Automatic Classification of Music Genre Using SVM

Nandkishor Narkhede, Sumit Mathur, and Anand Bhaskar

Abstract The growing number of music content online has opened up new possibilities for the introduction of successful digital knowledge access services known as music referral systems that help user groups in searching, finding, sharing, and creating. The music recovery approach based on specific similarity information combines several similarity features, including audio and contextual similarities, such as tone format features and melodic details. Audio classification is very important for recovering audio files quickly. To get the best results from audio classification, it is important to choose the best feature set and follow the best analysis method. Support vector machines (SVMs) are implemented by learning from input samples to classify music into separate classes of music genres. The SVM study excelled in the music category classification.

Keywords SVM · ZCR · STE · LPC · RMS · MFCC · MIR and genre

1 Introduction

Information retrieval (IR) requires very little discipline and music information retrieval (MIR) requires different approaches than other field subjects. Prior to the development of the Internet, musical compositions for libraries were arranged alphabetically and were technologically advanced. Around the world of digital music, numerous studies are being conducted and how the user experience can be improved. Many unlabeled music files can be downloaded, cached, and contain incorrect or suspicious tags [1]. Automated classification of genres, however, is not an easy task to do as music develops in a short period of time [1, 2]. In addition, developments in audio and video signal processing and data exploration have resulted in a comprehensive study of music signal analysis, such as content-based music retrieval, music genre classification, duet analysis, music interpretation, and music information retrieval and music instrument identification and classification [1]. Identification

N. Narkhede (✉) · S. Mathur · A. Bhaskar
Sir Padampat Singhania University, Udaipur, Rajasthan, India
e-mail: narkhede.nandkishor@spsu.ac.in

S. Smys et al. (eds.), *Computer Networks and Inventive Communication Technologies*,
Lecture Notes on Data Engineering and Communications Technologies 75,
https://doi.org/10.1007/978-981-16-3728-5_33

techniques for musical instruments include many applications such as detecting and analyzing solo lines, retrieving audio and video, music dictation, playlist creation, group of sound background, analyzing video scenes and tagging [1].

Advanced music libraries are gaining a reputation for being professional archives and private music collections. The number of people interested in audio libraries is also increasing due to improvements in Internet access and network bandwidth [1]. But warehouses are backbreaking with a large music archive and this is time consuming, especially when classifying audio style by hand. Music is divided into genres and subspecies, based not only on sound but also on lyrics [1, 3]. This interferes with the classification. To further complicate matters, the concept of music style may change over time [4]. For example, rock songs done 5 decades ago are very different now.

2 Related Work

This section describes the work done by researchers in the similar area. Xu et al. [5] applied a multi-layer classifier based on support vector machine to classify the music genre and achieved an accuracy of 93.14%. Mutiara et al. [6] used several kernels of nonlinear support vector machines (SVM) for classification of music genres extracting feature sets related to timbre, rhythm, tonality, and LPC from music files and achieved accuracy of 76.6%. In [7], authors used polynomial SVM to classify music using MFCC features and polynomial SVM classifier to achieve an accuracy of 78%. Aryafar et al. [8] performed automatic music genre classification using spasity-eager SVM and obtained accuracy of just 37%. Kyaw and Renu [9] used multi-layer SVM for music genre classification and obtained the accuracy of 93%.

3 Proposed System

The outline of the proposed system is shown in Fig. 1. To build a dataset of input audio signals, we considered the GTZAN dataset from the Marsyas site, which contains 1000 music signals in ten different categories [10]. All audio tracks in the GTZAN dataset are.au format, 16-bit, 30 s long, 22,050 Hz mono file. Figure 2 represents the signal given as input. These audio music signals are filtered using average or mean filters. As shown in Fig. 3, this process results in the amplitude normalization and Gaussian noise elimination in the audio signal.

