2018 17th IEEE International Conference on Machine Learning and Applications

Similarity Estimation for Classical Indian Music

Anusha Sridharan
Department of Computer Science
San Jose State University
San Jose, CA, USA
anusha.sridharan@sjsu.edu

Melody Moh
Department of Computer Science
San Jose State University
San Jose, CA, USA
melody.moh@sjsu.edu

Teng-Sheng Moh
Department of Computer Science
San Jose State University
San Jose, CA, USA
teng.moh@sjsu.edu

***Abstract*— Music is a complicated form of communication, where creators and cultures communicate and expose their individualities. Thanks to music digitalization, recommendation systems and other online services have become indispensable in the field of Music Information Retrieval (MIR). Classification of music is essential for music recommendation systems. In this paper, we propose an approach for finding similarity between music. Our approach is based on mid-level attributes like pitch, midi value, interval, contour, and duration, and applying text-based classification techniques. Performance evaluation has been done using the accuracy score of scikit-learn. As a preliminary study, our system first predicted jazz, metal, and ragtime for western music. The genre prediction system has been tested on 476 music files with a maximum accuracy of 95.8% across different n-grams. Then, we have analyzed and classified the Indian classical Carnatic music based on their raga. Our system has predicted Sankarabharam, Mohanam, and Sindhubhairavi ragas. The raga prediction system was tested on 68 music files with a maximum accuracy of 90.14% across different n-grams.**

Keywords*— *Evaluation Matrix, Machine Learning, Music Information Retrieval (MIR), Natural Language Processing, Raga Classification, Text-based Classification

I. Introduction

Classification of music based on genre has gained its popularity in both Music Information Retrieval (MIR) and Machine Learning (ML) domains in the past decade. Many music streaming platforms like Spotify, Pandora, and Saavn use automated music recommendation services. The accessibility of music has given rise to the necessity for developing tools to effectively manage and retrieve music that interests end users [1]. The primary step to manage and track these systems is to classify music. Classification of music is important because typically one tends to listen to varied music within a specific set of music pattern and classification would be beneficial to recommend and promote desired music[3]. Classification can be based on various parameters like genre, emotion, and mood. When a person listens to music, she recognizes the feel, culture, and emotion of the music and cannot relate to acoustics of the sound wave. Due to its acoustic and cultural complexity, music classification is an intricate process.

Genre classification in western music involves categorizing music by genre and tagging them as ‘blues’, ‘pop’, ‘rock’, or ‘classical’ for example. Music emotion classification is done through two scales: Valence-Arousal Scale and Geneva Emotional Music Scale (GEMS). Valence-Arousal Scale classifies music into four types, namely happy, sad, angry, and relaxed. GEMS classifies music into nine categories, namely power, joy, calmness, wonder, tenderness, transcendence, nostalgia, tension, and sadness [18].

Music Genre Classification is a familiar problem in the field of MIR. Common approaches for genre classification use low-level features like Fast Fourier Transform (FFT) and Mel Frequency Cepstral Coefficient (MFCC) [2]. To overcome the limitation of using low-level features, we use mid-level features like rhythm, pitch, and harmony. Mid-level features are widely used in problems like cover song detection and query by example. These features are relatable to music, but it is challenging to extract them[5].

On the other hand, Indian classical Carnatic music is disparate from western music[21]. Carnatic music has its origin in southern India. Classification of Indian classical Carnatic music is performed based on raga, tala, and artist. Raga is a sequence of swaras (notes) and is defined by arohana and avarohana. Arohana is ascending scale of notes, and avarohana is descending scale of notes. Tala refers to rhythmical cycle, similar to measure in western music. The swaras Sa Re Ga Ma Pa Da Ni correspond to C D E F G A B in western music[38].

