CLASSIFICATION OF MUSIC GENRE USING MACHINE LEARNING

Isha Pathania[1], Dr Anuj Bhardwaj[2]

[1] M.E. Dept. of Computer Science and Engineering, Chandigarh University, Gharuan
[2]Assistant Professor, Dept. of Computer Science and Engineering, Chandigarh University, Gharuan

---***---

Abstract - *The classification of music files in accordance with their genre is a very demanding task, particularly in the zone of mainly MIR(Music Information Retrieval).We will compare the performance of the two classes of models in this study. Deep learning approach is the first approach and for the purpose of predicting the genre label of a signal, a CNN model would be trained usually from end-to-end, spectrogram is used to carry out this process. Hand-crafted features were used in the second approach which was also the last approach. Four out of traditional machine learning classifiers will be trained beside these features and their performance would be compared afterwards. Identification of features that would help the classification of this task needs to be performed. A set data of Audio will be used for the experiments and after that an AUC value of 0:894 is reported for an ensemble classifier which merges the two proposed methods. The databases of online music are growing day by day and nowadays it is very simple to access online music. Nowadays music has become a very salient proportion of the Internet content: the most salient source of pieces of music is the net.*

***Key Words*: Machine Learning, Convolutional Neural Network, Music genre, Support Vector Machine, Deep Neural Network**

1.INTRODUCTION

Music has nowadays become a significant proportion of Internet content: the most important source of pieces of music is the net.In this context, the automatic procedures which are capable of dealing with huge amounts of music in digital formats are salient, and Music Information Retrieval (MIR) has become an imperative research area. Labels of music genre are very salient categories which are used to categorise and classify songs, albums and artists into broader groups which can share alike musical characteristics. The classification of music files in accordance with their genre is a very demanding task, particularly in the zone of mainly MIR(Music Information Retrieval).We will compare the performance of the two classes of models in this study. The first approach in this approach is a deep learning approach, where a CNN model is trained from end-to-end, generally to predict the genre label of a signal of an audio, exclusively using its spectrogram.

The signals of an audio can be automatically classified into a hierarchy of musical genres .Those genres of music are categorical labels which are mainly created by humans only to categorise many different pieces of music. They are classified using some of the most common characteristics. These characteristics are almost related to the different instruments that are used like rhythmic structures and mainly harmonic music content. Genre hierarchies are mostly used to structure very huge music collections which are available on web. There are three feature sets: firstly the timbral texture, secondly the rhythmic content and lastly the pitch content. The investigation of put forward features which are done in order to examine the performance and the relative importance, was usually done by training mostly the statistical pattern recognition classifiers by making utilisation of real-world audio collections.

The music can be automatedly categorised by providing various tags to the songs that are present in the library of the user. It surveys both the traditional method and the Neural Method of using ML(Machine Learning) algorithms to achieve their aim. The Neural Method approach uses CNN(Convolutional Neural Network) which is directed from end to end using the Spectrograms (images) features of the signal of audio. The second approach makes use of many ML(Machine Learning)algorithms just like Logistic Regression, Random forest and many more, here it utilizes features which are hand-crafted from the frequency domain and the time domain of the signal of audio. The physical features which are extracted like MFCC (Mel-Frequency Cepstral Coefficients),Chroma Features, Spectral Centroid and many more, can be used to categorise the music into its various genres by using Machine Learning(ML) algorithms like Logistic Regression, Random Forest, XGB (Gradient Boosting) and SVM (Support Vector Machines) .By making comparison of the two approaches isolatedly they came to a conclusion that model of VGG-16 CNN gave the most highest accuracy.

The various important features which generally contribute to build the optimal model which is utilized for Music Genre Classification can be easily understood. A novel approach can be made to take out musical pattern features of the audio file with the use of CNN(Convolutional Neural Network).The various chances of application of CNN in MIR (Music Information Retrieval) can be surveyed. By making use of many experiments and results, it can be concluded that CNN(Convolutional Neural Network) have the most powerful capacity to catch informative features from the musical pattern which are varying.