The segmentation process divides the audio signal into voice and silent frames. For partitioning, we used ZCR and STE as time domain properties as shown in Figs. 4 and 5, respectively, and frequency domain properties, spectral flux as depicted in Fig. 6 and spectral skewness [11]. Below are all the features used in the segmentation

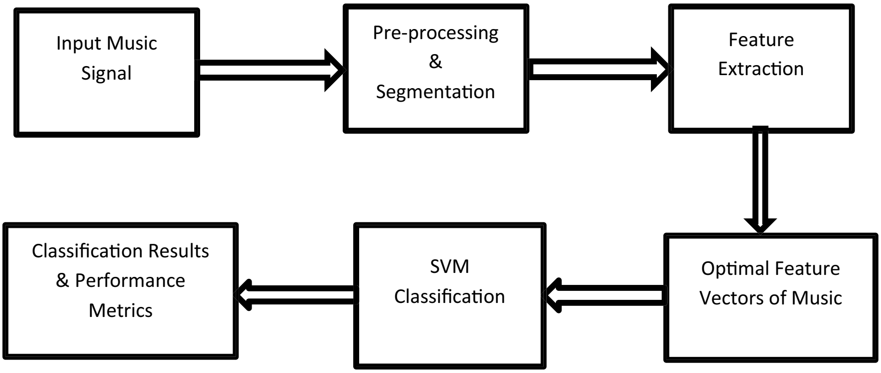

Input Music Signal
Pre-processing & Segmentation
Feature Extraction
Optimal Feature Vectors of Music
SVM Classification
Classification Results & Performance Metrics

Fig. 1 System schema

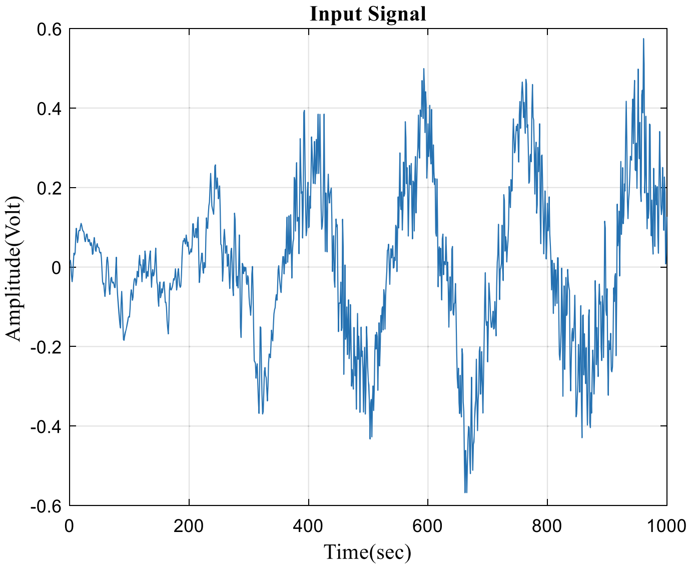


Fig. 2 Input signal

process to find the voiced segment in the input signal. The voiced segment is shown in Fig. 7.

Short-time energy [12]: Representation of Amplitude Differences. It is calculated using

