MUSICAL GENRE CLASSIFICATION OF RECORDED SONGS BASED ON MUSIC STRUCTURE SIMILARITY

Rajitha Peiris, Lakshman Jayaratne
University of Colombo School of Computing
No 35, Reid Avenue, Colombo 07, Sri Lanka
rajitha11ucsc@gmail.com, klj@ucsc.cmb.ac.lk

ABSTRACT: *Automatic music genre classification is a research area that is increasing in popularity. Most researchers on this research area have been focusing on combining information from different sources than the musical signal itself. This paper presents a novel approach for the automatic music genre classification problem using audio signal for the context of Sri Lankan Music. The proposed approach uses two feature vectors and Support Vector Machine (SVM) classifier with radial-basis kernel function. More specifically, two feature sets for representing frequency domain, temporal domain, cepstral domain and modulation frequency domain audio features are proposed through this work. Music genre classification accuracy of 74.5% was recorded as the highest overall classification accuracy on our dataset containing over 100 songs over the five musical genres. This approach shows that it is possible to implement a genre classification model with a reasonably good accuracy by using different types of domain based audio features.*

KEYWORDS: Musical Genre Classification, Audio Signal Analysis, Music Information Retrieval, Feature Extraction, SVM

INTRODUCTION

Music can be divided into many categories mainly based on style, tempo and cultural background. These styles are what is known as music genres. Musical genres are categorical labels created by human experts in the field of music and these are used for describing, storing, categorizing and even comparing songs, albums, or authors in the vast universe of music [1]. There are various high-level descriptors, such as genre, instrumentation, mood and artist to describe music. Musical genre is among the main high-level descriptors and it encapsulates semantic information of the given music track. Today, a personal music collection may contain thousands of songs, while professional music collections typically contain millions of songs in their databases [18]. Most of the current universal music databases are indexed based on artist, title, album and genre of a song [18]. When songs are indexed improperly in the databases, it can cause unexpected search results. The addition of genre as an index to a song has made browsing and searching such large music collections very efficient and effective.

Although genre classifications exist in the world music, there are no verified genres available for Sri Lankan music still. Hence there exists a need for a proper genre classification system for the Sri Lankan music context. Most music listeners are interested in listening to their favorite types of music. Therefore, a music genre classification system would enable them to search for the music they are interested in

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

and even enable them to create their own playlists comprising of their desired genres. Traditionally, music genres are labeled by human musical experts. Therefore these labels are based on the expert opinion of different human experts, hence one can argue on generalizing this process.

Human perception of music is dependent on a variety of personal, cultural and emotional aspects. Therefore its genre classification results may avoid clear definition and the boundaries among genres can be fuzzy [2]. Therefore a scientific approach to introduce a genre classification system for the Sri Lankan music would provide a solution to this problem. In this paper we present a novel approach for automatically classifying audio signals into musical genres using a supervised learning model.

We proposed frequency domain, temporal domain, cepstral domain and modulation frequency domain audio features and two types of feature vectors are designed for individual classification according to short term and long term based features. From the classifier comparison we carried out, the SVM emerged as the winning classifier in all experiments. Therefore SVM classifiers are employed as the base classifiers for each of the feature vectors. Using different feature sets we were able to identify the best set of features for our genre classification model and we achieved a 74.5% overall accuracy for that model.

A brief overview on related work is provided in section II. Feature extraction and the different domains of features used in this research are described in section III. Section IV discusses regarding the experiment strategies, feature selection process and about building the classifier model along with the evaluation of the proposed model and finally, Section V provides conclusions and an outlook on future work.

RELATED WORK

A. Musical Genre Classification

Automatic music genre classification does not have a long history but there has certainly been a rising interest in the last five to ten years. It is an interdisciplinary research area connected to especially from areas such as digital signal processing, machine learning, and music theory [18]. One of the most significant proposals specifically to deal with studies on automatic musical genre classification was proposed by Tzanetakis and Cook in [3]. In this paper, researchers have used timbre related features, pitch related features and rhythmic content features based on Beat Histogram. They have used Gaussian Mixture Model (GMM) and k Nearest Neighbour (KNN) classifiers for the evaluation of their model. The overall genre classification accuracy of the system produces a 61% of correct classifications over 10 musical genres. Sound analysis process was used for different sound representation techniques such as waveform, spectrum and spectrogram for the different purpose. In [4] Costa has proposed an alternative approach for musical genre classification which is based on texture images. They convert the audio signal into spectrograms and then

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

extract features from this visual representative image. However, larger musical structures other than the instantaneous surface features are difficult to identify by only viewing a spectrogram. Previously most of the music classification researchers worked on a fusion of feature subspaces. Lately some approaches were built on classifier ensemble techniques, where fusion of the genre labels are assigned separately by each single classifier [5], [7].

Music is an inherently multimodal type of data and Mayer in [6] has proposed an approach for multimodal classification of music using classifier ensemble techniques. In [7] researchers have presented a method, which combines the multiple feature vectors extracted from the beginning, middle and end parts of 30 second music segments [7]. In this research we tried to combine short term and long term based features of a music piece, using our novel classification approach for the Sri Lankan local music context focusing on five pre-identified genres.

Texture window was firstly introduced for musical genre classification in [3]. They used variances and means to capture the long term features of sound texture. A novel approach to musical genre classification using temporal information was proposed in [8]. In this paper they have introduced several different temporal evolution descriptors as features. The experimental results show that using only mean and standard deviation achieves the best accuracy and when adding more temporal evolution descriptors, it does not cause any improvement of the overall classification accuracy. They showed that standard deviation and mean are simple but powerful for discriminating different music genres.