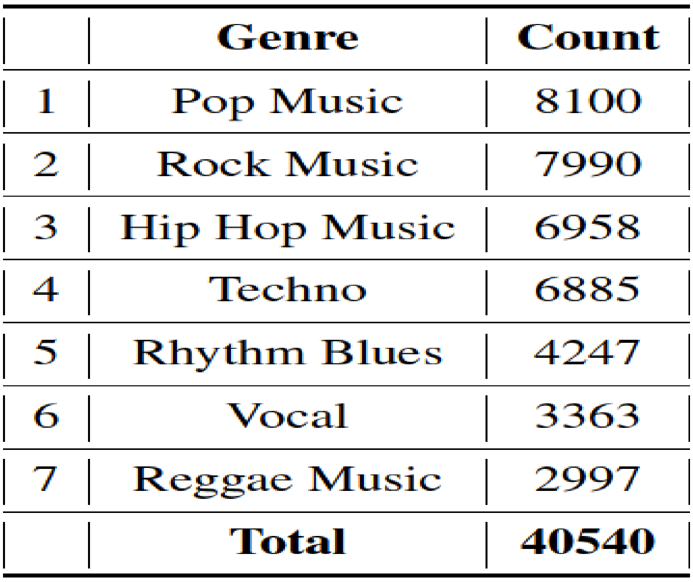
The various features that are extracted from the various audio clips like firstly the statistical spectral features,secondly the rhythmic content and lastly the pitch content are less dependable and produces very less accurate models.

2. CLASSIFICATION AND SEGMENTATION OF AUDIO CONTENT ANALYSIS

The classification and segmentation of the audio content analysis can be done in various ways, one way is in which a stream of audio is divided according to identity of the speaker or the audio type. The main approach is to build a strong model which will be capable of segmenting and classifying the audio signal given into the speech , music, environment sound and silence. This classification is processed in two main steps, which have made it satisfactory for many various other applications as well. The first step is non- speech and speech discrimination. Here, a novel algorithm which is mainly based on linear spectral pairs-vector quantization (LSP-VQ) and on KNN (K- nearest- neighbour) have been finally developed. The second and last step is to divide the class of non-speech into music, many environmental sounds, and silence with a classification of rule-based method. Here, utilisation is done by many rare features like the noise frame ratio. A speaker segmentation algorithm, is unsupervised and it mostly uses a novel scheme which is based on LSP correlation analysis and quasi – GMM . Without any previous knowledge of anything, this model could support the open-set speaker, the online speaker modelling and lastly the real time segmentation.

3. DATASET

Audio Set is used in this work, which is a large scale human annotated database of sounds. The database was made by extracting 10-second sound clips from a total of 2.1 million youtube videos. This study requires only the audio files that belong to the music category, specially having one of the seven genre tags, music genre are displayed in Table 3.1.



	Genre	Count
1	Pop Music	8100
2	Rock Music	7990
3	Hip Hop Music	6958
4	Techno	6885
5	Rhythm Blues	4247
6	Vocal	3363
7	Reggae Music	2997
	Total	**40540**

TABLE No. 3.1:NUMBER OF INSTANCES IN EACH OF GENRE CLASS

This table comprises of various music genre like reggae music, rock music, hip hop music, techno, rhythm blues and vocal. These music genres will be used for music genre classification which will be done by various machine learning approaches. Various data pre processing approaches and the various classifiers like Random forest will also be used. The training and testing of dataset will be used to find the accuracy of the various music genres. The F score will also be found. At last the confusion matrix will be plotted for Logistic Regression and other machine learning algorithms.

The total number of audio clips which is taken from Google will be tabulated in every category. The various audio clips in raw form of these sounds is not available in the Audio Set data release. However, the data produces the YouTubeID of the various communicating videos, beside the end and start times. Hence, the first task is to recover the given audio files. For this goal of the audio restoration from YouTube, the steps would be carried out which are given below:

A. Command line program named youtube-dl was made use of to download the required video in the format of mp4.

B. The files in the mp4 format are changed into the desired wav format by making the use of an audio converter whose name is Ffmpeg(it is a command line tool). Every wav file is approximately 880 KB in size, which concludes that the total amount of data used in this study is approximately thirty-four GB.