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \tag{1}$$

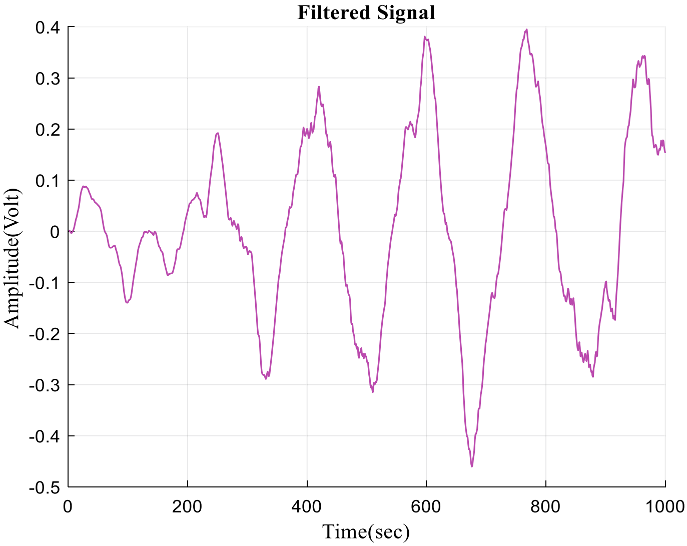


Fig. 3 Filtered signal

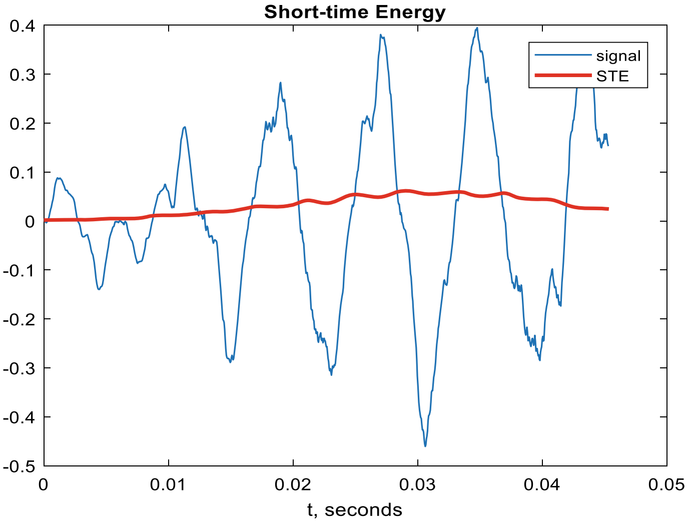


Fig. 4 Short-time energy

Zero crossing rate [12]: The ZCR in the signal reflects the mark change rate. It uses rectangular window function for measurement. Measured using this formula:

$$Z_n = \sum_{m=-\infty}^{\infty} |\text{sgn}[y(m)] - \text{sgn}[y(m-1)]| w(n-m) \qquad (2)$$

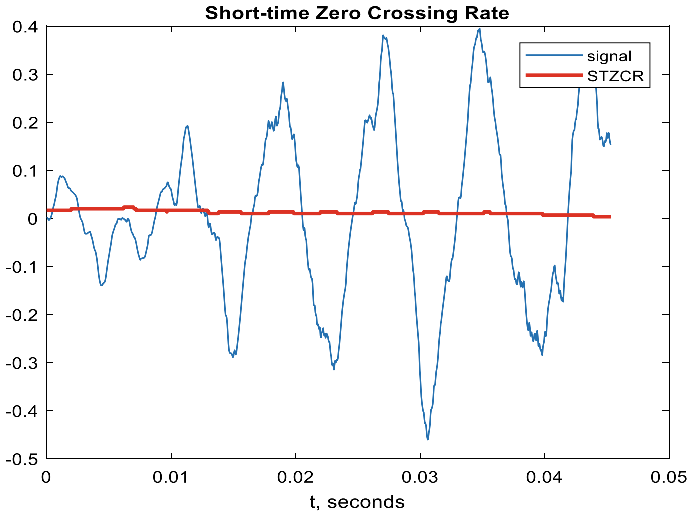


Fig. 5 Short-time ZCR

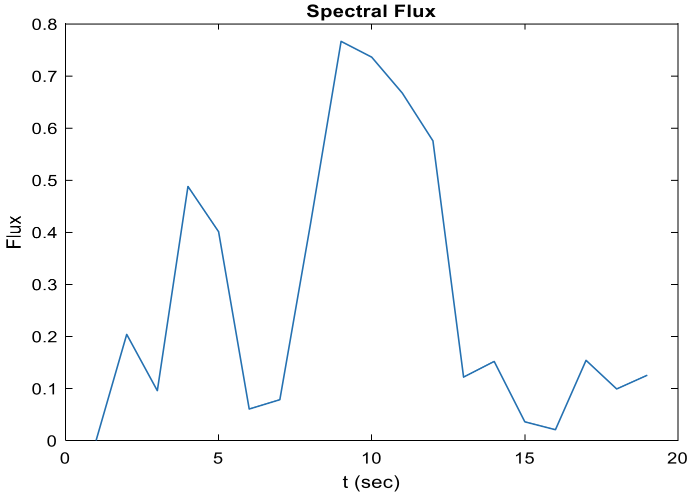


Fig. 6 Spectral flux

Spectral skewness [12]: The pitch portrayed in the music signal is skewness. In the upper and lower parts of the spectrum, the curve represents more energy [12]. Spectral skewness of input signal is −0.034447.