McKay in [9] has tried to improve musical genre classification performance using lyrical feature. They investigated the genre classification utility of combining features extracted from symbolic, audio, lyrical and cultural sources of musical information. The experimental results show that features extracted from lyrics were less effective than the other feature types.

B. Background on Sri Lankan Music Genres

The historical trajectory of Sri Lankan music can be traced to its roots in traditional folk music which predates the Indian and European influences on Sri Lankan music. Developed by the lay people the ancient folk music was framed by the Buddhist religious traditions and communal folklore including various rituals which were part of the daily life of the layman. Early music and song took the forms of folk poems (kavi) and improvised folk verse (virindu) and was essentially seen as a communal practice.

The influences of foreign artistic traditions, especially those of Indian origin, on the development of the genre of Sri Lankan music is evident in the early forms of theatrical music. From south Indian classical music (Karnataka music) to North Indian classical and tamil and Hindustani film music, the influence of Indian musical forms continue to shape the Sri Lankan musical genres.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

With the development of traditional theatrical rituals and pageantry, the emergence of cultural forms such as the Kolam, Noorthi and Naadagam traditions served to the expansion of Sri Lankan music. Based on low country folk tradition, Kolam revolved around traditional customs of exorcism and healing and was derived from masked comedy and drama [1], [2]. The south Indian influence on Sri Lankan music is exemplified through the form of Nadagam music, which was introduced by a south idian artists. Phillippu Singho in 1824. the Sinhala theatrical musicals such as “Maname” and “Sanda Kinduru” are derivatives of the south Indian street drama tradition [1].

Noorthy which owes its musical roots to the North Indian Musical tradition was influenced by Parsi theatre. However, one of the turning points in the evolution of Sri Lankan music genre, is the influence of Hindustani classical music.

With the arrival of Ravindranath Tagor in sri lanka in 1934, the foundation for the Hindustani classical music, also known as, ragadari music was cemented with the establishment of “sri Pali” at Horana. the cultural imprint created by the introduction of this musical tradition influenced not only music but also forms of art, theatre and dancing. nevertheless, its impact on the stylistic components of Sri Lankan music was significant as the origins of Sri Lankan classical music can be directly linked to the Hindustani ragadari music [2]. With many Sri Lankan artists visiting india for higher education, even at present, the influence of this musical tradition continues to mould Sri Lankan music. Pioneering the artistic tradition of Sri Lankan classical music, veteran artists such as ananda samarakoon derived musical inspiration from the North Indian classical ragas for their composings.

In the wake of western and Indian proliferation in music, composer and singer Ananda Samarakoon emerged from training at Rabindranath Tagore's school at Shanthiketalan to develop a uniquely Sinhalese music tradition in 1939. His work such as "Punchi Suda", "Ennada Manike" and notably "Namo Namo Maatha" (adapted as Sri Lanka's national anthem later) established the sarala gee genre [2]. Another artist Devar Surya Sena with his Western education was pivotal in popularizing folk songs of Sri Lanka to the English elite that bore higher status in the country at the time [1]. However, with the emerging national consciousness and the subsequent search for national identity in music saw the attempt to break away from the pervasive influence of the indian tradition. The musical genre introduced by sunil shantha exemplifies the development of a truly Sri Lankan musical form.

The arrival of the portugese also resulted in the introduction of the conscripted Africans and the formation of the community of the Sri Lankan Kaffirs[19]. This led to the development of Baila music, a form of folk art, which uses European instruments and rhythms. It entered into Sri Lanka's mainstream culture, primarily through the compositions of work of Wally Bastian who combined “kaffirhina” rhythms with Sinhala lyrics which gave birth to Sri Lankan Baila music. By the 1970s musicians, including MS Fernando and Maxwell Mendis, had helped Baila grow into a well-known and respected style of Sri Lankan popular music [1], [2].

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

The influence of both European and African traditions served to further diversify the musical roots of contemporary Sri Lankan music. Calypso-style is a genre of Sri Lankan music. It grew out of Sri Lankan musicians' fascination with the music of the Caribbean in the 1960s, particularly Harry Belafonte and calypso music. It typically uses acoustic guitars, rhumba shakers and conga/bongo drums.

Sri Lankan groups such as Los Cabelleros led by Neville Fernando (first ever Sinhala pop group), Las Bambas, The Humming Birds, Los Muchachos, and The Moonstones(whose members included Annesley Malewana and Clarence Wijewardane) practiced this music, which melded Caribbean rhythms to traditional Sri Lankan music. Noel Ranasinghe's Le Ceylonians became the most famous group of this genre [1], [2].

The above account of Sri Lankan music clearly identifies 5 popular modern genres in Sri Lankan music. They are namely; Ragadari, Classical, Western, Baila and Calypso music. We hope to use few or all of these genres for our genre classification based on the availability of good quality local songs with respect to each class but more importantly based on the expert opinion of musicians.

METHODOLOGY

Music structure analysis systems use symbolic and audio representation for collecting musical information. An audio format is a format that is used for storing digital audio data. This data can then be stored in few different formats, namely uncompressed, lossless compressed and lossy compressed. These three audio formats are used to reduce the file.

WAV, AIFF, and AU are uncompressed audio formats and the WAV format is based on the Resource Interchange File Format (RIFF). A lossless compressed format such as FLAC format stores data in less space by eliminating unnecessary data. Lossy compressions such as MP3, AAC formats enables even greater reductions in the size by removing some of the data.