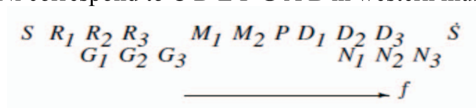

Fig 1. Frequency scale of swaras

Fig 1. shows swaras and in the increasing order of their frequencies of Carnatic music. Some of the swaras have different variations of it. E.g. Ga has three variations Ga1, Ga2, and Ga3. It is also important to note that Ri2 and Ga1 lie in the same frequency scale, so they cannot occur together in a raga. Every song is based out of a base pitch value called Sruti[38].

Raga identification is essential because raga can be associated with mood and emotion. Some ragas are also found to have therapeutic effects in our body [19][20]. Raga identification requires the in-depth knowledge in the raga, and precision can only be achieved through practice and experience. There are 72 parent ragas called Melakarta ragas. There are thousands of ragas derived from the Melakarta ragas, and they are called Janya ragas. Melakarta ragas are also called Janaka ragas. The Melakarta ragas have their arohana strictly in ascending order and their avarohana strictly in descending order. It is also implied that two variations of same swara cannot occur in the same raga.

Gamaka refers to the embellishment of swara (note) where the note is sung with oscillations imposed on it and the transition from one note to another is also sung with oscillations. Raga with same swaras can be distinguished by gamakas. Bhava is the

978-1-5386-6805-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMLA.2018.00130

emotion expressed in the song. The complexities present in Carnatic music makes it difficult to represent and classify Carnatic music.[22][23][27]

This paper attempts to propose an approach for classification of western and carnatic music by extracting musicological features and applying machine learning algorithms to identify the genre of a western music and raga of a Carnatic music.

This paper is organized as follows: the next section (section II) presents the most recent related works. Section III contains the description of data used for experiments and data preprocessing. Experiments conducted and results achieved are detailed in Section IV, and in section V, the results are analyzed. Finally, a conclusion is presented in VI.

II. RELATED WORKS

A. *Music Information Retrieval*

Music Information Retrieval (MIR) is the discipline of retrieving meaningful information from musicMIR aims at making music available to listeners based on their interests [3]. In the early stages, MIR research focused on illustrating music in a structured way, including representing music in digital format like midi, au, mp3 and wav. Markus et. al reason how the research on MIR started progressing. As digitization became prominent, various signal processing techniques helped in deriving various music qualities like rhythm, timbre, melody, and harmony [5]. As per Lopez et. al, a music piece was[21] treated as a sound wave, and system-centric features were derived from it to categorize it based on a musical or cultural aspect in the initial stages[4]. Recent research in MIR is making a shift by moving away from system-centric features to user-centric features. Casey et al. [5] insist on the importance, serendipity, and time-awareness of user-centric features which is widely used in recommendation systems these days.

B. *Music Classification*

When building a music genre classification method, researchers including, Fu et al [2], use three common descriptors: low-level features, mid-level features, and high-level content descriptors.

Fujinaga and McKay, and Lopez et. al suggest approaches by performing genre classification in a musicological way[4][10]. The first work on analyzing music using mid-level features started in 2004 by Fujinaga and McKay. Fujinaga and McKay worked on analyzing statistical distributions of feature set and applying various machine learning algorithms on it [10].

Zheng et al. [1] propose a solution which extracts sequential features from symbolic music representations. Melody and bass features of pitch, midi values, duration, contour, and interval are chosen. Each feature element is considered as a word and text-based technique called n-gram is performed. Based on the value of n given, word frequency count vector is derived. Multinomial Naïve Bayes classification is performed with word frequency count vector as the input.

The classification used for text-based data (converting music signals to notes) involve pattern recognition. K-Nearest Neighbor (KNN), Gaussian Mixture Model (GMM), Support Vector Machines (SVM), and Multinomial Naïve Bayes are best suited for finding patterns in text-based data [2]. KNN and SVM are suited for single vector representation and pairwise vector representation.

CNN can be directly used for classification using feature set. Therefore, CNN can be used for audio classification based on low-level as well as mid-level features. This is exhibited in [7] by applying a Convolutional Deep Belief Network (CDBN), which is a type of CNN having multiple layers.