Spectral flux [12]: The spectral flux (SF) is the magnitude of the average spectrum difference between two consecutive frames in the provided clip [12].

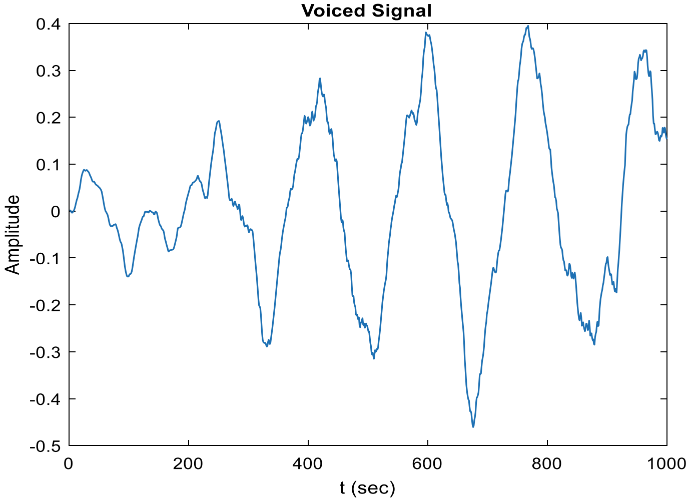


Fig. 7 Voiced segment

The next step in the proposed research is extraction of features from input signal in three separate domains. In the time domain, we use autocorrelation method to measure RMS, ZCR, and pitch salience ratio.

Root mean square (RMS) [12]: The RMS represents the square root of the mean audio amplitude over a given time period [12]. It checks the sound of the audio frame.

$$\mathrm{RMS}_j = \sqrt{\frac{1}{N}\sum_{m=1}^{N} x_j^2(m)} \tag{3}$$

Pitch saliency ratio [12]: This is the ratio of silent frames to maximum frames in the music signal [12]. If RMS < 10%, the frame is silent.

We calculated the characteristics of a frequency domain such as bandwidth, spectrogram, frequency centroid, spectral centroid, and pitch.

Bandwidth [12]: This refers to the frequency range of the signal containing data [12]. It is calculated according to the equation:

$$B_j = \sqrt{\frac{\int_0^{\omega_0} (\omega - \omega_c)\left|X_j(\omega)\right|^2 \mathrm{d}\omega}{\int_0^{\omega_0} \left|X_j(\omega)\right|^2 \mathrm{d}\omega}} \tag{4}$$

The bandwidth of the input signal is shown in Fig. 8.

Spectrogram [12]: This is a three-dimensional illustration as depicted in Fig. 9. The X-axis represents the properties of time. The Y-axis shows the frequency components of the audio signal. Dark region refers to the strength of an audio signal at that frequency [13]. The spectrogram divides the signal into overlapping segments, each