Lossy compression typically achieves far greater compression but it reduced quality than lossless compression by simplifying the complexities of the data. Most of the music available in the Internet and personal collections is stored in digital as WAV or MP3 audio formats.

A. Design

We witnessed a clear difference between the different genres of songs on their visual representation of music audio signals. This was clear evidence to understand that these genres are in fact different and hence the importance of identification of the low-level features which differentiates these genres is highlighted.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

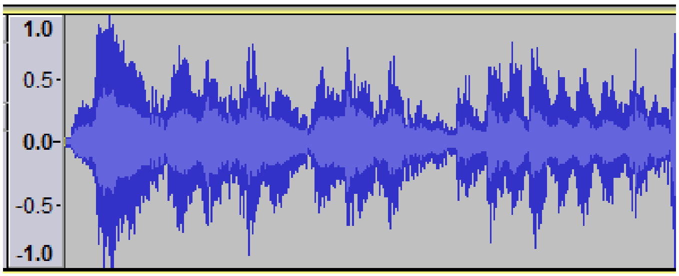

Figure 1. Visual representation of the raw audio file for 'Baila' genre

The above Figure 1 depicts the visual representation of an audio file belonging to 'Baila' genre in .WAV format. The pattern in which peaks appear in songs of this genre are quite different from the rest of the genres. Also another noticeable thing from this visual representation is that the shape of the signal is different to that of others. The objective of this research is to identify the low-level musical features which best recognizes these characteristics unique to these genres.

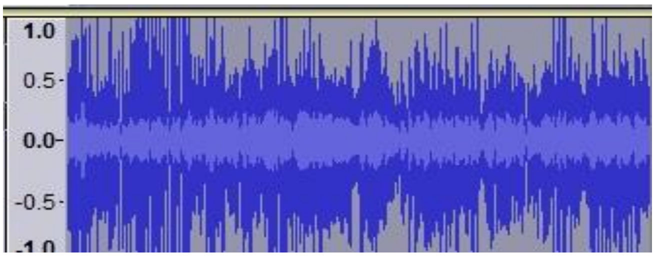

Figure 2. Visual representation of the raw audio file for 'Ragadari' genre

The above Figure 2 depicts the visual representation of an audio file belonging to the 'Ragadari' genre. The Ragadari genre is a musical genre which originated in India and these songs are very different from the other four types of songs we discuss in this research paper. The Ragadari singing style has a lot of sudden pitch variations which is also reflected by the high zero crossings shown in Fig 2. Therefore when you listen to a song in this genre, it feels like it's filled with lot of noise, the reason simply being a characteristic of the singing style.

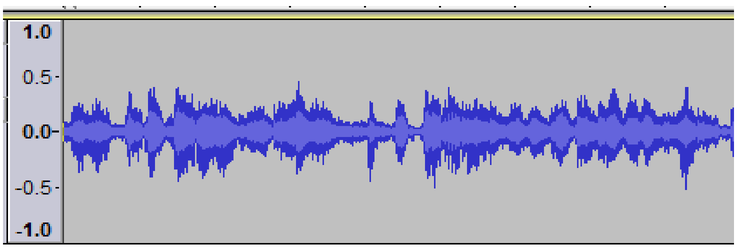

Figure 3. Visual representation of the raw audio file for 'Calypso' genre

The above Figure 3 depicts the visual representation of an audio file belonging to Calypso genre. The above figure has the most unique visual characteristics of all the 5 genres studied in this research. The songs of this genre according to its visual characteristics remain within a certain low frequency range which makes it very easy to distinguish this type of songs from the rest. Another thing that is noticeable is that its shape is also consistent with less variations.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

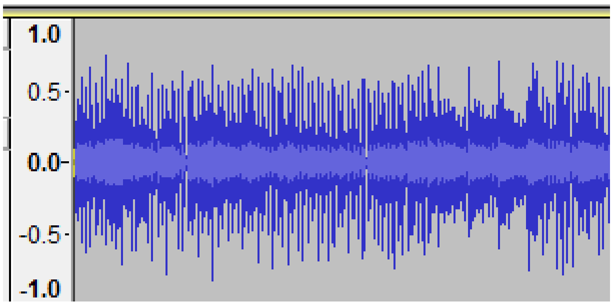

Figure 4. Visual representation of the raw audio file for 'Western' genre

The above Figure 4 depicts the visual representation of an audio file belonging to 'Western' genre in .WAV format. The shape of this song is also different to that of others. There are very thin spikes appearing all along the song which is only slightly noticeable again in Classical type of songs. Another thing easily recognizable from this visual representation is that the shape of the signal is consistent throughout the duration of the signal. It is important to experiment these classes against one another to test how effectively a classifier can recognize these genres when they're tested against each other.

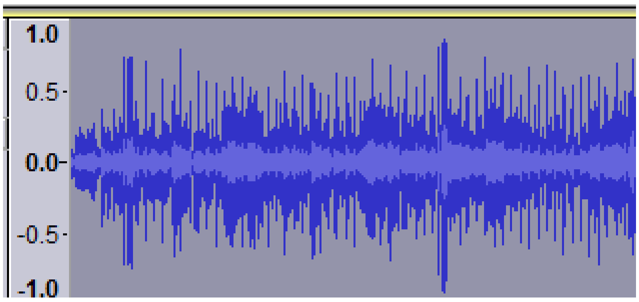

Figure 5. Visual representation of the raw audio file for 'Classical' genre

The above Figure 5 depicts the visual representation of an audio file belonging to 'Classical' genre in .WAV format. The shape of this signal is a bit similar to that of the signal of the Western genre. At the same time there are sudden peaks with big impact appearing at different places along the length of the song, which distinguish them from the earlier Classical genre. Also it would be interesting to see the energy levels in each of these genres as it will be a key differentiator for identification between these genres.