4. VARIOUS DATA PRE-PROCESSING STEPS

The various data pre-preprocessing steps followed by the description of the 2 given approaches for the classification of music genre are as follows:-

4.1) Data Pre-processing

In the aim of improving the SNR(Signal to Noise Ratio) of the various signals, a pre-emphasis filter which is given by Equation 1 shown below ,this equation is applied to the original audio signal.

x(t) -x(t 1)=x(t)….(1),here in this equation, x(t) is being referred to the original signal, and the y(t) is being referred to the filtered signal and alpha is set to 0.97.

4.2)Deep neural networks

Nodes are little fragments of the system, and they are just like the functioning of neurons present in the human brain.A process takes place in these nodes,when a stimulus hits them.Some of them may be connected and also marked, and maybe some are not marked or connected, but generally, nodes are usually grouped into various layers.

The system should process the various layers of data in between the input and output in order to solve a task.

The usage of deep neural network could be found in various applications in real life. For example, a company which is Chinese naming Sensetime, created a system that consisted of automatic face recognition system which is used to identify criminals by using the real-time cameras in order to find an offender usually in the crowd. Nowadays, it has become a famous practice in the police and various other governmental entities.

4.3)Spectrogram Generation

A spectrogram is a two dimensional representation of a signal,that have a frequency on the y-axis and the time on the x-axis. A colormap is made use of for quantifying the magnitude of frequency which will be given within a given time window.

A spectrogram is usually a visual representation of the various frequencies containing spectrum of a signal as it changes with time. When it is applied to a signal of audio, spectrograms are infrequently called sonographs, voiceprints, or voicegrams. When the data are performed in a 3D plot they might be called as waterfalls.

Spectrograms are used widely in the fields of linguistics, music, sonar, radar, seismology, and speech processing. Spectrograms of audio could be made use to analyze spoken words phonetically.

Amplitude, time and frequency are the three dimensional information of a spectrogram.

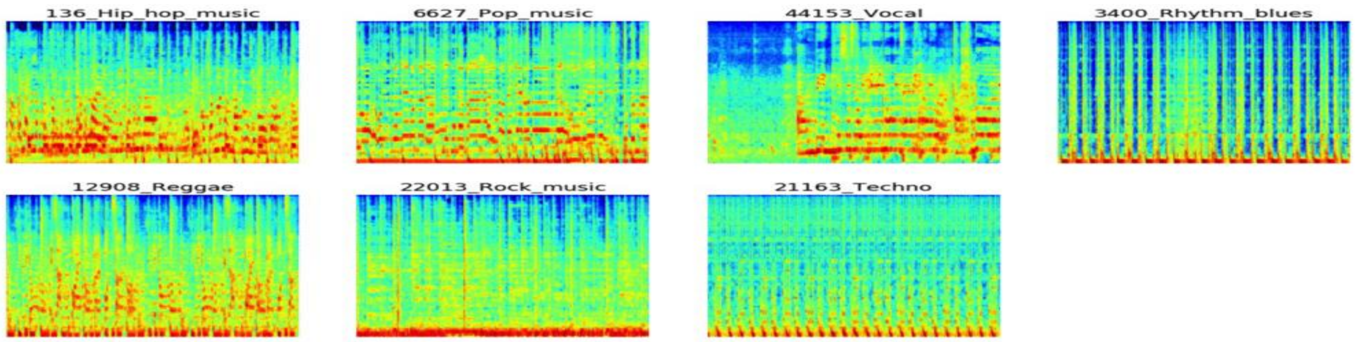


Figure 4.1:The sample spectrograms for one audio signal from every genre of music

