GACMIS - Genre Automated Classification using Machine learning for Indian Songs

Ujjwal Singh
IIIT Delhi
Delhi, India
ujjwal18113@iiitd.ac.in

Himanshu Garg
IIIT Delhi
Delhi, India
himanshu18337@iiitd.ac.in

Mayank Joshi
IIIT Delhi
Delhi, India
mayank18048@iiitd.ac.in

Abstract

Automated music genre classification is a very active research topic, as it is of significant importance in music information retrieval system. There is lots of research done in classifying western music genre. India has a heritage of rich music culture, from Purana's,Veda's, to modern times, music is an essential part of Indian households. In India, Music genre differs with the change in geographical area because of different cultures, religions, languages etc. This makes automating human classification abilities for classifying different music genre a difficult task. There is also very few data set to work on, for Indian Music Classification.

In this work, we are releasing a database for Indian songs genre (Bollywood Rap, Ghazal, Garhwali, Bhajan, Bollywood Romantic, Sufi and Bhojpuri) classification and compare the performance of different ML algorithms (SVM, Gradient Boosting, kNN, Light GBM and Neural Network) for the classification task. Light GBM classifier gives the best accuracy of 77.2%.

Link to Github repository.

1. Introduction

Recent internet and technological advancement has helped user to interact with music, by directly consuming the content through music streaming apps, Televisions, Radio etc. Now, users have access of millions of songs on their fingertips. This exploding digital music database needs to be correctly indexed and described so that end-user can interact appropriately with this database. In Music Information Retrieval systems, it analyses the various signals in the music. Genre information is a crucial part of Music Information Retrieval.

A lot of research has been done for the western music genre classification, but very little for Indian music. Since India has a lot of different languages and diversified musical culture, it is challenging for a human being to know all types of music, thus very difficult for a single human being to classify Indian music into different genres.

Any music can be characterized by defining its Top-level, mid-level and low-level feature set. Top Level feature set includes human-defined labels like mood, artist, genre, etc. Mid Level features are beat, pitch, tempo and other rhythmic features. Low Level or short term features include timbral features that are extracted from a small frame. Most commonly used short term features are **Spectral centroid, Spectral roll-off,Spectral flux, MFCC, octave-based Spectral contrast, zero-crossing and low energy, etc.** These features that are extracted from several short frames are integrated to form various temporal features like mean of all the calculated features to get the overall temporal evolution of features across the various frames.

For this work, we have selected more popular Indian music genres, namely **Bollywood-Rap, Bollywood-Romantic, Ghazal, Folk(Garhwali), Sufi, Bhojpuri and Bhajan**. The above genres cover most of the Indian Music culture. The Bhojpuri is popular Music of West India, Garhwali is the music of hilly areas, and Bollywood is the primary music genre of north India, most Indian films music is based on these two genres. Ghazal and Sufi are light classical or semi-classical music.

In this work, we explore the performance of various features extracted from the audio signal in terms of the separability of the seven classes of Indian music using various classification models like **SVM, kNN, ensemble model like Gradient Boosting, Light GBM, and basic Neural Network**
.

2. Related Work

The pioneering work by George Tzanetakis and P. Cook (2002)[11] is the primary motivation for this work. They proposed a music genre classification problem as one of the pattern recognition problems. They released the data

