International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-03, JUNE 2021

Classification of Music Genre using Machine Learning Techniques

Sunil Kumar R [1*], Prakruthi P S [2], Suhas Shetty [3]

[1] **B.E. Student, Dept. of Mechanical Engineering, Sapthagiri College of Engineering, Bangalore, India**

[2] **B.E. Student, Dept. of Telecommunication Engineering, Dayanand Sagar College of Engineering, Bangalore, India**

[3] **B.E. Student, Dept. of Mechanical Engineering, Sapthagiri College of Engineering, Bangalore, India *Corresponding author: sunilrudrakumar@gmail.com**

Abstract: Music has a very significant role in life. Music brings people together in common that keeps communities whole. Music genre classification is a key element in music information retrieval (MIR). So, the classification of genre may be helpful in explaining interesting problems like setting up song references, song recommendation or may look for societies who will like that specific song genre. Recently, the rapid growth in the use of digital music has received a great deal of attention. Various musical genres are characterized by unique attributes shared by its members. Musical annotation for this genre is today carried out manual. The automatic classification of musical genres may aid and eliminate human users in this process and would be an important addition to the music genre system. The object of our project and our research is to find a better machine learning algorithm compared to the pre-existing one templates that predict the genre of songs. We cross-referenced the performance of all these models and registered their results in prediction details. One of the models, a convolutional neural network, was shown to have the highest accuracy when given only the spectrograms as the dataset amid the rest of the models.

Keywords —Audio Classification, CNN, Feature Extraction, Music Information Retrieval, Neural Networks, Spectrograms

I. INTRODUCTION

Nowadays, a personal music collection can have hundreds of tracks, and a professional music collection can have tens of thousands. The majority of music files are listed by song title or artist name, which can make finding a song related with a specific genre challenging. When dealing with huge music databases, warehouses need a lot of effort and time, especially when manually classifying audio genres. Not just based on music, but also on the basis of words, music has been categorized into genres and subgenres. This makes categorization more difficult. Businesses currently use music classification, either to make recommendations or to their clients (like Spotify, Soundcloud) or just as a commodity (this type of. As a result, algorithms must be able to determine without having to ask the answer to the question, "What sort of music do you like?" Recognizing music genres has been the first step in this strategy.

Music analysis is done by means of a piece of digital signature to assess what kind of music a person is interested in hearing for several variables such as acoustics, danceability, tempo, energy, etc.

Classifying music genre is maybe the finest general information for clarification of the music content. Music engineering promotes the practice of categories and families to organize sound accumulations through this clarification, and therefore greatly improves the demand to unwarrantedly group audio recordings into categories.

Furthermore, the recent developments in category organization here are still a question of describing a type or whether they mainly communicate the understanding and taste of a consumer.

These genres are man-made. The traits typically shared by its members are separated from a genre of music. Typically,

the rhythms, instrumentation and harmonic content of the music have to do with these traits. In the Music Information Retrieval (MIR), the field dedicated to viewing, organizing and looking for big music collections, music files are classified into their particular genre.

Up to now genre classification has been accomplished manually for digitally available music. Through these techniques, the creation of sound information recovery systems for music would be an important contribution to the automatic genre classification system. A framework for

 DOI : 10.35291/2454-9150.2021.0335

development and evaluation of feature for any type of content-based analysis of music signals can be provided by automatic classification of music into genres.

II. Problem Statement

Music plays an awfully critical part in people's lives. Different communities and bunches tune in to diverse sorts of music. Communities can be recognized by the sort of tunes that they compose, or indeed tune in to. Music bring like-minded individuals together and is the stick that holds communities together. One primary highlight that isolates one kind of music from another is the class of the music.

1. To construct a machine learning model which classifies music into its particular sort.

2. Compare the details of this machine learning model with pre-existing models and draw conclusions from them.

OBJECTIVE: Creating a machine learning model that classifies music into different genres proves that there exists a way which automatically classifies music into its classes based on various distinctive highlights, rather than physically entering the genre.