In Indian classical music, the initial step involves extracting swaras from the melodic stream where shadja, the base note is identified and the relative notes to it derived based on the shadja. Ranjani et al [22] propose a technique to extract shadja from pitch values where shadja is decided based on the pitch values present in various windows in the music file. Based on the shadja, the other notes are calculated and raga prediction is done. Geetha et. al's [27] work involves extracting swaras using segmentation algorithm and prediction by using string matching algorithm. Vijay et. al [23] propose a technique where pitch values are extracted from the melodic stream, n-gram pitch histograms are obtained and raga prediction is done using pitch histograms.

On the other hand, using low-level features for raga identification has been researched by Srinath et. al [38]. In their research, frequencies are converted to specmurt which is a fast Fourier transformation of the linearly scaled spectrum. Shadja is then found based on the specmurt, and based on shadja the other swaras are obtained. The prediction is done using hidden Markov models and the system is proved to perform well for Melakarta ragas. Anita et. al [34] have performed classification of ragas using neural networks. Spectral, timbre, and tonal features are extracted, and classification is performed with back propagation neural network and counter propagation network. The experiments are restricted to Melakarta ragas.

III. DATA PREPARATION

A. *Dataset Preparation*

In this section, we first describe dataset used for our experiment and the operations performed on the dataset which constituted midi and mp3 files.

1) *Dataset for western music*

The dataset used for experiments is a set of 476 files picked from 130000_Pop_Rock_Classical_Videogame_EDM_MIDI_Archive dataset[36]. The files constituted of three genres – jazz, metal and ragtime. The dataset contains piano and guitar files with a mix of various artists. Due to the difficulty in using a universal/commonly used MIDI dataset, we have handpicked the files and used it in our experiments. MIDI format files are chosen because it records the musicological aspects of a music which includes notes of each instrument, type of instrument, loudness, finishes, pitch, etc.

Table 1 contains the genres used and the number of files used in each genre.

TABLE 1. DATASET – WESTERN MUSIC

Genre	Number of Records
Jazz	244
Metal	118

Ragtime	114

2) Dataset for Carnatic music

The dataset used for experiments contains mp3 files from three different ragas taken from dunya corpmusic[21]. The mp3 files are converted to midi because mp3 is lossy compressed encoding which makes it difficult to extract features. [24] is used for conversion. The bpm values used for the experiments is 120. The shadja and swaras in Carnatic music are extracted using music21 python package. Apart from the extracting the swaras, we also extract the duration a swara lasts. The ragas used for experiments are listed in table 2. Out of the three ragas, sankarabharanam is Melakarta raga and sindhubhairavi and mohanam are Janya ragas.

TABLE 2. DATASET – CARNATIC MUSIC

Raga	Number of Records
Sankarabharanam	19
Mohanam	23
Sindhubhairavi	26

The midi files are read as data streams using music21 python package and labeled with genre name. The highest and lowest score of stream extracted from each music file is used to extract melodic and bass features. The features are then converted to text and n-gram technique is applied to the text features. Classification techniques are applied on n-gram count vector and the genre is predicted.

B. Data Pre-processing

In the data pre-processing step, we derive the melodic and bass features from the western music dataset, convert them to text, apply rules to handle special characters and write the data to a CSV file. Fig 2. shows diagrammatic representation of data flow in the pre-processing phase.

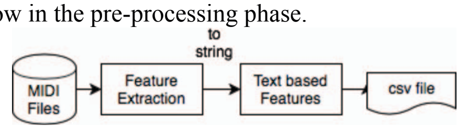

Fig 2. Pre-processing data flow pipeline

Fig 3. Shows the pre-processing data flow pipeline of Carnatic music. The difference in preprocessing is that the count features are separated and only sequential features are converted to text. After the conversion, the two data frames are concatenated and the entire contents are fed to csv file.