set named GTZAN, for further research in the Music Genre Classification domain. In there work, they made use of statistical features which are derived from the pitch, rhythmic, and timbral content extracted from 30 seconds of music excerpts and employed classifiers to classify the genres. In 2003, Tao li and Tzanetakis proposed that if we mix these timbral features with features like MFCC, results in much better accuracy. Xi Shao, Maddage, M.C., Changsheng Xu and Kankanhalli, M.S. (2005) used Linear prediction cepstrum coefficients, MFCC, zero- crossing rates, Spectrum Flux/power amplitude envelope, and cepstrum flux with Support Vector Machines Classifiers for classifying music.
Recent Work in Classifying Indian Music genres includes the work of S. Jothilakshmi , N. Kathiresan (2012)[5] , they took various Indian music genres namely Hindustani, Carnatic, Ghazal, Folk and Indian western, and try to classify them. They proposed to calculate the performance of classification based on temporal (zero-cross rate, linear prediction coefficients), energy (RMS Energy, the energy of harmonic component of power spectrum), spectral shape (centroid, skewness, kurtosis) and perceptual features (relative specific loudness, sharpness). They got the best classification accuracy for kNN and GMM were 61.25% and 80.63% respectively for the feature combination **MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, irregularity, Rolloff, Spread.**
In work by H. Sharma, RS.Bali (2015)[10]. They have proposed a method for Hindustani raga identification. They classified ragas into four sub Genres i.e. **DES, Bhupali, Yaman, Todi** and Compared four machine learning classifiers on the normalized dataset of ragas. Classifiers used were **Random forest classifier, C4.5(Decision Tree), Bayesian network and K-star**. The highest results accuracy of 93.38% was achieved for K-star learning classifier followed by C4.5(Decision Tree), with 88.6%, Random Forest with 87.6% and Bayesian with 85.1%.
Our work is an extension of above approaches in Classification of Indian Music Genre. None of the above work has taken the mainstream Indian Music into account. Thus, our work focuses on implying above techniques with different ML models to the Major Indian Music ”Hindi”, to perform the automated classification of genres.

3. Feature Set

Features extracted from audio signals give meaningful information, as the audio is characterized by a compact numerical representation. The features extracted are as follows:

3.1. Timbral Features

These features are obtained from the frequency of the audio signal. After converting an audio signal to a frequency domain, the following features are extracted from the spectrum and can help in music genre classification [11].

- **Spectral Centroid:** It is mainly associated with the measure of the brightness of a sound. It is obtained by calculating the “center of gravity” using the Fourier transforms’ frequency and magnitude information. Each centroid of a spectral frame is given as -

$$Spectral\ Centroid\ =\ \frac{\sum_{k=1}^{N} kM[k]}{\sum_{k=1}^{N} M[k]}$$

where M[k] is the magnitude of the Fast Fourier Transform(FFT) at frequency bin k and N is the number of frequency bins.

- **Spectral Roll off:** It is mainly associated with the shape of the spectral and is defined as -

$$Roll\ off\ =\ \sum_{k=1}^{N} M[k]$$

where M[k] is the magnitude of the Fast Fourier Transform(FFT) at frequency bin k and N is the number of frequency bins.

- **Spectral Flux:** Also known as onset strength, It is associated with the rate of change of power spectrum. It denotes how quickly the power changes in between frames and is calculated as -

$$Flux\ =\ F\big\|M[k] - M_p[k]\big\|$$

where M_p[k] denotes the FFT magnitude of the previous frame in time and M[k] represents FFT of the present frame.

- **Zero crossings rate:** It denotes the measure of the noisiness of the audio signal and is defined as -

$$Z_t = \frac{1}{2}\sum_{n=1}^{N}\big|sign(x[n]) - sign(s[n-1])\big|$$

where the sign function is 1 for positive arguments and 0 for negative arguments and x[n] is the time domain signal for the frame.

- **Mel-Frequency Cepstral Coefficients:** MFCC describes the overall shape of a spectral envelope. It is calculated after taking the log-amplitude of the magnitude spectrum, the FFT bins are grouped and smoothed according to the perceptually motivated Mel-frequency scaling. Finally, in order to decorrelate the resulting feature vectors a discrete cosine transform is performed[4]. In Figure 1, we can see that there is huge difference in graphs between various genres. These are graph plotted on one sample song from each of these three genres.