III. Related Work

The predominance of the classification of programmed musical genres has been growing steadily for some years.

Some of the related work is outlined below.

I. In Tom LH Li et al., (2010). Automatic extraction of music patterns by means of a convolutional neural network, they tried to understand the main characteristics that really help to construct the optimal model for the classification of the musical genre. The primary objective of this article is to propose a new approach to extract the characteristics of musical patterns from the audio file using Convolution Neural Network. Their primary goal is to explore the possibilities of applying CNN in music information retrieval (MIR). Their findings and experiences show that CNN has a strong ability to capture informative characteristics of the variable music model. Characteristics extracted from sound clips such as statistical spectral characteristics, the tempo and pitch are less reliable and produce less precise models. Hence, the approach made by them to CNN, where the musical data have similar characteristics to image data and mainly it requires very less prior knowledge. GTZAN was the dataset that was considered. It is divided into ten genres, each with 100 audio clips. Each audio clip is 30 seconds long, with a sample rate of 22050 Hz and a 16-bit resolution. Multiple categorization models were used to analyses the musical patterns using the WEKA tool. The accuracy of the classifier was 84 percent at first, and it gradually improved. In compared to the MFCC, chroma, and temperature features, the CNN features yielded better results and were more trustworthy. Parallel computation on different combinations of genres can still improve accuracy.

II. In Lu L. et al., (2002). They presented their research on audio content categorization and segmentation. An audio stream is divided into segments based on the kind of audio or the identity of the speaker. Their strategy is to create a strong model capable of identifying and segmenting an audio signal into speech, music, ambient sound, and quiet. This categorization is broken down into two primary phases, making it useful for a variety of diverse uses. Discrimination between speech and non-speech is the initial stage. A unique approach based on KNN (K-nearestneighbor) and linear spectral pairs-vector quantization (LSP-VQ) has been devised in this paper The non-speech class is then divided into three categories using a rule-based categorization method: music, ambient noises, and quiet. They've used a few unusual and novel characteristics like noise frame ratio and band periodicity, which are not only presented but also explained in depth. A speech segmentation method was also incorporated and built. This is an unsupervised situation. It employs a unique correlation analysis methodology based on quasi-GMM and LSP correlation analysis. The model can enable open-set speakers, online speaker modelling, and real-time segmentation without any prior knowledge of anything

III. In Tzanetakis G. et al., (2002). They have mostly investigated how to categories audio signals into a hierarchy of music styles using automated classification. They argue that these music genres are categorization designations devised by humans only for the purpose of organizing musical compositions. Some of the common qualities are used to classify them. These features are usually linked to the instruments employed, the rhythmic frameworks, and, most importantly, the harmonic content of the song. Genre hierarchies are commonly used to organize very vast music collections available on the internet. Three feature sets have been proposed: timbral texture, rhythmic content, and pitch content. Training statistical pattern recognition classifiers with some real-world sound collections was used to analyze proposed features in order to analyses performance and relative importance. Both entire file and real-time frame-based classification techniques are given in this research. This model can accurately classify around 61 percent of ten music genres using the proposed feature sets.

IV. In Hareesh Bahuleyan, (2018). Music Genre Classification using Machine Learning techniques the research presented a method for automatically classifying music by assigning tags to the songs in the database. The library of the user. It looks at both Neural Networks and Artificial Intelligence. Machine Learning algorithms have been used in the past in a traditional way.as well as to attain their objective. The first method employs a Convolutional Neural Network that is trained from start to finish utilizing the features of audio signal Spectrograms (pictures). The second method employs several Machine Learning techniques such as Logistic Regression, Random Forest, and others, as well as hand-crafted characteristics from the audio signal's temporal and frequency domains.ML algorithms

 DOI : 10.35291/2454-9150.2021.0335

such as Logistic Regression, Random Forest, Gradient Boosting (XGB), and Support Vector Machines are used to classify the music into genres utilizing manually extracted features such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, and Spectral Centroid (SVM). They found that the VGG-16 CNN model provided the highest accuracy after analyzing the two approaches separately. The optimal model with 0.894 accuracy was produced by developing an ensemble classifier with VGG-16 CNN and XGB.A further aim is to achieve good precision so that the model correctly categorizes the new music in its kind.