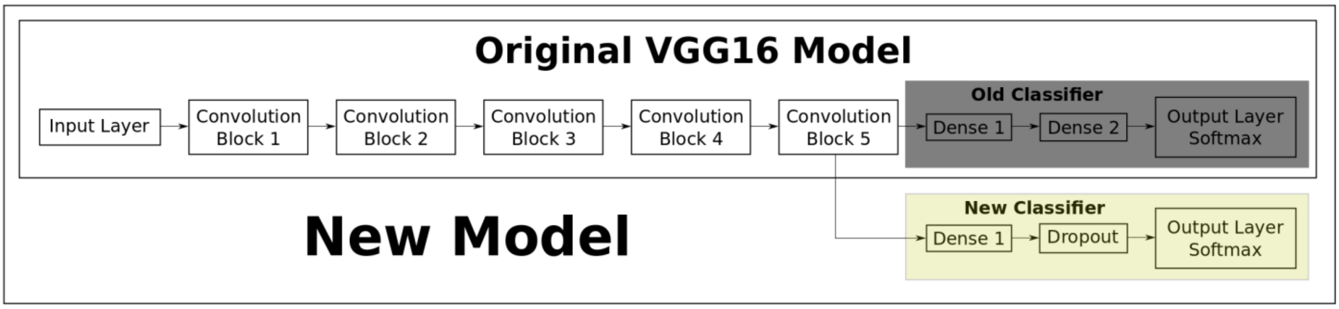


Figure 4.2:Convolutional neural network architecture

5. LITERATURE REVIEW

5.1.In the research paper of Music Genre Classification using Machine Learning techniques, In Hareesh Bahuleyan, (2018).The total work which was carried out by the author gives us an approach to categorize music automatedly by providing various tags to the songs that are present in the library of the user. It surveys both traditional method and Neural Method of using ML(Machine Learning) algorithms to achieve their aim. The Neural Method approach uses CNN(Convolutional Neural Network) which is directed from end to end using the Spectrograms (images) features of the signal of audio. The second approach makes use of many ML(Machine Learning)algorithms just like Logistic Regression, Random forest and many more, The physical features which are extracted like MFCC (Mel-Frequency Cepstral Coefficients),Chroma Features, Spectral Centroid and many more can be used to categorise the music into its various genres by using Machine Learning(ML) algorithms like the Logistic Regression, Random Forest, XGB (Gradient Boosting) and lastly the SVM (Support Vector Machines) .By making comparison of the two approaches isolatedly they came to a conclusion that model of VGG-16 CNN gave the most highest accuracy.

5.2. In the research paper Musical genre classification of audio signals, Tzanetakis G. et al., (2002). They have mostly explored about how the audio signals are automatically classified into a hierarchy of musical genres and how it can be done. They generally believe that those genres of music are categorical labels which are mainly created by humans only to categorise different pieces of music. They are classified using some of the most common characteristics. These characteristics are almost related to the different instruments that are used like the rhythmic structures, and mainly the harmonic music content. Genre hierarchies are mostly used to structure very huge music collections which are available on web. They have put forward three feature sets:the timbral texture, secondly the rhythmic content and lastly the pitch content. The investigation of put forward features which are done in order to examine the performance and the relative importance was usually done by training mostly the statistical pattern recognition classifiers by making utilisation of mostly real-world audio collections. Here, in this paper, both the real time frame-based and file classification schemes are been given a brief description. Using the put forward feature sets, this model can classify about 61% of 10 music genre correctly.

5.3.In the research paper of Content analysis for audio classification and segmentation, Lu L. et al., (2002).

They have proposed their study of classification and segmentation of the audio content analysis. Here, in this paper a stream of audio is divided according to speaker identity or the audio type. Their dominant approach is to construct a powerful model which will be efficient of dividing and categorising the audio signal that is specified into the speech , music, environment sound and silence. This classification is processed in two main steps, which have made it satisfactory for many various other applications as well. The first step is non- speech and speech discrimination.Here, a novel algorithm which is mainly based on linear spectral pairs-vector quantization (LSP-VQ) and on KNN (K- nearest- neighbour) have been finally developed. The second and last step is to divide the class of non-speech into music,environment sounds and silence with a classification of rule-based method. Here, they have made utilisation of new and many rare features like the noise frame ratio, the band periodicity which are not just mainly introduced, but are also discussed in much detail. They had also involved and produced a speaker segmentation algorithm that is mostly unsupervised and it most of the time uses a novel strategy that is centered on LSP correlation analysis and quasi.