The main objective of this research is to identify the low-level musical features which best represents these 5 musical genres. By analyzing the visual representations of these audio files belonging to different genres, it can be said that these genres show more differences to each other than similarities.

Therefore the next section will describe how the dataset was created in order to carry out the experiments to achieve the research objectives.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

B. Dataset

Dataset of a study plays a vital role to its end result. The appropriateness of the data and the nature of the data gathered is key when thinking of the dataset of a study. Sometimes data cannot be processed as it is, hence preprocessing is used as a technique to make sure that all data used is of the same standard, same level so as to easily compare them. This is true for any study regardless of the field or the scope of that study. A good dataset will result in a good system, hence it has a higher probability of producing results with high accuracies. There are some general issues to be considered about and addressed to setup a good dataset. In a classification system, in order to have a good classification outcome, sufficient amount of music files for each category has to be found and the classifier needs to be trained using them. The larger the training dataset, higher the accuracy of the classification it produces. On the other hand, the training set has to be labelled and it should be the ground truth of music structure discovery. Since, the structure is a subjective factor, a universal ground-truth for music structure does not exist and getting reliable labels for the data is often a serious practical problem that researchers have to consider. Determining the number of different structural categories for the training dataset is another important design consideration. Usually, accuracy is decreased when the number structural categories are increased. Hence it is recommended to carry out the study with few structural categories and choosing data wisely would increase the accuracy of the results. This study focuses on Sri Lankan music, hence Sinhalese songs from the golden era is best suited for this type of study. Since there is no research work has been conducted for music classification based on structural similarity for Sri Lankan music before, no data sets are available. No matter what the field of the study in Sri Lankan music, no data sets are available. Therefore, a data set for this study has to be generated from the very beginning. For Western music, generating a new data set would not be that cumbersome since user tags for music files are publicly available on the web. But, we are unlucky to not have such user tags available for Sri Lankan songs. In this case, we have to take the assistance of some resource persons like experts in the music field. Or else, we can do a subjective test just like Yi-Hsuan Yang did for their studies in [9, 10]. In those studies, authors have got the help of interested people in order to generate a data set by labeling the songs for different structures represented in songs. Sri Lankan music today is not in a good standard. As we all know, some of the newer generation people killing the spirit of the good music. As a result the current situation of Sri Lankan music is pathetic. As described in the introduction chapter too, all the music styles and categories have been fused by Sri Lankan people and hence, there are no standards visible, like there are in western music. Because of this reason, it is hard to find out a good collection of songs in order to generate a new dataset. Therefore, as some musicians pointed out, we have to generate the dataset using the songs from 60’s, 70’s and 80’s. According to the music experts in Sri Lanka, Sri Lankan music was at a very high standard in terms of maintaining consistent structures with respect to different song categories. Different level of standards hence called the “Golden Era of Sri Lankan music”. According to them, the melodies of the songs of that era is best suited when studying/analyzing the structure of music and classifying them based on structural similarity. Therefore the dataset will be generated using the songs from 60’s, 70’s and 80’s. In order to decide

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

on the number of genres for this classification we consulted the expert opinion of musicians in Sri Lanka. As the literature also identifies, the expert opinion of musicians was to limit the research to 5 genre classes namely Baila, Classical, Ragadari, Western and Calypso. Therefore the dataset used for this research consists 108 songs belonging to the 5 genres Baila, Classical, Ragadari, Western and Calypso with the class Baila having 20 songs and rest of the classes having 22 songs each.

C. Domain based audio features

Whatever the format music is stored, the data it contains can be decoded and transformed into a succession of digital samples to represent the waveform. But this data cannot be used directly by automatic systems because pattern matching algorithms cannot deal with such an amount of information and formats. So it is necessary to extract some features that describe the audio wave using a compact representation. For applications of audio signal processing such as music classification, feature extraction is one of the most important steps. In this section, we review four domain of audio features and explain how they are extracted. These features can be roughly classified into short term and long term features. After pre-processing the dataset, the set of musical features representing short time features of music was extracted from the audio using the Marsyas3 (Music Analysis, Retrieval and Synthesis for Audio Signals) framework. Marsyas is an open source software framework for audio processing with specific emphasis on MIR applications. Then these musical audio features can be grouped together into feature vectors that serve as the input to classification systems. Computational features are extracted from digital audio signals but they do not have a musical meaning to any human perceptual measure.

1) Chroma *:* Human perception of pitch is periodic in the sense that two pitches are perceived as similar in color but di_er by an octave. Based on octave a pitch can be separated into two components which are referred to as chroma and tone height. Chroma features show a high degree of robustness to correlate closely in the musical aspect of harmony and variations in timbre.

Chromagram: The chromagram is a spectrogram that represents the spectral energy of each of the twelve pitch classes by maps all frequencies into one octave. Chromas set consists of the twelve pitch spelling attributes: A, A♯, B, C, C ♯ D, D ♯, E, F, F♯ G, G♯ as used in Western music notation. Chroma is a pitch based feature that projects the frequency spectrum into 12 bins, with one bin for each of the 12 distinct pitches of the chromatic musical scale. The conversion of an audio music into a chromagram representation can be performed by using STFT in combination with binning strategies [11].