In case of Carnatic music, we derive the melodic attributes, convert the pitch, duration, midi, pitch contour, interval, and rest duration to text. The rest of the features which are midi counts and note counts remain as numeric in the data frame. We then combine both text and numeric features and write it to a csv file. In the next step where we derive n-grams, separate text features and apply n-grams only on text features.

1) Feature extraction

a. Western Music Dataset

We are proposing an approach to extract mid-level features like pitch, midi, duration, interval, contour, and rest duration of both melody and bass. The features we extract here are short-term features as the live for a very short time. From these short-term features, we try to understand the long-term effect of the music piece. Melody features are obtained by parsing highest notes and bass features are obtained by parsing lowest notes.

The features extracted are pitch, midi, duration, contour, interval, and rest duration. Pitch is the property which refers to highness or lowness of note[15]. There could be duplicate pitch names like Eb5 and D#5. We are treating them differently as appearing in the midi file. Each pitch has a numerical value associated with it and that is the midi value. Midi value ranges from 0 to 127. The higher the value of midi, the higher the pitch. Duration refers to the time which a note lasts. This feature is to understand the tempo of a song. Duration is expressed in quarter length. Contour is a signed integer that represents the difference between two consecutive midi values. It is to denote whether the notes are increasing or decreasing. Contour is obtained by subtracting midi value of previous note from that of the current note. Interval refers to the name of pitch interval between two consecutive notes. Interval is derived by combining previous and current notes and obtaining their pitch interval. Rest refers to the interval of silence in the music stream and is expressed as “rest” to denote the pause. Rest duration is the time taken by a pause and is measured in quarter length. These two attributes are used to understand the tempo of the song.

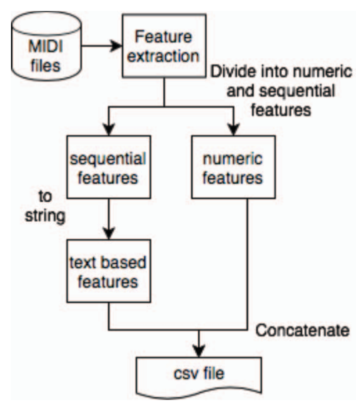

Fig 3. Preprocessing data flow pipeline of Carnatic music

The above attributes are calculated for both melody and bass streams. All the above attributes are extracted and stored in a panda data frame. The values in each of the fields are converted to text. The special characters and space in each of the values are replaced with underscore to eliminate discrepancies. The final data frame obtained is stored in a CSV file.

b. Carnatic Music Dataset

Raga classification is done based on the melodic notes of the song. So instead of extracting the features for both bass and melodic stream, the features are extracted only from melodic stream. This is because unlike western music, raga is determined by melodic notes of the song. Apart from extracting the features mentioned in the previous section, the count of midi values and note values are also calculated. Emphasis is given to note values because each raga has its own melodic flow of notes. The sequence of notes is spread across the song and it is challenging even for an experienced musician to find which part of the song has the notes identifying the raga. To overcome this difficulty, we compute different values of n-grams of note and pitch sequences and the duration at which each of it occurs. We also calculate the count of midi values and note values as songs of the same raga are found to revolve around those pitch values and notes.

C. *N-gram count vector*

N-gram refers to a continuous sequence of n items in given text. We use CountVectorizer function from scikit-learn to calculate n-gram matrix. We specify the values as (2,2) and (1,3) for 2-gram and 1,2 and 3-gram vectors respectively. We conducted our experiments for various values of n-gram and maximum accuracy is found to be achieved with 3-gram in western music. In case of Carnatic music, the maximum accuracy is achieved using 4,6 grams because a raga's arohana and avarohana have five to eight notes.