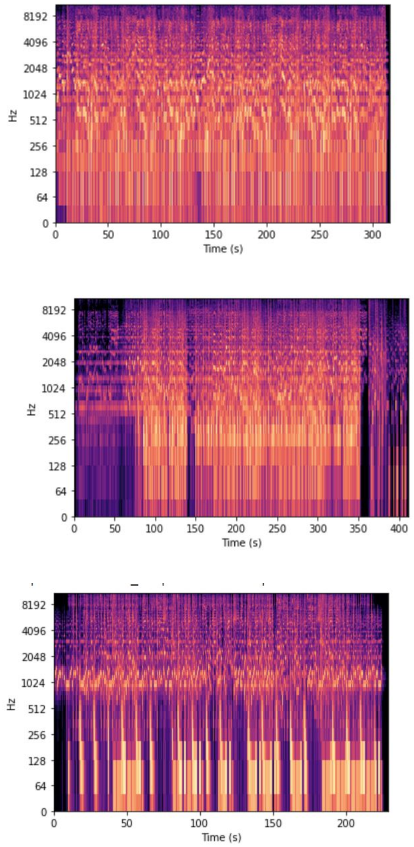


Figure 1. (Top to Bottom) Ghazal, Sufi, Bollywood Rap Log Mel Spectrograph

3.2. Chroma related Features:

- **Chroma stft:** It denotes the chromagram of a waveform or power spectrogram. [1]

- **chroma cens:** It denotes the chroma variant ”Chroma Energy Normalized” (CENS) of an audio signal. It is robust to dynamics, timbre and articulation, thus can be helpful in audio matching [7].

3.3. Other Features:

- **Tonnetz:** Tonnetz or the Harmonic Network is a well known planar representation of pitch relations of an audio signal and can be helpful in detecting Harmonic Change In Musical Audio [4].

4. Dataset

The created dataset consists of the following seven genres: Bollywood Rap, Ghazal, Garhwali, Bhajan, Bollywood Romantic, Sufi and Bhojpuri. The class distribution is represented in Figure 2.

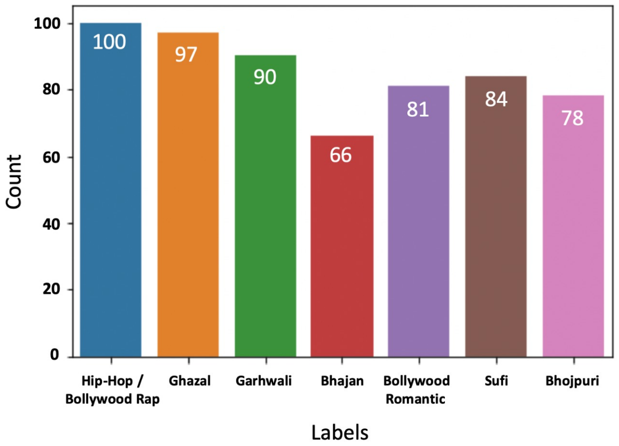


Figure 2. Class Distribution

The songs in each genre are selected if it is available in at least three manually verified Spotify playlists of the same genre. Once again, after manual verification, the complete songs are downloaded from Youtube.

The python library Librosa [7] has been used to create the following features related to each song - **onset strength, chroma stft, chroma cqt, chroma cens, Mel spectrogram, Mel-Frequency Cepstral Coefficients (MFCC), spectral centroid, spectral bandwidth, spectral contrast, spectral roll off, tonnetz, and zero crossing rate.** Then, for the features like MFCC, chroma stft, etc. which are represented using a vector, the mean of the vector is taken as the feature instead of the whole vector.

PCA (Principal Component Analysis) has been used to remove the correlation between the features; in combination with StandardScaler transformation (Standardize features by subtracting the mean and scaling to unit variance) before running SVM. However, StandardScaler transformation without PCA gives better results for KNN classifier.

.