2) Rhythm features: Rhythm is the timing pattern of musical sounds and silences. These musical sound and silences are put together to form a pattern of regular or irregular pulses caused in music by the occurrence of weak and strong melodic and harmonic beats to create a rhythm. The above block diagram illustrates the different steps in the calculation from raw audio signal to the final Rhythm Patterns, Statistical

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Spectrum Descriptor and Rhythm Histogram features. These three type of audio features are extracted using the java audio feature extraction packages.

3) Timbral Features: These features extracted from an audio signal is used to represent timbral texture which are based on standard features proposed for music-speech discrimination [16] and speech recognition [17]. The following features extracted from music audio signals for this research falls into the category of timbral features.

Time Domain Zero Crossings: The following equation depicts the function for calculating the zero crossings of a music audio signal.

$$Z_t = \tfrac{1}{2} \sum_{n=1}^{N} |\mathrm{sign}(x[n]) - \mathrm{sign}(x[n-1])| \qquad (1)$$

The sign function produce 1 for positive arguments and 0 for negative arguments and x[n] is the time domain signal for the frame of duration t. This feature is a good indication for the noisiness of a signal.

Mel-Frequency Cepstral Coefficients: Cepstral features are frequency smoothed representations of the logarithm of the estimated spectrum of a signal and capture pitch and timbre characteristics. Mel-frequency Cepstral Coefficients (MFCCs) are compact, short time descriptors of the spectral envelope audio feature set and typically computed for audio segments of 10-100ms [12]. MFCC are one of the most popular set of features used in pattern recognition. MFCC was originally developed for automatic speech recognition systems, lately have been used with success in various musical information retrieval tasks[13], [15]. Although this feature set is based on human perception analysis but after calculated features it may not be understood as human perception of rhythm, pitch, etc. Normally first 13 MFCCs are used for musical information retrieval tasks. This illustrates the different steps in the calculation from raw audio signal to the final MFCC features. Normally violin's sounds has much higher values in the third and fifth MFCC than the flute and the fork so mel-frequency information may be better suited to discriminate between the different sound sources or different instruments.

The first step is dividing the speech signal into frames, usually by applying a windowing function at fixed intervals. The aim here is to model a small typically 20ms section of the signal, which are statistically stationary. The window function typically a hamming window and its removes edge effects.

Then for each short analysis window a spectrum is obtained using FFT. In the next stage the Spectrum is passed through Mel-Filters to obtain Mel-Spectrum. This mel-band step is also a smoothing of the spectrum and a Cepstral analysis is performed on Mel-Spectrum to obtain MFCC. In the Cepstral analysis stage, take the logs of the value of the mel bands and apply a set of discrete cosine transform (DCT) filters on the mel bands as if they were signals [13].

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Finally the result is a lower dimensional feature of MFCCs. Thus music is represented as a sequence of Cepstral vectors and which are given to pattern classifiers for musical genre recognition purpose.

RESULTS AND DISCUSSION

A. Experiment Strategies

This research was evaluated using two main experiment strategies: One vs One and One vs All. Also known as 'OVR or One Vs Rest', One vs All is a very common experiment strategy used in multiclass classification problems. In this strategy we compare the detailed accuracies by class to examine how well each class differentiates each other [14]. This experiment method is used to compare and contrast the overall accuracy of the classification model by their individual class accuracies. It can be stated that the overall accuracy has dropped with the increasing number of output classes. This kind of experiment is very important in deciding the best classifier model for classification of Sri Lankan songs. However 'One vs One' (OVO) approach treats the classification problem as a set of multiple binary classification tasks. Every time two output classes are selected and their accuracy by class statistics which are produced by WEKA is used to make a critical analysis on the decision boundary of individual classes. This technique allowed us to recognize the classes which can be distinguished quite easily which means there is a healthy class boundary/ decision region between these two classes. But most importantly it gave us the opportunity to recognize the classes which are harder to distinguish and which have more overlapping instances between them. This is a clear indication that there is no healthy decision boundary between such classes. OVO experiment strategy has often proved to be useful when identifying the fact that the reason for the overall accuracy of a multiclass genre classification model to drop is because of the presence of overlapping instances in only two genre classes. Once these kind of information is discovered then we can opt for a hierarchical classification model, sometimes known as multi-level classification which will increase the overall accuracy of the classification model. Therefore the identification of key relationships that exist between the output classes is important to implement the best model for a given classification problem. As it is 'genre classification' in this instance, the probability of overlapping occurring between output classes is much higher. Therefore we have used OVO as one of our experiment strategies in this research.

B. Feature Selection

Generally, all features in our feature vector can be characterized as relevant, irrelevant or redundant features. Relevant features have an influence on the output result and their role cannot be assumed by the rest. Irrelevant features not having any influence on the output and which values are randomly generated for each example.

Redundant features can take the role of another. For example, if values of two features are completely correlated then they are redundant to each other. In machine learning,

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Feature Selection (FS) also known as attribute selection is the technique of selecting a subset of relevant features for building robust learning models by removing most redundant and irrelevant features from the feature vector. The main goal of FS is to determine minimal feature subset without affecting the high accuracy in representing the original features. FS method is extremely useful in reducing the dimensionality of the feature vector to be processed by the SVM classifier, improving predictive accuracy by removing irrelevant features or noise data, and speeding up the running time of the learning algorithms

C. *Classification*

Each experiment under 'One vs One' classification was experimented with 3 different learning algorithms as the base classifier; SVM with PCA (Principal Component Analysis), MLP (Multilayer Perceptron) and J48 decision tree algorithm for critical analysis between the learning algorithms.