IV. DATA SET

We employ a portion of the Free Music Archive (FMA) , an open and easily accessible dataset suited for evaluating numerous tasks in MIR (Music Information Retrieval), a topic concerned with browsing, searching, and organizing huge music collections. The increased interest in feature and end-to-end learning in the community is hampered by the scarcity of large audio datasets.

The FMA intends to overcome this barrier by offering 917 GB and 343 days of Creative Commons-licensed audio from 106,574 tracks by 16,341 artists and 14,854 albums, organized into 161 genres. It includes full-length and highquality audio, pre-computed features, metadata, tags, and free-form text such as biographies at the track and user level.

The audio and metadata in the collection are from the Free Music Archive, a free and open resource hosted by WFMU, the country's longest-running freeform radio station.

Genres:

The FMA is ideal for MGR because it includes fine genre information, such as various (sub-)genres connected with individual tracks, a built-in genre hierarchy (Table 1), and is commented on by the artists.

Subsets:

The dataset has been split into the following sets, each of which is a subset of the larger set, to make it useful as a development set or for those with limited computational resources.

1.Full: It has all 161 genres, with 1 to 38,154 tracks per genre and up to 31 genres per track and is uneven.

2. Large: the whole dataset, with audio limited to 30-second segments selected from the center of the tracks (or the complete track if the track is less than 30 seconds). The data is reduced by a factor of ten as a result of this reduction.

3. Medium: This includes 25,000 30-second clips, but only one of the 16 top genres per clip, and is genre imbalanced with 21 to 7,103 clips per top genre.

4. Small: To create a balanced subset, the top 1,000 clips from the medium set's eight most popular genres were chosen using the same approach. The subset is made up of 8,000 30-second clips from eight major genres, with 1,000 clips per genre and one root genre per clip. This subset is identical to the widely used GTZAN, except it includes the FMA's metadata, pre-computed features, and copyright-free audio.

id	parent	top_level	title	#tracks
38	None	38	Experimental	38,154
15	None	15	Electronic	34,413
12	None	12	Rock	32,923
1235	None	1235	Instrumental	14,938
25	12	12	Punk	9,261
89	25	12	Post-Punk	1,858
1	38	38	Avant-Garde	8,693

Table 1: An excerpt of the genre hierarchy, stored in genres.csv. Some of the 16 top-level genres appear in the top part, while some second- and third-level genres appear in the bottom part.

S.No	Genre Name	Count
1	Electronic	1000
2	Experimental	1000
3	Folk	1000
4	Hip-Hop	1000
5	Instrumental	1000
6	International	1000
7	Pop	1000
8	Rock	1000
	Total	**8000**

Table -2: Number of instances in each genre class

dataset	clips	genres	Length [s]	Size [GiB]	Size #days
small	8000	8	30	7.4	2.8
medium	25000	16	30	23	8.7
large	106574	161	30	98	37
full	106574	161	278	917	343

Table -3: Variants of FMA dataset

dataset	#clips	#artists	year	audio
GTZAN	1000	~300	2002	yes
MSD	1,000,000	44745	2011	no
AudioSet	2,084,320	-	2017	no
Artist20	1413	20	2007	yes
AcousticBrainz	2,524,739	-	2017	no

Table -4: List of other audio datasets

Table 2 shows the total number of audio clips in each category. Because each audio file is about 1 megabyte in size,

 DOI : 10.35291/2454-9150.2021.0335

the FMA small dataset is around 8 GB in size. The fma medium dataset (25,000 tracks of 30s, 16 unbalanced genres (23 GiB)), fma large dataset (106,574 tracks of 30s, 161 unbalanced genres (98 GiB), and fma full dataset (106,574 tracks of 30s, 161 unbalanced genres (917 GiB) are all available forms of the FMA dataset.