5.4. In the research paper of Automatic musical pattern feature extraction using convolutional neural network, Tom LH Li et al., (2010):They made a huge effort which will be able to understand the important features which generally contribute to build the optimal model utilized for Music Genre Classification. The dominant reason of this research paper is to recommend a novel strategy to eliminate musical pattern characteristics of the audio file by making the usage of CNN(Convolutional Neural Network). Their experiments and results shows us that CNN(Convolutional Neural Network) have the most powerful capacity to catch informative features from the musical pattern which are varying. The various features that are extracted from the various audio clips like firstly the statistical spectral features, secondly the rhythmic content and lastly the pitch content are less dependable and produces very less accurate models. Therefore, in this approach which is made by them , the musical data have alike characteristics to that of image data and it requires mostly less previous knowledge. The dataset which was considered was GTZAN. It comprised of 10 genres with hundred audio clips each. Each audio clip is 30 seconds, sampling rate 22050 Hz at 16 bits. The musical patterns which were estimated using WEKA tool, here many classification models were considered. The classifier accuracy was eighty-four percent and in the end got higher. The accuracy can be expanded by parallel computing done on dissimilar combination of genres.

5.5. In the research paper Evaluation of the feature extractors and the psycho-acoustic transformations for music genre classification by Thomas Lidy and Andreas Rauber(2005): The authors here discussed the advantages of the psycho-acoustic features for recognization of music genre, mostly the significance of STFT taken at the Bark Scale, Mel-frequency cepstral coefficients (MFCCs) ,spectral contrast and spectral roll-off were several features utilised by the mixture of visual and acoustic features which are used to train the Support Vector Machine(SVM).

5.6.In the research paper of Musical genre classification of audio signals by George Tzanetakis and Perry Cook.(2002): They introduced three sets of features which will be used for this task (addressing to the the classification of music genre with the supervised machine learning approaches such as the Gaussian Mixture model and the KNN(k-nearest neighbour classifiers))classified as firstly the timbral structure,secondly the rhythmic content and lastly the pitch content. (HMMs)Hidden Markov Models,that has been tremendously used for the speech recognition tasks,these have also been explored for the classification of music genre.

5.7. In the research paper Convolutional neural networks for speech recognition, Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang,Gerald Penn, Dong Yu. 2014:- They discussed the most recent success of the deep neural networks,a number of studies had applied these techniques to speech and the various other forms of audio data.

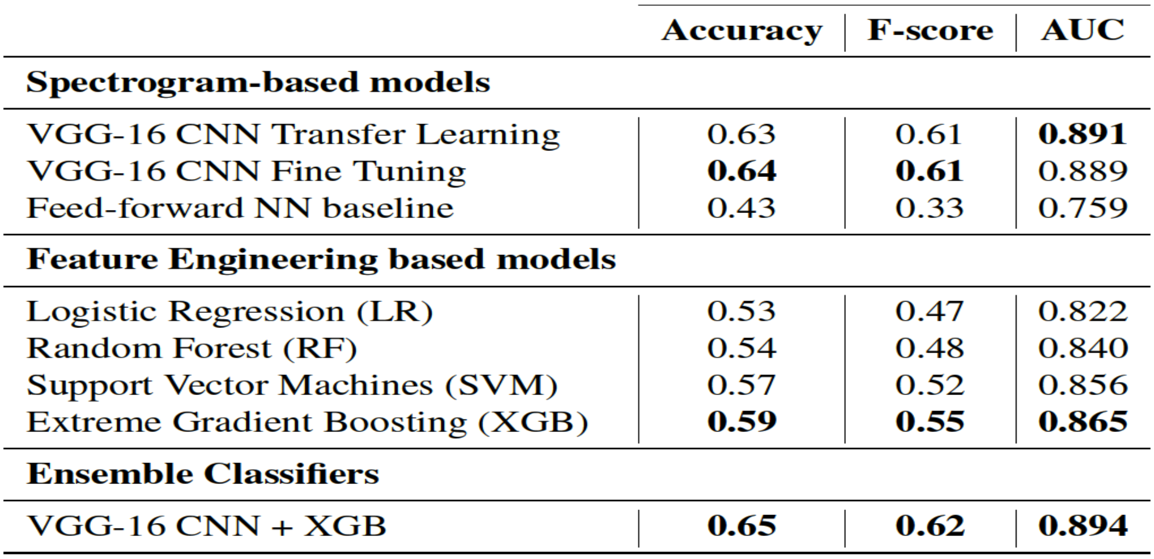
5.8.In the research paper of A generative model for raw audio , Aaron Van Den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Nal Kalchbrenner and Koray Kavukcuoglu. 2016., They gave a brief description of how the audio generation tasks will be carried out. A very common substitute representation is the spectrogram which is present in a signal that captures both of the frequency and time information. Spectrograms could be considered as images which are mainly used to train convolutional neural networks(CNNs).