5. Methodology

The following three classifiers have been used for the purpose of classification:

- **Support Vector Machine (SVM):** SVM is a supervised machine learning algorithm which is widely used for both classification and regression purposes. It is based on the idea of finding a hyperplane that maximizes the margin between the two classes. Here, the hyperplane is the decision boundary and support vectors are the points closest to the hyperplane. SVM algorithms use a set of mathematical functions called kernels. These can be linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. We have used RBF for this experiment as it is more flexible and gives access to all infinitely differentiable functions. RBF kernel of two samples x and y is defined by:

$$k(x,y) \;=\; exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$$

where, $\|x-y\|$ is the euclidean distance between the feature vectors and σ is the the Gaussian radial basis function. Here, the cost function 'C' is a parameter indicating the soft margin that controls the influence of each support vector [9].

- **Gradient Boost :** It is based on the fundamental idea that the best possible next model when combined with the previous model, minimizes the overall predicted error. The target outcome of the next model is chosen such that the error is minimized. The decision trees learnt are relatively simple during the initial iterations. As training progresses, the classifier becomes more powerful as it is made to focus more on the wrongly classified instances. The final prediction, in the end, is a weighted linear combination of the output from the individual models [2].

- **KNN (k-nearest neighbors):** KNN is a non-parametric classification method. To classify a data record t, it retrieves its k nearest neighbors, called the neighbourhood of the record t. It is a lazy learning algorithm that doesn't assume any training data generalization. The algorithm works well even when the data is not linearly separable. Only the value of 'K' and the distance metric alone needs to be tuned [3].

- **Neural Network :** It is the network of artificial neurons unit, which tries to mimic the biological neural network, to learn a vast amount of data. A neural network learns through processing the train data samples, which contains the known input and output result. The training of neural network happens, as it calculates the error between the processed output result and the actual result, and then adjust its weights using this error and learning rule specified [8]. Basic intuition is given in Figure 3.

- **Light GBM :** LightGBM comes into light recently. Light GBM model stands for (Gradient Boosting Machine Light), and it was recently introduced by Microsoft team. It is a gradient boosting framework, which uses tree-based decision algorithms. It has many advantages over other gradient methods mentioned above as it is light, easy to use, have lots of parameters and supports parallel learning. It gives higher accuracy than traditional gradient boosting methods [6].

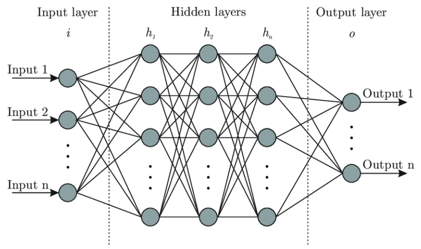


Figure 3. Neural Network Architecture (Credit: A.A. Jimenez)

5.1. Models

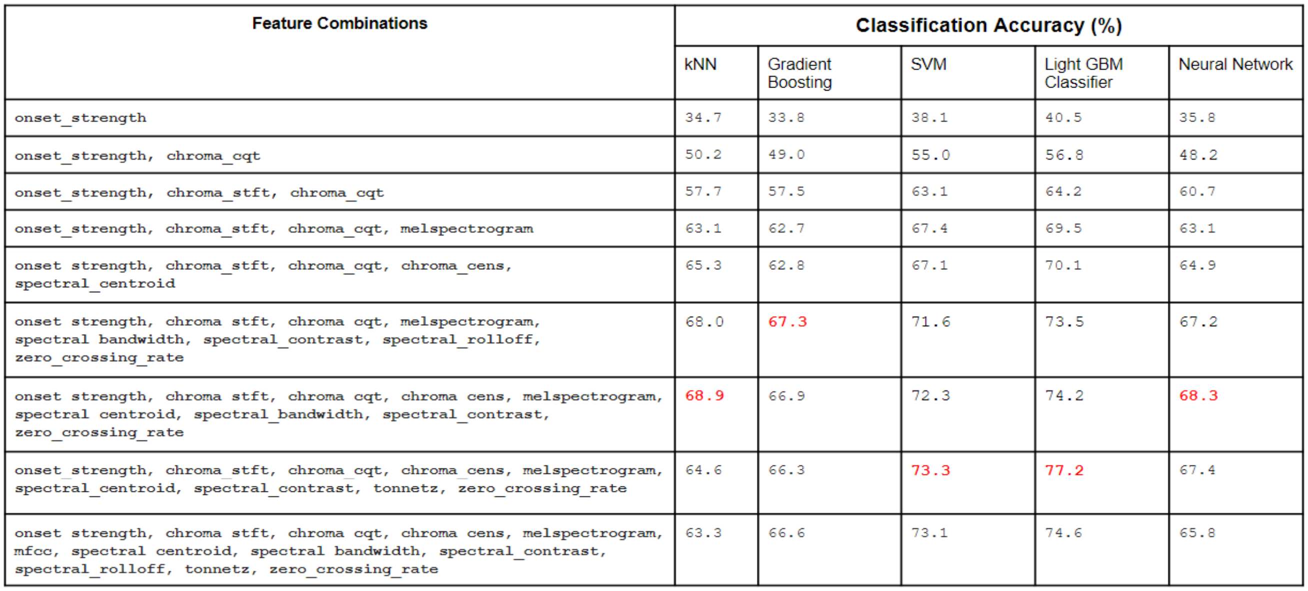
- **Model 1:** It denotes the KNN classifier with 'K' = 7 because it was found to yield higher accuracy. Before training the KNN classifier on the data, data was standardized using Standard Scalar technique (Standardize features by removing the mean and scaling to unit variance) to give better results.