In addition, the dataset includes metadata that allows users to experiment without having to deal with feature extraction. The librosa Python library, version 0.5.0, was able to extract all of these features. Each feature set (excluding zero-crossing rate) is calculated on 2048-sample windows separated by 512-sample hops. The mean, standard deviation, skew, kurtosis, median, minimum, and maximum were then computed across all windows. The 518 precomputed features are dispersed across features. All tracks have a csv file (included in the fma metadata). Other open source and freely available datasets are listed in Table 4.

V. METHODOLOGY

The details of data pre-processing are described in this section, followed by a description of the recommended strategy to this classification challenge.

1.Deep Neural Networks

We can classify music genres without using hand-crafted attributes thanks to deep learning techniques. Convolutional neural networks (CNNs) are an excellent choice for picture classification. A CNN is given the 3-channel (R-G-B) matrix of a picture, which it uses to train itself on those images. The sound wave is represented as a spectrogram in this work, which can be viewed as an image (Nanni et al., [4]). (Lidy and Schindler, [15]). The CNN's objective is to predict the genre label using the spectrogram (one of eight classes).

2.Spectrogram Generation

A spectrogram is a two-dimensional representation of a signal, with the x-axis representing time and the y-axis representing frequency. Each audio sample was transformed into a MEL spectrogram (with MEL frequency bins on the y-axis) in this investigation. The following are the STFT settings used to construct the power spectrogram:

Sampling rate (sr) = 22050

Window size (n_fft) = 2048

Hop length = 512

X_axis: time

Y_axis: MEL

Highest Frequency (f_max) = 8000

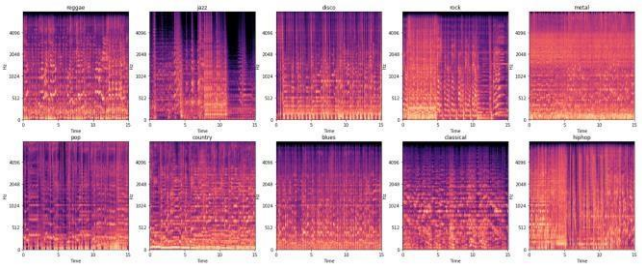

Fig- 1: Sample spectrograms for 1 audio track from each music genre

3.Convolutional Neural Networks

Figure 2 shows that the spectrograms of audio signals from various classes have some distinct characteristics. As a result, spectrograms can be thought of as "pictures" that may be fed into a CNN. Figure 3 depicts a preliminary framework for the CNN model.

4.Feed Forward Network

A CNN is a feed-forward network, which means that input samples are fed into the network and transformed into an output; in supervised learning, the output is a label, which is a name given to the input. That is, they map raw data to categories, recognizing patterns that indicate whether an input image should be classified "folk" or "experimental," for example. A feedforward network is trained on tagged images until it can guess their categories with the least amount of error.

The network uses the trained set of parameters (or weights, collectively known as a model) to categories input it has never seen before. A trained feedforward network can be subjected to any random collection of photos, and how it classifies the first shot will not necessarily change how it classifies the second. The net will not perceive a spectrogram of an experimental song after seeing a spectrogram of a traditional tune. That is, a feedforward network has no concept of time order and just considers the current example it has been exposed to as input. Feedforward networks have amnesia about their recent past, remembering only the early phases of training nostalgically.

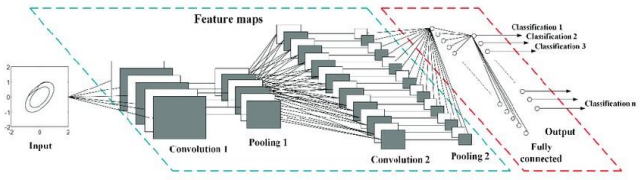

Fig -2: CNN Framework

5.Operations of CNN

Each block in a CNN consists of the following operations:

Pooling: The dimension of the feature map acquired from the convolution step is reduced using this method. We only keep the element with the highest value among the four elements of the feature map that are covered in this window by max

 DOI : 10.35291/2454-9150.2021.0335

pooling with a 2x2 window size. With a preset stride, we move this window across the feature map.