D. *Classification Pipeline*

After pre-processing, n-gram count vectors are obtained and 5-fold cross-validation is applied. The accuracy is averaged. Data is shuffled while dividing data for training and testing. Train data has 80% of the total set and test data has 20% of the total set. We use Multinomial Naïve Bayes and Random forest classification algorithms for classification. Fig 4. Shows the diagrammatic representation of N-gram count vector and classification pipeline.

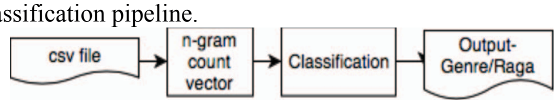

Fig 4. Classification Pipeline data flow

The classification pipeline for Carnatic music slightly varies when compared to western music. The columns in the csv file are divided into text and numeric attributes and n-gram count vectorizer is applied only on text-based columns. The other columns are combined to counter vectorized text columns. The data is shuffled and divided into test and train sets in the ratio 80:20. Dimensionality reduction is performed using SelectKBest function from scikit learn and the number of columns involved in training and prediction is reduced to 6932 columns. The number 6932 is decided by experimenting with random values of k between 50 and 15000 and the best performing value is selected. After dimensionality reduction, the model is trained and prediction is performed. The process is repeated five times for each split in test sets. The result is the average of five different accuracy values obtained. Multinomial Naïve Bayes and Random Forest are the classification algorithms used on preprocessed dataset.

Apart from the two algorithms we already used for western music dataset, we also performed classification using neural networks. There are 3 layers in the neural network. Sigmoid function is applied or the first layer. Sigmoid function works well for classification. It identifies the gradient which decides the output. Relu function is applied on the second layer because it does not saturate gradients and tend to converge faster. Softmax function is applied on the output layer because it gives categorical distribution for classification problems. The error is measured by 'categorical_crossentropy'.The Sigmoid function is defined as a real-valued, differentiable and a monotonic function to introduce non-linearity. The relu is defined as the maximum of the given value and 0. The softmax function is usually used in final layer of neural networks for multiclass classification. Dropout is used as the regularization method and is applied in layer 1 and layer 2. The maximum accuracy achieved using neural networks is 80.28% and it is achieved with n-gram (4,4).

Python package music21[15] is used to derive musicological features and scikit-learn[14], pandas, scipy and numpy packages are used for data pre-processing, prediction and results evaluation. Keras[37] is used for building neural networks. The libraries used are free and open source. We evaluate our results using accuracy. Accuracy is defined as the ratio of number of music pieces identified correctly by the classification algorithm to the total number of music pieces fed to the model for testing.

IV. EXPERIMENTS AND RESULTS

The data from csv file is read and classification is performed for various values of n-gram.

A. *Western Music*

For each value of n-gram the prediction is performed and results are evaluated. The values of n-grams used in the experiments are (1,1), (1,2), (1,3), (2,2) and (3,3). Table 3 provides details about the results obtained by various n-gram values.

TABLE 3. RESULTS EVALUATION – WESTERN MUSIC

N-Gram	State of art approach	Multinomial Naïve Bayes	Random Forest
(1,1)	87.24%	86.37%	94.79%
(1,2)	86.45%	84.37%	94.41%
(2,2)	87.5%	82.29%	93.75%
(1,3)	88.54%	84.37%	92.7%
(2,3)	88.02%	87.34%	94.79%
(3,3)	86.45%	87.5%	95.83%

B. *Carnatic Music*

The experiments of Carnatic music are conducted for a wide range of n-grams ranging from (1,1) to (8,8). This is because arohana and avarohana have around 8 notes in it and having a n-gram vector upto length 8 would be useful. Table 4. shows the top 6 results obtained for the exhaustive values of n-grams tested for midi extracted with 120 bpm value.