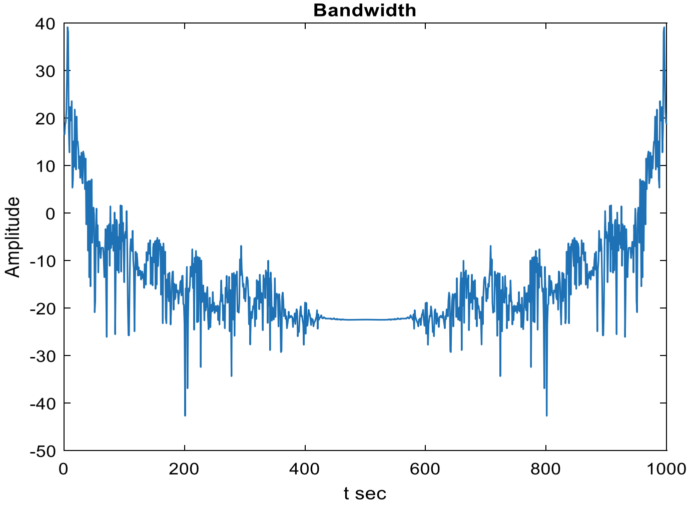


Fig. 8 Bandwidth of input signal

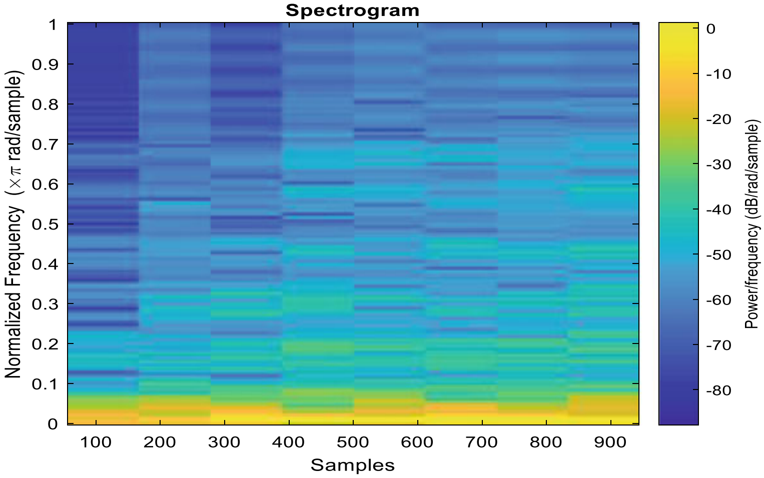


Fig. 9 Spectrogram of input signal

segment is filtered by a Hamming window, and the output is provided using N-point DFT [12].

Frequency Centroid [12]: It maintains signal brightness. It is computed using equation:

$$\omega_c j = \frac{\int_0^{\omega_0} \omega \left|X_j(\omega)\right|^2 \mathrm{d}\omega}{\int_0^{\omega_0} \left|X_j(\omega)\right|^2 \mathrm{d}\omega} \quad (5)$$

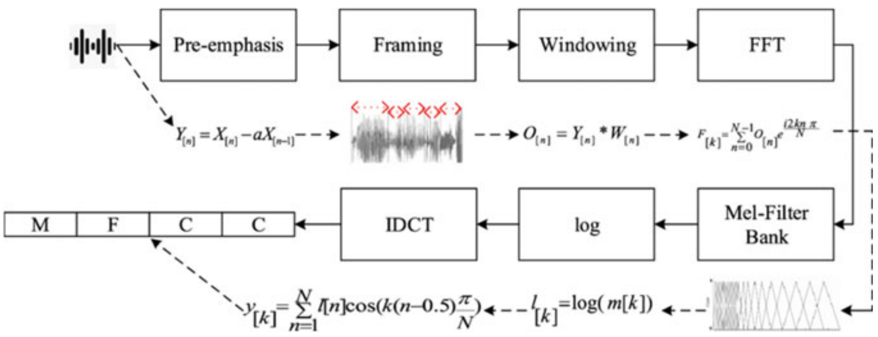

Pre-emphasis
Framing
Windowing
FFT
$Y_{[n]} = X_{[n]} - aX_{[n-1]}$
$O_{[n]} = Y_{[n]} * W_{[n]}$
$F_{[k]} = \sum_{n=0}^{N-1} O_{[n]} e^{-\frac{j2kn\pi}{N}}$
M
F
C
C
IDCT
log
Mel-Filter Bank
$y_{[k]} = \sum_{n=1}^{N} l[n] \cos(k(n-0.5)\frac{\pi}{N})$
$l_{[k]} = \log(m[k])$

Fig. 10 MFCC coefficient estimation process

Spectral centroid [12]: It deals with the chromatic variation of sound, i.e., the high frequency components of the spectrum. It is calculated using the formula:

$$C_r = \frac{\sum_{k=1}^{N/2} f[k]|X_r[k]|}{\sum_{k=1}^{N/2}|X_r[k]|} \tag{6}$$

Pitch [12]: Pitch or tone refers to the basic wavelength of the human voice [12]. Input signal pitch is 22.050 kHz.