Non-linear Activation: The convolution procedure is linear, and we need to introduce some non-linearity to make the neural network more powerful. On each element of the feature map, we can use an activation function like Rectifier Linear Unit (ReLU) for this.

Convolution: This step involves a matrix filter (say 3x3 size) that is moved over the input image which is of dimension image width x image height. The filter is first placed on the image matrix and then we compute an element-wise multiplication between the filter and the overlapping portion of the image, followed by a summation to give a feature value.

Convolutional Recurrent Neural Network:

We train a Convolutional Recurrent Neural Network, which is a mix of convolutional and recurrent neural networks, to evaluate the performance increase that CNNs may accomplish.

Recurrent Neural Networks: Recurrent nets are a form of artificial neural network that recognizes patterns in data sequences like as text, genomes, music, video, or numerical time series data from sensors, stock exchanges, and government agencies. These algorithms have a temporal component since they consider time and sequence.

Even pictures, which can be deconstructed into a number of patches and processed as a sequence, may be used with RNNs.

Recurrent networks differ from feedforward networks in that they include a feedback loop related to their previous judgments, absorbing their own outputs as input moment by moment. Recurrent networks are frequently described as having memory. Adding memory to the neural network

Long Short-Term Memory Networks (LSTMs) are a kind of RNN that can learn long-term dependencies. Hochreiter & Schmidhuber were the ones who presented them (1997).

CRNN Model Specifications:

The CNN-RNN model, also known as the CRNN model, consists of three 1-dimensional Convolution Layers, an RNN's LSTM layer, and a fully linked dense layer, which serves as the output layer. To make the evaluation and comparison fair, the batch size utilized in this model is 32, and the number of epochs is kept at 30. ReLU is the activation function utilized. Dropout of 0.2 is introduced in the hidden layers to avoid data overfitting.

CNN-RNN Model in Parallel:

This model runs the CNN and RNN models in tandem while maintaining the same metrics and convolution factors as the prior models. The goal is to compare two siblings.

Specifications for Implementation:

The spectrogram pictures are 150 × 150 pixels in size. A 512-unit hidden layer is developed for the feed-forward network connected to the conv base. In neural networks, over-fitting is a prevalent problem. To avoid this, we implemented the following strategy:

Dropout [21]: This is a regularization method in which certain neurons are turned off (their weights are adjusted to zero) at random during training. We anticipate the final output with a new mix of neurons in each iteration, therefore randomizing the training cycles. A 0.2 dropout rate is utilized, which means that a particular weight is set to zero with a chance of 0.2 throughout an iteration.

The dataset is divided into three sections: training (80%), validation (10%), and testing (10%). (10 percent). All of the comparisons utilize the same split.

TensorFlow is used to implement the neural networks in

Python. With a batch size of 64, all models were trained for 30 epochs. The ADAM optimizer was used to optimize these neural networks. One iteration of the whole training set is referred to as an epoch.

VI. Feature Extraction

This section describes the features taken from previous models that have been compared to the proposed model. Time domain and frequency domain features are the two types of features. Librosa, a Python package, was used to extract the features.

Time Domain Features

These are the characteristics that were retrieved from the raw data.

1.Central moments: This consists of the mean, standard deviation, skewness and kurtosis of the amplitude of the signal.

2.ZCR (Zero Crossing Rate): The signal's sign shifts from positive to negative at this moment. The amount of zerocrossings present in each frame is calculated after dividing the 30-second signal into smaller frames. Representative characteristics are the average and standard deviation of the ZCR over all frames.

3.Root Mean Square Energy (RMSE): RMSE is calculated frame by frame and then the average and standard deviation across all frames is taken.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

4.Tempo: The tempo of a piece of music refers to how quick or slow it is. The tempo is measured in BPM (beats per minute) (BPM). We use Tempo's aggregate mean, which fluctuates from time to time.