TABLE 1. Summary of all accuracies - One vs One (Using Chroma Features)

	Classical vs Western	Classical vs Calypso	Classical vs Baila	Calypso vs Western	Baila vs Calypso	Western vs Baila
1	0.838	**0.543**	0.738	0.781	**0.541**	0.736
2	0.709	0.778	0.724	0.936	0.94	0.829

Table I depicts the summary of the performance of one vs one experiments with respect to each genre class. The performance of these experiments were measured by the classification accuracy of each experiment. By looking at the table below, it can be understood that all classes display good overall accuracies, hence it indicates there is a clear decision boundary between these two classes and very less number of misclassifications. The same experiment conducted using Rhythm features and Timbral features turned out with much higher overall accuracies. Hence that was identified as the best classification model from the two feature vectors. In order to get a better understanding of its performance several experimental results of both winning classifier and the other is shown below.

TABLE II. Summary of all recall values - One vs One (Using Chroma Features)

Classical vs Western	Classical vs Calypso	Classical vs Baila	Calypso vs Western	Calypso vs Baila	Western vs Baila
70.1%	**95.5%**	**86.4%**	**95.7%**	**86.4%**	**78.3%**

In the above table II, the individual recall values obtained for each class using only Chroma features also known as feature set 1 is compared and shown. In this table, Row '1' displays the recall value of the first mentioned class while Row '2' displays the recall values of the second mentioned class.

Interestingly, while all other results are consistent two values show deviations from the rest. Class Classical has a recall value of 0.543 when tested against Class Calypso,

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Vol.4, No.5, pp.70-88, September 2016
Published by European Centre for Research Training and Development UK(www.eajournals.org)

while Class Baila shows a recall value of 0.541 when tested against Class Calypso. This result shows that the above classifier model is unable to distinguish the two classes correctly on a number of occasions. This is because the class boundary is not very clear between the classes Calypso and Baila and Calypso and Classical. These results could occur due to the fact that the extracted features are not enough to differentiate the two classes. Hence the same experiment was carried out by using the second feature set, which was built using a combination of features from Timbral and Rhythm domains.

TABLE III. Summary of all accuracies - One vs One (Using a combined feature vector of Rhythm & Timbral features)

True Positive Rate	False Positive Rate	Precision	Recall	Receiver Operating Characteristics (ROC) Area	Class
0.3	0.182	0.6	**0.3**	0.559	Ragadari
0.818	0.7	0.563	0.818	0.559	Classical

The above table III depicts the new results of the experiment in table 1 using a combined feature vector using features from two different domains; Rhythm and Timbral. These results prove that a combined approach is better in performance than extracting features from a single domain.

The following table IV shows the new recall values of the classes which showed confusions in the experiment shown in table II. The results clearly indicate use of a combined feature vector shows definite improvement in identifying class boundaries over the earlier approach.

	Classical vs Calypso	Baila vs Calypso
1	**1**	**0.818**
2	0.909	0.909

TABLE IV. Summary of all recall values - One vs One (Using Chroma Features)

According to the above table the new approach has drastically improved the performance of the classifier. Both recall values which were little above 0.5, now have values of 1, 0.818 respectively. This improvement is further evident when we analyze the final confusion matrix of the multiclass classification conducted using this approach. Table V depicts the results of the multiclass classification using the combined domain feature approach.

Another significant finding of this research was that we were able to provide a scientific basis to the long standing expert opinion within Sri Lankan Music that the Sri Lankan classical music has its roots in the Hindustani ragadari music.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Classical vs Western	Classical vs Calypso	Classical vs Baila	Calypso vs Western	Calypso vs Baila	Western vs Baila
77.9%	71.46%	73.1%	87.2%	78.7%	79.4%

TABLE V. Summary of all statistics values – Ragadari vs Classical (Using a combined feature vector of Rhythm & Timbral features)

It can be noted that new classifier model shows a deviation from its usual good performance as it gives a 0.3 recall value for the class Ragadari. This result has occurred because a large portion of Ragadari instances has been misclassified as Classical music. Table VI depicts the results of that confusion matrix.

	Ragadari	Classical
Ragadari	6	14
Classical	4	18

TABLE VI. Confusion Matrix – Ragadari vs Classical (Using a combined feature vector of Timbral, Rhythm features)

The above table VI depicts the resulted confusion matrix for the classification between Classical and Ragadari genre classes. The attention should be focused to the value highlighted in bold as it leads to a significant scientific discovery. By looking at the confusion matrix we can see that the classifier has confused and mapped instances of Ragadari music to instances of Sri Lankan Classical music. Because this is a binary classification only instances present are those belonging to these two genre classes. Hence if there is similarity between them there is a higher probability of confusions occurring. Although normally this kind of misclassification between any other genre classes can be considered as poor performance, this misclassification proves otherwise.