- **Model 2:** It denotes Gradient Boosting classifier with loss function as 'deviance' and the number of boosting stages = 100.

- **Model 3:** It represents an SVM classifier. The Radial Basis Function (RBF) was used as the kernel function as it effectively handles multiclass problems. The value of gamma for RBF kernel was chosen to be 1/number-of-features, and the cost function C = 5 was chosen for this model.
  After applying the PCA technique to the data to remove the correlation between the features, standardization (Standardize features by subtracting the mean and scaling to unit variance) was done before training the SVM classifier.

- **Model 4:** It represents Light GBM classifier. In this model, we have used Gradient Boosting Decision Tree as a boosting function, as it is highly optimized for handling multiclass classification problems. We have also optimized some hyperparameters like learning rate was set to be 0.1, min child samples are set as 20, etc. We got these hyperparameters by using grid search CV.
  Before applying LightGBM, data preprocessing as well as standardisation was done. PCA techniques were applied as well to remove any correlation and redundant features.

- **Model 5:** It represents a Neural Network. Sigmoid is used as hidden layers activation function and Softmax for the output layer. The number of input parameters was 12, and the output parameters were 7 (7 classes in our dataset).
  Size of the hidden layer is 512, 256, 128, respectively.



Feature Combinations	Classification Accuracy (%)				
	kNN	Gradient Boosting	SVM	Light GBM Classifier	Neural Network
onset_strength	34.7	33.8	38.1	40.5	35.8
onset_strength, chroma_cqt	50.2	49.0	55.0	56.8	48.2
onset_strength, chroma_stft, chroma_cqt	57.7	57.5	63.1	64.2	60.7
onset_strength, chroma_stft, chroma_cqt, melspectrogram	63.1	62.7	67.4	69.5	63.1
onset strength, chroma_stft, chroma_cqt, chroma_cens, spectral_centroid	65.3	62.8	67.1	70.1	64.9
onset strength, chroma stft, chroma cqt, melspectrogram, spectral bandwidth, spectral_contrast, spectral_rolloff, zero_crossing_rate	68.0	67.3	71.6	73.5	67.2
onset strength, chroma stft, chroma cqt, chroma cens, melspectrogram, spectral centroid, spectral_bandwidth, spectral_contrast, zero_crossing rate	68.9	66.9	72.3	74.2	68.3
onset_strength, chroma_stft, chroma_cqt, chroma_cens, melspectrogram, spectral_centroid, spectral_contrast, tonnetz, zero_crossing_rate	64.6	66.3	73.3	77.2	67.4
onset strength, chroma stft, chroma cqt, chroma cens, melspectrogram, mfcc, spectral centroid, spectral bandwidth, spectral_contrast, spectral_rolloff, tonnetz, zero_crossing_rate	63.3	66.6	73.1	74.6	65.8

Figure 4. The accuracy score obtained by various models on the dataset using 5 fold cross-validation with stratified sampling.

