Table 4.16: Precision Recall for 42 dimension dataset

Chapter 5

Conclusion and Future Work

5.1 Conclusion

We have examined effect of different features on the classification of five regions' Folk songs: Assamese, Kannada, Kashmiri, Marathi and uttrakhandi(Pahadi) folk songs. Also we checked for the different Sampling rate: 16 KHz and 44.1 KHz of the wav file. After experiment of that, results clearly show that we can achieve better accuracy results using 16 KHz audio files than with the use of 44.1 KHz wav files. which shows that the sampling rate is not directly proportional to classification accuracy. Also the combination of different features gives better results rather taking single feature. On some regions, loudness affects to classification that shows the presence of loudness is very important in those regions. After applying the feature selection techniques, we got better or nearer accuracy results with the accuracy results of original full dataset. It shows that there is a presence of some unwanted feature values that reduce the classification functionality of the model. From the prediction of proposed methods, it can be seen that dataset of 39 dimension feature set matches with the original labels than any other dataset prediction labels.

5.2 Future Work

For the future work, we can compare the similarity in music of neighborhood regions which shows the effect of neighborhood region in the culture. Another scope is by recording the voice of folk instruments, we can identify that any folk instrument is used in any song or not. Also we can check if any similar instrument is present in any other region.

References

[1] R. Typke, F. Wiering, R. C. Veltkamp, *et al.*, "A survey of music information retrieval systems," *ISMIR*, pp. 153–160, 2005.

[2] http://myodiasongs.net/category/3168/Marathi_Folk_Songs_Mp3_Songs.html

[3] J. K. H. K. ChulYong Yang*, JongTak Shin, "Korean folk song retrieval using rhythm pattern classification," *Fifth International Symposium on Signal Processing and its Applications, ISSPA '99, Brisbane, Australia*, 1999.

[4] J. Salamon and E. Gomez, "Melody extraction from polyphonic music signals using pitch contour characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 6, pp. 1759 – 1770, 2012.

[5] e. a. R. Heshi, Rushiraj and T. A. for Carnatic Music Processing., "Rhythm and timbre analysis for carnatic music processing," *Proceedings of 3rd International Conference on Advanced Computing, Networking and Informatics. Springer India*, pp. 603–609, 2016.

[6] R. Sridhar and T. Geetha, "Music information retrieval of carnatic songs based on carnatic music singer identification," *Computer and Electrical Engineering, 2008. ICCEE 2008. International Conference on. IEEE*, pp. 407 – 411, 2008.

[7] P. Chordia, "Automatic rag classification using spectrally derived tone profiles," *Proc. of the International Computer Music Conference*, vol. 129, pp. 83 – 87, 2009.

[8] P. Chordia and A. Rae, "Automatic raag classification using pitch-class and pitch-class dyad distributions," *7th International Conference on Music Information Retrieval (ISMIR)*, 2007.

[9] J. J. Parag Chordia and A. Rae, “Automatic carnatic raag classification,” *Journal of the Sangeet Research Academy (Ninaad)*, 2009.

[10] T. C. Nagavi and N. U. Bhajantri, “Overview of automatic indian music information recognition, classification and retrieval systems,” *Recent Trends in Information Systems (ReTIS), 2011 International Conference on. IEEE*, pp. 111 – 116, 2011.

[11] P. Chordia, M. Godfrey, and A. Rae, “Extending content-based recommendation: The case of indian classical music,” *ISMIR*, pp. 571 – 576, 2008.

[12] S. Shetty and K. Achary, “Raga mining of indian music by extracting arohana-avarohana pattern,” *International Journal of Recent Trends in Engineering*, pp. 362–366, 2009.

[13] Y. Liu, J. Xu, L. Wei, and Y. Tian, “The study of the classification of chinese folk songs by regional style,” pp. 657–662, 2007.

[14] S. Khoo, Z. Man, and Z. Cao, “Automatic han chinese folk song classification using the musical feature density map,” *6th International Conference on IEEE*, pp. 1 – 9, 2012.

[15] S. Khoo, Z. Man, and Z. Cao, “German vs. austrian folk song classification,” *Industrial Electronics and Applications (ICIEA), 8th IEEE Conference on IEEE*, pp. 131 – 136, 2013.

[16] C. K. Bassiou, Nikoletta and A. Papazoglou-Chalikias, “Greek folk music classification into two genres using lyrics and audio via canonical correlation analysis,” *Image and Signal Processing and Analysis (ISPA), 2015 9th International Symposium on IEEE*, pp. 238 – 243, 2015.

[17] W. Chai and B. Vercoe, “Folk music classification using hidden markov models,” *Proceedings of International Conference on Artificial Intelligence*, vol. 6, no. 6.4, 2001.

[18] http://www.ourgarhwal.com/uttaranchali-songs.php

[19] http://djsmarathi.com/category/3257/Marathi%20Folk%20Songs.html

[20] https://thinkbangalore.blogspot.in/2013/06/kannada-folk-mp3-songs.html#.WM6SkDuGPIU

[21] http://www.aiomusica.com/music/jammu-kashmir-folk-dance-kashmir-folk-song.html

[22] https://mr-jatt.com/album/assamese/assamese-bihu-songs-zfr.html

[23] http://spicymp3.com/site_goalparia-lokogeet-mp3-songs-download.xhtml

[24] https://mr-jatt.com/album/assamese/assamese-goalparia-lok-geet-jdr.html

[25] https://mr-jatt.com/album/assamese/rameswar-pathak-kamrupi-lokgeet-pvb.html

[26] http://spicymp3.com/site_kamrupi-lokogeet-mp3-songs-download.xhtml

[27] http://spicymp3.com/site_assamese_tokari_geet_free_download.xhtml

[28] X. Huang, A. Acero, H.-W. Hon, and R. Foreword By-Reddy, “Spoken language processing: A guide to theory, algorithm, and system development,” 2001.