5.9. In the research paper of Parallel convolutional neural networks for music genre and mood classification by Thomas Lidy and Alexander Schindler. 2016:They proposed a usual alternative representation,which is the spectrogram of a signal ,the spectrograms have the ability to capture both frequency and time information. Spectrograms could be considered as images that could be utilised to train mainly the convolutional neural networks .A CNN was basically developed to foresee the music genre with the use of the raw MFCC matrix as the main input in this research paper and a constant Q-transform.

6. RESULTS ANALYSIS AND FUTURE SCOPE

RESULT ANALYSIS

Confusion Matrix for the Extreme Gradient Boosting, relative importance of features in the XGBoost model; the top 20 most contributing features are displayed and the models are evaluated. The models are evaluated on the basis on firstly AUC, secondly accuracy and lastly F-score.



	Accuracy	F-score	AUC
Spectrogram-based models			
VGG-16 CNN Transfer Learning	0.63	0.61	**0.891**
VGG-16 CNN Fine Tuning	**0.64**	**0.61**	0.889
Feed-forward NN baseline	0.43	0.33	0.759
Feature Engineering based models			
Logistic Regression (LR)	0.53	0.47	0.822
Random Forest (RF)	0.54	0.48	0.840
Support Vector Machines (SVM)	0.57	0.52	0.856
Extreme Gradient Boosting (XGB)	**0.59**	**0.55**	**0.865**
Ensemble Classifiers			
VGG-16 CNN + XGB	**0.65**	**0.62**	**0.894**

FIGURE 6.1:The models are evaluated on the basis on AUC, accuracy and F-score.