Frequency Domain Features:

The Fourier Transform is used to convert the audio signal into the frequency domain. The following characteristics are then extracted.

1.Mel-Frequency Cepstral Coefficients (MFCC):
MFCCs were first introduced in the early 1990s by Davis and Murmelstein and have shown to be quite effective for tasks such as voice recognition.

 DOI : 10.35291/2454-9150.2021.0335

2.Chroma Features: This is a vector that represents the signal's total energy in each of the 12 pitch classes. (C, C#, D, D#, E, F, F#, G, G#, A, A#, B, C, C#, D#, D#, D#, D#, D#, D#, D#, D#, D#, D#, D# The mean and standard deviation are calculated using the sum of the chroma vectors.

3.Spectral Centroid: This is the frequency at which the majority of the energy is centered. It's a magnitudeweighted frequency, which is computed as follows:

$$f_c = \frac{\sum_k S(k) f(k)}{\sum_k fk},$$

where S(k) is the spectral magnitude of frequency bin k and f(k) is the frequency corresponding to bin k.

Spectral Contrast: Each frame is split into a number of frequency bands that are predetermined. The spectral contrast is determined as the difference between the greatest and minimum magnitudes within each frequency range.

Spectral Roll-off: This characteristic refers to the frequency at which 85 percent of the total energy in the spectrum is found.

The mean and standard deviation of the values collected across frames is regarded as the representative final feature that is supplied to the model for each of the spectral characteristics mentioned above.

These definitions are based on (Hareesh Bahuleyan's [3])

VII. CLASSIFIER

A quick summary of the machine learning classification utilized in this work is provided in this section.

1.Logistic Regression (LR):

In general, this linear classifier is employed for binary classification tasks. The LR is a one-vs-rest approach for this multi-class classification issue. In other words, eight different binary classifiers are trained. The highest probability class from the 8 classifiers is selected as the projected class during the test period.

2.Simple Artificial Neural Network (ANN): A computational model based on the structure and functions of biological neural networks is known as an artificial neuron network (ANN). Information flowing through the network alters the ANN structure since a neural network, based on this input and output, alters or learns to a certain extent. ANNs are regarded non-linear methods for statistical modelling of data, which model or find patterns for complicated interactions between inputs and outputs. This model uses a CSV file of the hand-made features derived from the librosa library audio clips and provides an output with a feature comparable to the above-described logical regression.

VIII. EVALUATION

Metrics

The following statistic will be used to assess the performance of the models.

Accuracy: Refers to the percentage of test samples that are correctly categorized. This measure assesses how close the model's prediction is to the actual data.

IX. RESULTS:

The findings of the different modelling techniques described in Section 4 and their accuracies are reviewed in this section.

Spectrogram-based models	**Accuracy**
CNN model	**88.54%**
CRNN model	53.5%
CNN-RNN model	56.4%
Feature based models	
Logistic Regression (LR)	60.892%
Simple Artificial Neural Network (ANN)	**64.0625%**

Table 5: Displays these accuracies.

CRNN model was trained for 30 epochs using ReLU activation function with Dropout of 0.2 to introduce it in the hidden layers to avoid data overfitting. As seen in the result, the model has low bias but high variance implying the model is overfitting a bit to training even after using several regularization techniques. Overall, this model got to around 53.5% accuracy on the validation set.

CNN-RNN model had an accuracy of around 56.4%. Both models have very similar overall accuracies which is quite interesting, but their class wise performance is very different. CNN-RNN model has a better performance for Experimental, Folk, Hip-Hop and Instrumental genres. The ensembling of both these models should produce even better results.

With a test accuracy of 88.54 percent, the CNN model that utilizes only the spectrogram as an input to predict the music genre has the greatest performance in terms of accuracy. Even when the regularization criteria are changed or the epochs are increased, the CRNN and CNN models do not provide excellent accuracy, regardless of how robust and sophisticated their designs are. The small dataset of 8000 audio recordings might be the cause of the poor test accuracy rate. An increase in the dataset might help these models be more accurate.