In coefficient domain, we computed Mel Frequency Cepstral Coefficients (MFCCs). Firstly, speech data is emphasized, then emphasized data is framed according to a defined time. Then, it is applied a Hamming windowing function. Next, discrete Fourier transform is carried out to the data. Logarithm is applied to the processed data by applying mel-scale. Finally, MFCC data is obtained by applying inverse discrete cosine transform. The entire process of MFCC coefficient estimation is shown in Fig. 10.

4 SVM Classification

The "support vector machine" (SVM) is an inspected machine learning algorithm that can be used for classification or regression challenges. However, it is mostly used in classification problems [14]. Of the SVM algorithm, we plot each data item as a point in *n*-dimensional space (where *n* is the number of feature vectors we have) and the value of each feature is the value of the specific coordinate. Next, we classify it by finding a hyperplane that separates the two classes as in Fig. 11. The support vector is a compilation of individual observations. The SVM classifier is the boundary that best separates the two classes (hyperplane/line). [14] The kernel approach plays a key role in correctly classifying a new object (test case) from the available examples (train cases). The kernels use a collection of mathematical operations to change the order of real objects. The process of rearranging an object is called the mapping

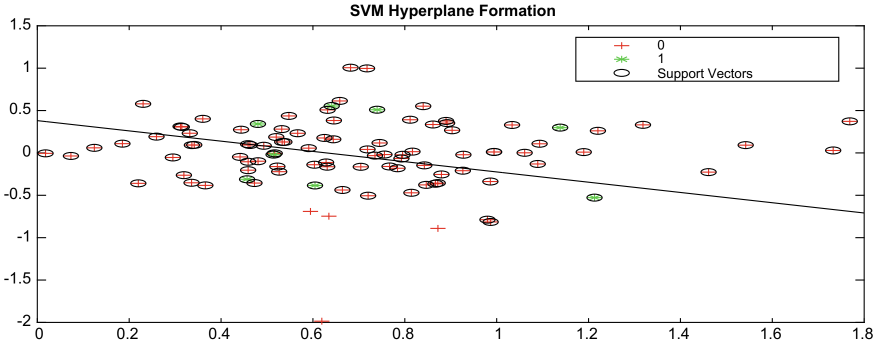
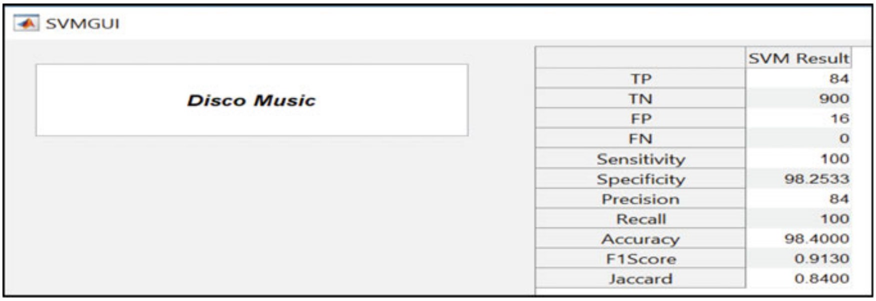


Fig. 11 SVM hyperplane

process [15]. The function of the kernel refers to the dot product of the input data points, which are transformed [16] and mapped to the high dimensional function space. Feature vectors input into the SVM classification include STE, ZCR, pitch, spectral flux, spectral centroid, and three MFCC modules. 70% of total signals in dataset are taken as training samples and remaining 30% are taken as test samples. There are 10 different genres of music in the used dataset. The linear kernel function is used for the experiment.