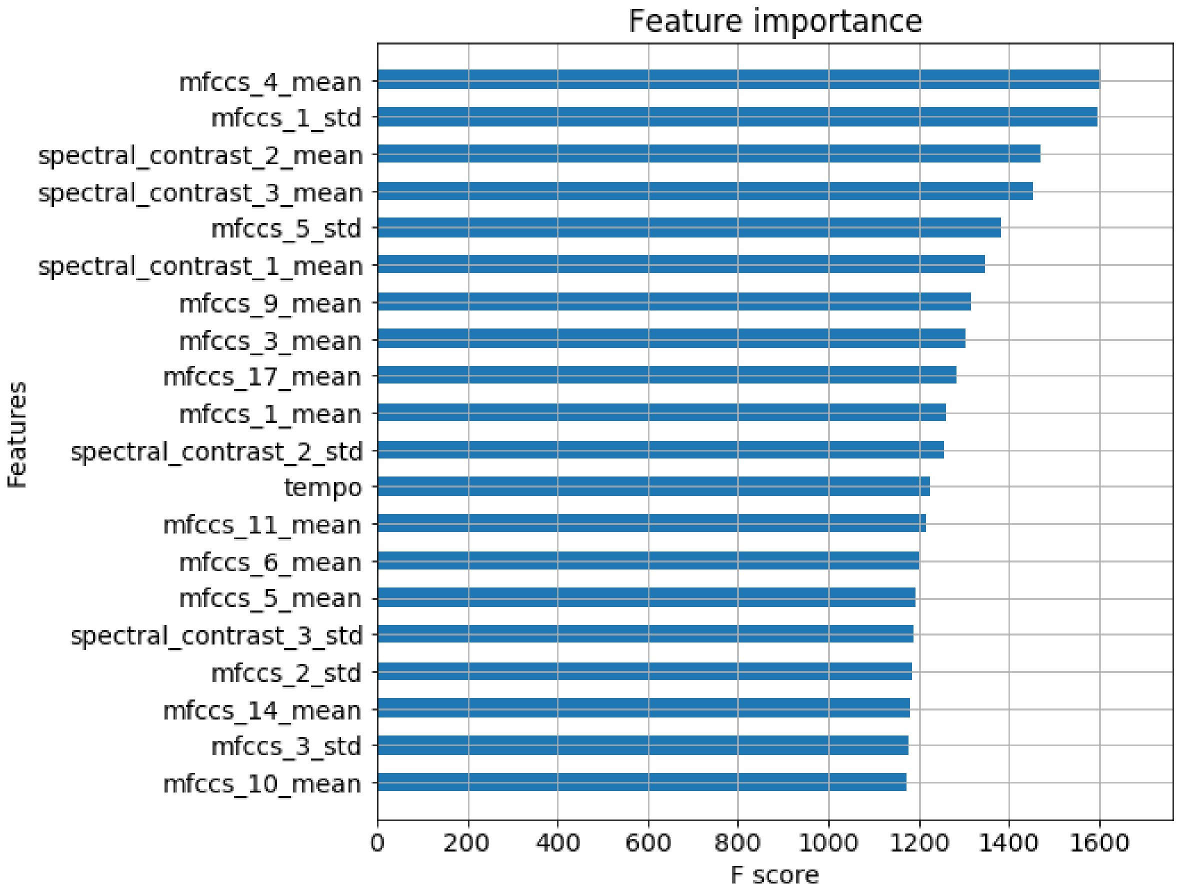
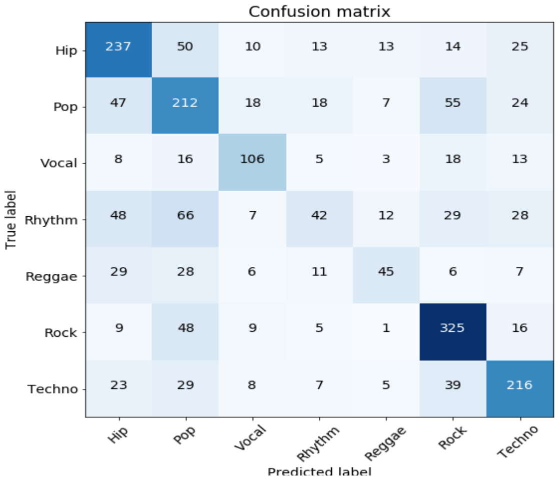


Figure 6.2: Relative importance of the features present in XGBoost model, in this figure top twenty best contributing features are shown



	Hip	Pop	Vocal	Rhythm	Reggae	Rock	Techno
Hip	237	50	10	13	13	14	25
Pop	47	212	18	18	7	55	24
Vocal	8	16	106	5	3	18	13
Rhythm	48	66	7	42	12	29	28
Reggae	29	28	6	11	45	6	7
Rock	9	48	9	5	1	325	16
Techno	23	29	8	7	5	39	216

FIGURE 6.3:Confusion Matrix for Extreme Gradient Boosting

FUTURE SCOPE

Following are some of the future scopes I think can be listed for now:

1)Music genre classifier is used for predicting the genre of a particular piece of music which is in the format of audio. These devices could be used for tasks like the automatically tagging of music for distributors such as Spotify and Billboard and determining appropriate background music for events.

2) If enough audio data is given, out of which many large amounts could be easily gathered from mostly freely available music online, machine learning could find it and observe and then make predictions using ill-defined patterns.

3)Music genre classification can be used to produce the confusion matrix for algorithms like logistic regression.

4)Music genre classification can be used to compare the accuracy of different machine learning models.

5)Music genre classification can be used to compare the f-score of various machine learning algorithms

7. CONCLUSION

Deep learning approach was the first approach and for the purpose of predicting the genre label of a signal ,a CNN model was trained usually from end-to-end, spectrogram was used to carry out this process. Hand-crafted features were used in the second approach which was also the last approach. Four out of traditional machine learning classifiers were trained beside these features and their performance was compared afterwards. Identification of features which would help the classification of this task was performed. A set data of Audio was used for the experiments and after that an AUC value of 0:894 was reported for an ensemble classifier which merged the two proposed methods. The databases of online music are growing day by day and nowadays it is very simple to access online music.