X. CONCLUSION

The Free Music Archive tiny dataset is used to investigate music genre categorization in this article. We suggested a simple solution to the classification problem and compared it to several more complicated, reliable models. We also compared the models according on the type of data they were getting. Spectrogram pictures for CNN models and audio characteristics contained in a csv for Logistic Regression and ANN models were used as inputs to the models. With a test accuracy of 64%, Simple ANN was judged to be the best feature-based classifier among Logistic Regression and ANN models.

The spectrogram shows not only the signal's frequency content, but also its energy. The vertical axis of the

 DOI : 10.35291/2454-9150.2021.0335

spectrogram depicts frequency, with the lowest frequencies at the bottom and the highest frequencies at the top, and the horizontal axis represents time, running from left to right. Colors add a third dimension to the spectrogram display; different colors signify different energy levels. So, by combining it with a classifier like CNN, a powerful image classifier, we managed to get 88.5% accuracy.

REFERENCES

[1]. Tom LH Li, Antoni B Chan, and A Chun. Automatic musical patterns feature extraction using convolutional neural networks. In Proc. Int. Conf. Data Mining and Applications, 2010

[2]. Music Genre Classification using Machine Learning Algorithms: A comparison Snigdha Chillara1, Kavitha A S2, Shwetha A Neginhal3, Shreya Haldia4, Vidyullatha K S5

[3]. FMA: A DATASET FOR MUSIC ANALYSIS

Michaël Defferrard† Kirell Benzi† Pierre Vandergheynst† Xavier Bresson‡

[4]. HASITHA B. ARIYARATNE, ZHANG D., "A

Novel Automatic Hierarchical Approach to Music

Genre Classification", 2012 IEEE International Conference on Multimedia and Expo Workshops, 2012.

[5]. ZHANG T., "Semi-Automatic Approach for Music Classification", Proceedings of SPIE Vol. 5242 Internet Multimedia Management Systems IV, 2003.

[6]. Hareesh Bahuleyan, Music Genre Classification using Machine Learning Techniques, University of Waterloo, 2018

[7]. SCARINGELLA N., ZOIA G. and

MLYNEKTHEM D., "Automatic Genre Classification of Music Content", IEEE SIGNAL PROCESSING MAGAZINE, MARCH 2003.

[8]. TZANETAKIS G. & COOK P., "SOUND ANALYSIS USING MPEG COMPRESSED

AUDIO", 2000.

[9]. Thomas Lidy and Alexander Schindler. Parallel convolutional neural networks for music genre and mood classification. MIREX2016, 2016.

[10]. T Bertin-Mahieux, D PW Ellis, B Whitman, and P Lamere. The million-song dataset. In ISMIR, 2011.

[11]. PYE D., "Content-Based Methods for the Management of Digital Music", 2000

[12]. Lonce Wyse, Audio Spectrogram representations for processing with Convolutional Neural Networks, National University of Singapore, 2017.

[13]. Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audiobased music classification and annotation. IEEE Transactions on Multimedia, 13(2), 303–319

[14]. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, 2014.

[15]. "Automatic Music Genres Classification using Machine Learning" -(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 8, 2017-Muhammad Asim Ali,Zain Ahmed Siddiqui

[16]. Lonce Wyse, Audio Spectrogram representations for processing with Convolutional Neural Networks, National University of Singapore, 2017.

[17]. A Porter, D Bogdanov, R Kaye, R Tsukanov, and X Serra. Acousticbrainz: a community platform for gathering music information obtained from audio. In ISMIR, 2015.

[18]. J F Gemmeke, D PW Ellis, D Freedman, A Jansen, W Lawrence, R C Moore, M Plakal, and M Ritter. Audio set: An ontology and humanlabeled dataset for audio events. In ICASSP, 20

 DOI : 10.35291/2454-9150.2021.0335