TABLE 4. RESULTS EVALUTION – CARNATIC MUSIC – 120 BPM

N-gram	Multinomial Naïve Bayes	Random Forest	Neural Networks
2,8	88.73%	75.1%	69.01%
4,4	83.37%	70.42%	80.28%
4,5	86.25%	76.05%	75.64%
4,6	90.14%	72.1%	78.87%
4,8	84.5%	71.83%	66.19%
5,5	87.275%	78.87%	69.01%

5,8	88.73%	66.19%	72.1%
6,7	82.7%	70.42%	67.61%

Best results are obtained with values of n-grams as 4,6. Table 5. shows the top 6 results obtained for n-grams from (1,1) to (8,8) tested for midi extracted with 180 bpm value.

TABLE 5. RESULTS EVALUTION – CARNATIC MUSIC – 180 BPM

N-gram	Multinomial Naïve Bayes	Random Forest	Neural networks
2,5	84.5%	77.47%	66.19%
3,6	83.09%	81.69%	73.24%
4,4	83.09%	76.05%	78.87%
4,5	88.73%	77.46%	70.42%
5,7	84.5%	78.34%	70.42%
6,7	85.91%	71.83%	72.1%
7,7	87.42%	85.49%	67.61%

The results of 180 bpm are found to be slightly lesser than results of 120 bpm, highest being 88.73% using n-gram (4,5).

V. Results Evaluation

From the results obtained for western music, the best results for genre classification is obtained with (3,3) using Random Forest algorithm. Other values of n-gram have consistent values with an average of 95.83%. Fig 5. gives a comparison of accuracy obtained using approach [1] and proposed approach. When we compare the results, we see that the proposed approach gives consistent results compared to [1].

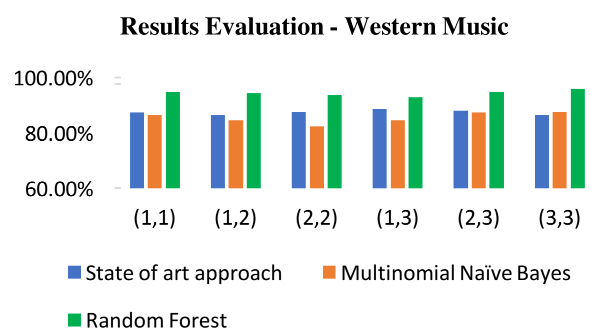

Fig 5. Comparison of Accuracy [1] vs proposed approach

From the comparison, we can clearly see that the accuracy has been constantly increased when compared to [1]. Including new features and performing classification using Random Forest algorithm has improved the prediction of genre when compared to using Multinomial Naïve Bayes algorithm. Genre of the music is determined by speed, rhythm, and progression of the song and using ensemble techniques works better than Multinomial Naïve Bayes.

On the contrary, with Indian classical music, the best results are obtained using Multinomial Naïve Bayes algorithm. This is because Multinomial Naïve Bayes works well when data with same class values have repeated patterns of text values. Here in our case, the arohana and the avarohana of music pieces having the same raga files is supposed to repeat. The values of n-gram are also much higher when compared to western music because Carnatic music raga has the same pattern of notes in an octave and that is the reason for testing values of n-grams upto 8. The length of arohana and avarohana is usually of length 8 but due to the improvisations done by musicians, the exact arohana and avarohana do not appear in the extracted pitches. The best results with the value of 90.14 is obtained for n-gram (4,6) for bpm value of 120. This implies that notes of 4 to 6 length appear in the music file following the pattern of the raga. The baseline results for Carnatic music dataset is obtained from [22] and the best achieved by baseline is 84.6 % using 4-gram. The results obtained by the proposed approach shows an improvement of 6%. Fig. 6. shows accuracy obtained by machine learning algorithms implemented.