Model is trained using Keras library and gives relatively low performance due to the small amount of training data.

6. Experimental Results

This section exhibits the results obtained from the experimentation described in the above section. Around 600 songs features are extracted, and then, these features are compiled in CSV files. The training and testing data set is validated by 5 folds cross-validation system with stratified sampling. In Figure 5, we have summarized our all experimentation results.

Model 4 (Light GBM) gives the best results and the highest accuracy of 77.3% using **onset-strength, chroma-stft, chroma-cqt, chroma-cens, melspectrogram, spectral-centroid, spectral-contrast, tonnetz, zero-crossing-rate** as features.

7. Conclusion and Future Work

In this work, we have shown that our Model 4, which is based on LightGBM, is producing the best results with a certain feature set. Interesting observation after our experimentation is that our all classifiers model are getting confused while classifying bhajan and garhwali genre, little EDA shows that these both genres are different from each other, still they share common roots in form instruments being used in them. We have already made our data set along with features public. Each member has contributed equally towards this project, we have divided work by assigning different model to each member, and everyone has done their part. We have learned a lot about ML in practice. We learned a lot about models that we have used for this work, and how to pre-process data and do feature engineering.

A plausible extension of our work is in the field of Music Streaming services. We have shown that even though these 7 genres have so much in common, still can classify them using basic ML techniques. Platforms like Spotify, YouTube Music, Amazon Music can incorporate such models in their apps to get an even better classification of these indigenous music genres. This work also aims to draw the attention of researchers around the world, to work on this problem using the robust dataset that we have made public for their use.

References

[1] Daniel P.W. Ellis. ”chroma feature analysis and synthesis”. http://labrosa.ee.columbia.edu/matlab/chroma-ansyn.

[2] J.H. Friedman. ”stochastic gradient boosting”. Computational statistics data analysis, 1997.

[3] Wang H. Bell D. Bi Y. Guo, G. and K. Greer. ”knn model-based approach in classification”. In OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”, 2003.

[4] Sandler M. Gasser M. Harte, C. ”detecting harmonic change in musical audio.”. In Proceedings of the 1st ACM Workshop on Audio and Music Computing Multimedia, 2006.

[5] S. Jothilakshmi and Nagarajan Kathiresan. ”automatic music genre classification for indian music”. International conference on software and computer applications (ICSCA 2012), IPCSIT. Vol. 41, 2012.

[6] Meng Q. Finley T. Wang T. Chen W. Ma W. Ye Q. Ke, G. and T.Y. Liu. ”lightgbm: A highly efficient gradient boosting

decision tree.". In Advances in neural information processing systems (pp. 3146-3154).

[7] Colin Raffel Dawen Liang Daniel PW Ellis Matt McVicar Eric Battenberg McFee, Brian and Oriol Nieto. "librosa: Audio and music signal analysis in python.". In Proceedings of the 14th python in science conference, pp. 18-25, 2015.

[8] M.M.. Poulton. "neural networks as an intelligence amplification tool: A review of applications". Geophysics, 67(3), pp.979-993.

[9] Sung K. K. Burges C. J. Girosi F. Niyogi P. Poggio T. Vapnik V. Scholkopf, B. "comparing support vector machines with gaussian kernels to radial basis function classifiers.". IEEE transactions on Signal Processing, 45(11), 1997.

[10] Hiteshwari Sharma and Rasmeet S.Bali. "comparison of ml classifiers for raga recognition". in International Journal of Scientific and Research Publications, October 2015.

[11] G. Tzanetakis and P. Cook. "musical genre classification of audio signals". in IEEE Transactions on Speech and Audio Processing, 2002.