Music has nowadays become a significant proportion of Internet content: the most important source of pieces of music is the net. In this context, the automatic procedures which are able to deal with very huge amounts of quantities in formats such as digital formats are salient. Labels of music genre are very essential categories which are used to categorise and classify the songs, albums, and artists into broader groups which can share alike musical characteristics. The classification of music files in accordance with their genre has become a demanding task in the zone of mainly music information retrieval (MIR).

REFERENCES

[1]Anirudh Ghidiyal, Komal Singh and Sachin Sharma, Music Genre Classification Using Machine Learning, 2020.

[2]Tzanetakis, Musical genre classification of the various audio signal,2002.

[3]Nitin and Punam Singh, Confusion Matrix in Machine Learning,2021.[online]Available: https://www.geeksforgeeks.org/confusion-matrix-machine-learning/

[4]Akhand Pratap Mishra, Introduction to Convolutional Neural Network,2021.[online]Available: https://www.geeksforgeeks.org/introduction-convolution-neural-network/

[5]Ahmet Elbir, Hilmi Bilal Cam,Mehmet Emre Iyican and Berkay Ozturk,Music Genre Classification and Recommendation Using Machine Learning Techniques,2018

[6] George Tzanetakis and Perry Cook., Musical genre classification of the different audio signals,2002.

[7] Thomas Lidy and Alexander Schindler, Parallel convolutional neural networks for the music genre,2016.

[8] Sawan Rai, Music Genre Classification Using Deep Learning Techniques,2021.[online] Available: https://www.analyticsvidhya.com/blog/2021/06/music-genres-classification-using-deep-learning-techniques/

[9]Parul Pandey, Music Genre Classification with Python,2018. [online]Available: https://towardsdatascience.com/music-genre-classification-with-python-c714d032f0d8

[10] Rajeeva Shreedhara Bhat,Rohit B. R.,Mamatha K. R.,"Music Genre Classification", 2020.

[11] Corinna Cortes and Vladimir Vapnik. 1995. Support vector networks. Machine learning 20(3):273–297.

[12] Pawan,XGBoost,2021.[online]Available: https://www.geeksforgeeks.org/xgboost/

[13] Trevor Hastie, Robert Tibshirani, and Jerome Fried-man," The elements of statistical learning", 2001

[14] Hagen Soltau, Tanja Schultz, Martin Westphal, and Alex Waibel,"Recognition of music types", 1998

[15] Yandre M.G.Costa, Luiz S.Oliveira, Carlos N.Silla Jr., "An evaluation of Convolutional Neural Networks for music classification using spectrograms", Applied Soft Computing, Volume 52, March 2017, Pages 28-38

[16] Elizabeth Nurmiyati Tamatjita, Aditya Wikan Mahastama, "Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours", 2016 International Conference on Information Management and Technology (ICIMTech), 18 May 2017

[17] Machine Learning GeeksforGeek - https://www.geeksforgeeks.org/machine-learning/

[18] Tzanetakis, G., & Cook, P, Musical genre classification of audio signals, 2002.

[19]Leland Roberts, Musical Genre Classification with Convolutional Neural Networks,2021.[online]Available:https://towardsdatascience.com/musical-genre-classification-with-convolutionalneural-networks-ff04f9601a74

[20]Leland Roberts, Understanding the Mel Spectrogram, 2020.[online] Available: https://medium.com/analyticsvidhya/understanding-the-mel-spectrogram-fca2afa2ce53

[21] Convolutional Neural Network Tutorial by Simplilearn,[online]Available: https://www.simplilearn.com/tutorials/deep-learningtutorial/convolutional-neural-network

[22]Plotting a Spectrogram using Python and Matplotlib, Zaware Sumedha,2021.[Online]. Available:

https://www.geeksforgeeks.org/plotting-a-spectrogram-using-python-and-matplotlib/

[23]Alokesh, Introduction to Support Vector Machines, 2020.[online]Available: https://www.geeksforgeeks.org/introduction-to-support-vector-machines-svm/

[24] Anannya Uberoi,K-Nearest Neighbours,2021.[online]Available: https://www.geeksforgeeks.org/k-nearest-neighbours/

[25]Ali Karatana and Oktay Yidiz, Music Genre Classification with Machine Learning Techniques, 2017