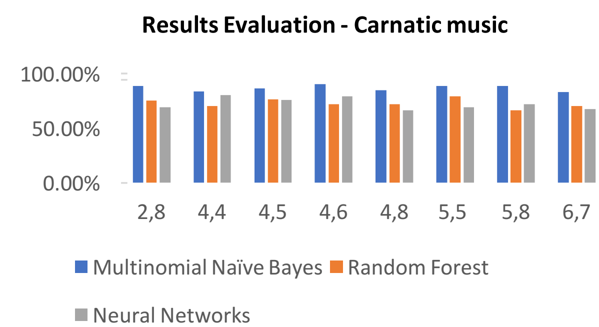

Fig. 6. Accuracy comparison of various approaches in Carnatic music

VI. Conclusion and Future work

With the advancement in the field of MIR, music similarity estimation in a musicological way has proved to provide better solutions for many MIR problems. The research validates that the approach of using mid-level features can overcome the limitation of using low-level features and build a solution which could be logically related to features understood by listeners.

The research has demonstrated that the approach of using mid-level features has proven its concept, and has a promising future for further developments. This includes extracting, computing, and combining more features, using comprehensive instrumental music datasets, and incorporating other advanced text classification techniques. The techniques discussed in the paper is not only useful for genre classification but can also be extended to other classification systems like emotion classification and query by humming problems.

On the other hand, for carnatic music, previous researchers have experimented on mid-level features as well as low-level features. In the proposed approach we have explored the idea of using midi counts and note counts. This could be further improvised by using the duration of a swara. For every n-gram, instead of incrementing the count, the duration for which the swara pattern lives can be incremented. This would help us to identify the movement of swaras, which varies from one raga to another raga. We have seen that using mid-level features in Carnatic music has proved its concept. We can also experiment on a combination of low-level and mid-level features and use it for prediction. The current research is limited to the compmusic dataset and training has been done only with less than a hundred files. The train data could be expanded or the same experiments can be conducted with a different dataset.

REFERENCES

[1] E. Zheng, M. Moh, and T. Moh, "Music genre classification: A n-gram based musicological approach," in *the Proc. of IEEE 7th Int. Advance Computing Conf.,* 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7976875/

[2] Z. Fu, G. Lu, K. Ting and D. Zhang, "A Survey of Audio-Based Music Classification and Annotation," in *the Proc. of IEEE Transactions, 2011.*

[3] T. Huang, and P. Chang, "Large-scale cover song retrieval system developed using machine learning approaches," in the *Proc. of IEEE Int. Symp. on Multimedia*, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7823695/

[4] A. T. Lopez, E. Aguiar, and T.O. Santos, "A comparative study of classifiers for music genre classification based on feature extractors," in *the Proc. Of Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)*, 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7806258/

[5] M. Schedl, E. Gomez, J. Urbano, "Music information retrieval: recent developments and applications," in Music Information Retrieval vol.8, no 2-3, pp 127-261, Jan 2014. [Online]. Available: http://www.nowpublishers.com/article/Details/INR-042

[6] A. R. Rajanna et. al, "Deep neural networks: a case study for music genre classification," in *Proc. Machine Learning and Applications (ICMLA), IEEE 14th Int. Conf*, 2016 [Online]. Available: http://ieeexplore.ieee.org/document/7424393/

[7] http://cmc.music.columbia.edu/MusicAndComputers/chapter2/02_01.php

[8] H. Lee, Y. Largman, P. Pham, and A. Y. Ng, "Unsupervised feature learning for audio classification using convolutional deep belief networks," in *Proc. Advances in Neural Information Processing Systems,* 2009.

[9] M. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: Current directions and future challenges," in *Proc. IEEE, vol. 96, no. 4, pp. 668–696,* 2008.

[10] C. McKay and I. Fujinaga, "Automatic genre classification using large high-level musical feature sets." in *ISMIR*, vol. 2004, 2004, pp. 525– 530.

[11] https://www.nap.edu/read/6323/chapter/11

[12] E. Pampalk, T. Pohle, and G. Widmer, "Dynamic playlist generation based on skipping behaviour," In the Proc. of *ISMIR 2005 Sixth Int. Conf. on Music Information Retrieval*, September 2005.