In the literature review in section 2, it was mentioned that the Sri Lankan Classical music is said to have its roots in Ragadari music**.** This assumption is also accepted by the musicians in Sri Lanka and they only justify it through their expertise on the domain. So up until now, there was no scientific basis to this assumption apart from the expert opinion of musicians. The result above shows that the classifier has identified majority of the Ragadari music as Sri Lankan Classical music. This can only happen because the classifier identifies the similarity between these two genre classes. Therefore this result provides a scientific basis for the assumption that "the Sri Lankan music is said to have its roots in Ragadari music".
This is a significant discovery of this research, and this finding alone adds a lot of value for this research

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

European Journal of Computer Science and Information Technology
Vol.4, No.5, pp.70-88, September 2016
Published by European Centre for Research Training and Development UK(www.eajournals.org)

This approach proved to be better in terms of the multiclass classification as well. By using all 5 genres at the same time we created this combined feature vector by extracting features from the above mentioned two feature domains. When the dataset was tested with the improved classification model the following results were obtained. These results depicted in below table VII are quite remarkable compared to the results obtained by extracting Chroma features only.

	Baila	Ragadari	Classical	Western	Calypso
Baila	**18**	0	2	0	2
Ragadari	0	**16**	6	2	0
Classical	0	4	**18**	0	2
Western	4	0	0	**18**	2
Calypso	0	0	2	0	**21**

TABLE VII. Final Confusion Matrix – Multiclass classification (Using a combined feature vector of Timbral, Rhythm features)

By looking at the values along the diagonal of Table VII, we can clearly see that a vast majority of inputs have been correctly classified for all genre classes. There is hardly any confusion visible with regard to all the genres. This is a clear indication that this approach has captured the characteristics belonging to all genres of Sri Lankan Music. Hence this becomes an acceptable genre classification model.

By focusing our attention to the class Ragadari you can see that this time most of the instances have been correctly in this experiment. As 6 is the highest number for misclassifications, here also we can notice evidence of their similarity as 6 instances of Ragadari has been misclassified as Classical. The reason for the reduced confusions is because it is a multiclass classification there is less chance of confusions occurring between these two genres as it is mixed with other data. Earlier the similarity was recognized strongly when only those two classes were present.

This concludes the evaluations carried out using the two feature sets; Chroma feature set and the combined feature set using features from Rhythm and Timbral domains. All features were extracted from the same dataset and same experiments were carried out for both feature sets to conduct a fair comparison between their results. Although both feature sets were able to justify the existence of different genres in Sri Lankan music, combined feature set with multiple domain features also known as feature set 2 showed better performance in terms of the multiclass genre classification model. Thus the classification model built combining features from multiple feature domains can be identified as the better classification model for this research.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

D. Bagging & Boosting Effect on the winning classifier

It was identified that the classification model built combining features from multiple feature domains was the winning classifier for the Sri Lankan Music context. The final experiment conducted was to check the effect of bagging and boosting techniques on this classifier.

Boosting and bagging are known techniques for constructing ensemble classifiers. The effect of boosting and bagging techniques were also tested on the winning classifier which was built using feature set 2. In general boosting and bagging techniques are applied on unstable classifiers when performance needs to be further improved. But in order to make this a complete research, both boosting and bagging techniques were applied on the winning model.

TABLE VIII. Effect of boosting & bagging on winning classifier model

	Result Before Applying Boosting or Bagging	Result After Applying Bagging	Result After Applying Boosting
Correctly Classified Instances	**74.5 %**	**45.1 %**	**46.09%**
Incorrectly Classified Instances	25.5 %	54.9%	53.91%
Kappa statistic	0.68	0.31	0.36
Mean absolute error	0.10	0.24	0.22
Root mean squared error	0.32	0.38	0.43
Relative absolute error	31.8%	75.24%	69.0%
Root relative squared error	79.6%	93.94%	106.76%

By looking at table VIII, it can be easily noted that the performance has decreased having applied boosting and bagging techniques on our classifier.

The resultant confusion matrix having applied boosting technique to the classifier is given on table IX.

TABLE IX. Confusion Matrix - After applying boosting to the classifier

	Baila	Ragadari	Classical	Western	Calypso
Baila	**10**	2	4	4	0
Ragadari	2	**11**	7	2	0
Classical	0	4	**10**	4	4
Western	0	6	4	**8**	4
Calypso	0	0	6	4	**12**

Although a majority of the instances have been correctly identified, there is a considerable amount of confusions visible in the above table.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

Table X depicts the confusion matrix of the classification after applying bagging technique to the classifier.

TABLE X. Confusion Matrix - After applying boosting to the classifier

	Baila	Ragadari	Classical	Western	Calypso
Baila	**16**	0	2	2	0
Ragadari	3	**10**	7	2	0
Classical	0	**6**	**4**	4	**6**
Western	4	**8**	4	**0**	4
Calypso	0	0	0	6	**16**

Values highlighted in bold which are not along the diagonal of the above table shows that data is behaving unpredictably and this is usually expected if these techniques are tried on a stable classifier model. Boosting algorithms perform better when the classifiers are weak and the data does not have much noise. Using a classifier like SVM with radial basis kernel function that is already strong as the base learning algorithm in boosting or bagging does not seem to provide any advantages. Therefore the performance decrease of the classier can be justified.

CONCLUSION & FUTURE WORK

In this paper we addressed the problem of music genre classification for Sri Lankan music using audio signals generated from music audio tracks. Most researchers in this research area are very concerned with the classification accuracy of the used model. Most of the times it is what determines the success of a particular genre classification model. The winning classification model of our research achieved a highest overall classification accuracy of 74.5% which is just short of 75% which provides significant motivation for future researchers of this research area especially within our local context as this research is the first of its kind to address this problem for Sri Lankan music.