5 Result Analysis

The input signal can be also classified as blues, classical, country, rock, reggae, jazz, metal, hiphop, pop, disco, etc., depending upon which genre signal is taken. In this paper, we show the classification of input signal as disco. The performance metrics of SVM classifier computed for the experiment with respect to this input signal are shown in Fig. 12. SVM learning in the proposed system demonstrated better results



Disco Music

	SVM Result
TP	84
TN	900
FP	16
FN	0
Sensitivity	100
Specificity	98.2533
Precision	84
Recall	100
Accuracy	98.4000
F1Score	0.9130
Jaccard	0.8400

Fig. 12 Classification of music genre and performance metrics of SVM classifier

in classification of music genres with an accuracy of 98.4% that is comparatively higher.

6 Conclusion

In this paper, we demonstrated the classification of music genres using SVM classifier with linear kernel function. The music signals from GTZAN dataset were taken for experiment. The feature vectors from three different domains like time, frequency and cepstral domain were computed and given as the inputs to the classifier. The SVM classifier outperformed well giving the classification accuracy of 98.4% which is higher as compared to accuracies obtained by researchers in literature review.

References

1. Thiruvengatanadhan, R.: Music genre classification using SVM. Int. Res. J. Eng. Technol. (IRJET) **05**(10), 1059–1061 (2018)
2. Joder, C., Essid, S., Richard, G., Member, S.: Temporal integration for audio classification with application to musical instrument classification. IEEE Trans. Speech Audio Process. **17**(1), 174–186 (2009)
3. Serwach, M., Stasiak, B.: GA-based parameterization and feature selection for automatic music genre recognition. In: Proceedings of 2016 17th International Conference on Computational Problems of Electrical Engineering, CPEE (2016)
4. Van Dijk, L.: Bachelorthesis Information Science: Finding Musical Genre Similarity Using Machine Learning Techniques, pp. 1–25. Radboud Universiteit Nijmegen (2014)
5. Xu, C., Maddage, N.C., Shao, X., Cao, F., Tian, Q.: Music genre classification using support vector machines
6. Mutiara, A.B., Refianti, R., Mukarromah, N.R.A.: Musical genre classification using support vector machines and audio features. TELKOMNIKA **14**(3), 1024–1034
7. Patil, N.M., Nemade, M.U.: Music genre classification using MFCC, K-NN and SVM classifier. Int. J. Comput. Eng. Res. Trends **4**(2), 43–47 (2017)
8. Aryafar, K., Jafarpour, S., Shokoufandeh, A.: Automatic musical genre classification using sparsity-eager support vector machines. In: 21st International Conference on Pattern Recognition (ICPR 2012), 11–15 Nov 2012, Tsukuba, Japan
9. Kyaw, L.Y., Renu: Using support vector machine for music genre classification
10. http://marsyas.info/downloads/datasets.html
11. Pradeep Kumar, D., Sowmya, B.J., Chetan, K.G.S.: A comparative study of classifiers for music genre classification based on feature extractors. IEEE, pp. 190–194 (2016)
12. Patil, N.M., Nemade, M.U.: Content-based audio classification and retrieval using segmentation, feature extraction and neural network approach. In: Bhatia, S., Tiwari, S., Mishra, K., Trivedi, M. (eds.) Advances in Computer Communication and Computational Sciences. Advances in Intelligent Systems and Computing, vol. 924. Springer, Singapore. https://doi.org/10.1007/978-981-13-6861-5_23
13. Yandre, M.G.C., Oliveira, L.S., Silla, Jr., C.N.: An evaluation of convolutional neural networks for music classification using spectrograms. Appl. Soft Comput. 1–39 (2016)
14. https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

15. https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989
16. Mutiara, A.B., Refianti, R., Mukarromah, N.R.A.: Musical genre classification using support vector machines and audio features. TELKOMNIKA **14**(3), 1024–1034 (2016)