[13] K. West and S. Cox, "Finding an optimal segmentation for audio genre classification," *In the Proc. ofISMIR 2005 6th Int. Conf. on Music Information Retrieval,* September 2005.

[14] scikit-learn: machine learning in python. [Online]. Available: http://scikit-learn.org/

[15] music21: a toolkit for computer-aided musicology. [Online]. Available: http://web.mit.edu/music21/

[16] C. D. Manning, P. Raghavan, and H. Schutze, Introduction to information retrieval. Cambridge university press Cambridge, 2008, vol. 1.

[17] R. Hillewaere, B. Manderick, and D. Conklin, "String methods for folk tune genre classification." *in the Proc.* ISMIR, vol. 2012, 2012, p. 13th.

[18] J. Jakubik H. Kwaśnicka , "Music emotion analysis using semantic embedding recurrent neural networks," in *Proc. Innovations in Intelligent Systems and Applications (INISTA), 2017 IEEE International Conf*[Online]. Available: http://ieeexplore.ieee.org/document/8001169/

[19] http://pilu.in/raga-therapy.html

[20] https://ayurveda-foryou.com/music/raga_chikitsa.html

[21] A. Srinivasamurthy., & X. Serra. "A Supervised Approach to Hierarchical Metrical Cycle Tracking from Audio Music Recordings". *In Proceedings of the 39th IEEE International Conference on Acoustics*, Speech and Signal Processing (ICASSP 2014) (pp. 5237–5241). Florence, Italy.

[22] H. G. Ranjani, G. Arthi, T.V. Sreenivas," Carnatic music analysis: shadja, swara identification and raga verification in alapana using stochastic models," *In Proc. of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2011.*

[23] V. Kumar, H. Pandya and C. V. Jawahar, "Identifying Ragas in Indian music," *In Proc. of International Conference on Pattern Recognition*), 2014

[24] https://github.com/justinsalamon/audio_to_midi_melodia

[25] http://essentia.upf.edu/

[26] https://librosa.github.io/librosa/

[27] N. A. Jairazbhoy, "C. SubrahmanyaAyyar: Acoustics for music students",[viii], 71 pp., 3 plates., *Bulletin of the School of Oriental and African Studies*, vol. 24, no. 01, p. 71, Dec. 2009.R. Sridhar and T. V. Geetha, Swara identification of Carnatic music, *in Proc. IEEE Computer Society Press, Proceeding of ICIT 2006*

[28] K. R. Scherer M. Zentner "Emotion effects of music: Production rules" in Music and emotion: Theory and research Oxford University Press pp. 361-392 2001.

[29] S. Sigtia S. Dixon "Improved music feature learning with deep neural networks" *in Proc. of the 38th International Conference on Acoustics Speech and Signal Processing* (ICASSP) pp. 6959-6963 2014.

[30] http://harmonyom.blogspot.com/2012/05/

[31] https://www.youtube.com/

[32] M. Henaff K. Jarrett K. Kavukcuoglu Y. LeCun "Unsupervised learning of sparse features for scalable audio classification" ISMIR vol. 11 no. 445 2011

[33] http://harmonyom.blogspot.com/2012/05/

[34] N. Glazyrin "Mid-level features for audio chord recognition using a deep neural network" UchenyeZapiskiKazanskogoUniversiteta. SeriyaFiziko-MatematicheskieNauki vol. 155 no. 4 pp. 109-117 2013.

[35] R. Anita, K. Gunavathi, A. Asokan, "Classification of Melakartha Ragas Using Neural Networks,"in the proc. 2017 International Conference on Innovations in Information, Embedded and Communication Systems

[36] https://www.reddit.com/r/datasets/comments/3akhxy/the_largest_midi_collection_on_the_internet/

[37] https://keras.io

[38] Krishna, Arvind S. et al. "Identification of Carnatic raagas using Hidden Markov Models." *2011 IEEE 9th International Symposium on Applied Machine Intelligence and Informatics (SAMI)* (2011): 107-110.