Our main objective of this research was to scientifically justify the existence of different genres in Sri Lankan music. Although this was accepted by musicians, there was no scientific basis for this statement except for their expert opinion. The results from this research was able to scientifically justify the existence of different genres based on their musical features which is a significant milestone of this research. Another of our objectives was to identify different domain features which are extractable from a music audio signal that could correctly recognize the characteristics belonging to our local music. We used two different types of domain based feature sets to achieve this purpose. We proposed two types of feature vectors based on chroma features and timbre, cepstral and rhythm domain features. Thus we were able

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

to utilize multiple feature domains which are relevant to music audio retrieval for classification of our local music.

Furthermore we tried both boosting and bagging techniques on the winning classifier to see the effect these techniques will have. The final objective of this research was to come up with a genre classification model with a reasonable accuracy and recall. For this task we identified 5 genre classes based on the expert opinion of several musicians in Sri Lanka. They are namely Baila, Classical, Western, Ragadari and Calypso. We came up with two classification models using the two different feature vectors we extracted. Both classification models were evaluated according to these genres by carrying out the same set of experiments for each feature set individually based on two key experiment strategies: binary & multiclass. The success of the classification model was evaluated by considering both overall classification accuracy as well as the recall values produced by the classifier for each genre. Along the course of this research, we were also able to make a significant discovery with related to Sri Lankan music as we managed to provide a scientific basis for the expert opinion of Sri Lankan musicians that “Sri Lankan classical music has its roots in Ragadari Music”.

As future work we intend to develop new features which are able to extract more useful information from the audio signals and use more feature sets such as Temporal Statistical Spectrum Descriptor, Temporal Rhythm Histograms, etc. Also we intend to improve the classifier with respect to its overall classification accuracy.

REFERENCES

[1] “http://www.mysrilanka.com/travel/lanka/art/music.htm
[Online].Available: http://www.mysrilanka.com/travel/lanka/art/music.htm.
[Accesssed 12 2015]
2] “Sri Lankan Music, “[Online]. Available :
https://en.wikipedia.org/wiki / Music_of_Sri_Lanka_Sinhala_light_music
[Accessed 12 2015]
[3] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” IEEE Trans. on speech and audio process., vol. 10, no. 5, pp. 293–302, 2002
[4] Y. Costa, L. Oliveira, A. Koericb, and F. Gouyon, “Music genre recognition using spectrograms,” in 18th Int. Conf. on Systems, Signals and Image Process. (IWSSIP), 2011, pp. 151–154.
[5] R. Mayer, A. Rauber, P. Ponce de Leon, C. P ´ erez-Sancho, and ´ J. Inesta, “Feature selection in a cartesian ensemble of feature ˜ subspace classifiers for music categorisation,” in Proc. of 3rd int. workshop on Music and Machine learning (MML), Firenze, Italy, 2010, pp. 53–56
[6] R. Mayer and A. Rauber, “Music genre classification by ensembles of audio and lyrics features,” in 12th Int. Conf. on Music Information Retrieval (ISMIR), Miami, Florida, USA, 2011, pp. 675–680.
[7] C. Silla, C. Kaestner, and A. Koerich, “Automatic music genre classification using ensemble of classifiers,” in Proc. of the IEEE Int. Conf. on Systems, Man and Cybernetics (SMC 2007), Montreal, Canada, 2007, pp. 1687–1692.

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)

[8] R. Tao, Z. Li, Y. Ji, and E. Bakker, "Music genre classification using temporal information and support vector machine," in ASCI Conf., vol. 77, 2010.
[9] C. McKay, J. Burgoyne, J. Hockman, J. Smith, G. Vigliensoni, and I. Fujinaga, "Evaluating the genre classification performance of lyrical features relative to audio, symbolic and cultural features," in 11th Int. Conf. on Music Information Retrieval (ISMIR), vol. 10, Utrecht, Netherlands, 2010, pp. 213–218.
[10] E. Pampalk, A. Rauber, and D. Merkl, "Content-based organization and visualization of music archives," in Proc. of the 10th ACM Int. Conf. on Multimedia, Juan-les-Pins, France, 2002, pp. 570–579.
[11] T. Lidy and A. Rauber, "Evaluation of feature extractors and psycho-acoustic transformations for music genre classification," in 6th Int. Conf. on Music Information Retrieval (ISMIR), London, U.K, 2005, pp. 34–41.
12] C. Snoek, M. Worring, and A. Smeulders, "Early versus late fusion in semantic video analysis," in Proc. of the 13th Annu.ACM Int. Conf. on Multimedia, Nov. 2005, pp. 399–402.
[13] G. Peeters, A. Burthe and X. Rodet, "Toward Automatic Music Audio Summary Generation from Signal Analysis," in International Conference on Music Information Retrieval, October 2002
[14] J. J. Aucouturier and M. Sandler, "Segmentation of Musical Signals using Hidden Markov Models," in Proceedings of AES 110th Convention, May 2001
[15] C. Roads, The Computer Music Tutorial, Cambridge: MIT Press, 1994.
[16] S. Davis and P. Mermelstein, "Experiments in syllable-based recognition of continuous speech," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 28, pp. 357–366, Aug. 1980.
[17]. Scheirer and M. Slaney, "Construction and evaluation of a robust multi-feature speech/music discriminator," in Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1997, pp. 1331–1334.
[18] C.McKay, Automatic Music Classification with jMIR, Montreal, 2010

[19]"https://en.wikipedia.org/wiki/Sri_Lanka_Kaffirs[Online]
.Available: http://www.mysrilanka.com/travel/lanka/art/music.htm. [Accesssed 12 2015]

ISSN 2054-0957 (Print), ISSN 2054-0965 (Online)