TABLE OF CONTENTS

ABBREVIATIONS

TABLE LIST

FIGURE LIST

SUMMARY

Today,music production has increased greatly with the development of technology,and tens of thousands of new music are released every year. The consumption of these music has also become easier and faster. With the increase in the use of mobile devices, our habits of listening to music have changed and new applications have occurred. It is known that there are an estimated 1300 music genres. For this reason, classifying the music in terms of genre becomes a current problem.

Throughout the project, suitable data sets were researched to solve this problem. It is quite difficult to create or find a proper data set in which the music has no copyrights. For this reason, GTZAN data set, which we frequently come across in literature research, was used. CNN models that classify 10 different music genres have been designed and the model that gives the best results is described in the report. For the best 2000 samples, an accuracy rate of 67% was obtained for the test data.

ÖZET

Günümüzde teknolojinin gelişmesiyle birlikte müzik üretimi büyük oranda arttı, her yıl on binlerce yeni müzik çıkmaktadır. Bu müziklerin tüketimi de aynı şekilde kolaylaşmış ve hızlanmıştır. Mobil cihazların kullanım oranının artmasıyla birlikte müzik dinleme alışkanlıklarımız değişmiş, yeni uygulamalar popüler uygulamalar çıkmıştır. Tahmini olarak 1300 tane müzik türü olduğu bilinmektedir. Bu sebeple dinlediğimiz müziklerin tür bakımından sınıflandırılması güncel bir problem haline gelmektedir.

Proje boyunca, bu problemi çözmek için uygun veri setleri araştırılmıştır. Müziklerin isim hakları olduğu uygun veri setini oluşturmak oldukça zordur. Bu sebeple literatür araştırmasında sıklıkla rastladığımız GTZAN veri seti kullanılmıştır. 10 Farklı müzik türünü sınıflandıran CNN modelleri tasarlanmış ve en iyi sonuç veren model raporda anlatılmıştır. En iyi 2000 örnek içerek test verisi için %67 doğruluk oranı elde edilmiştir.

1. INTRODUCTION

Sound is one of the most fundamental physical properties that humankind interacts with. Throughout mankind, humans interact with sound in different ways, they imitate animal sounds, they use sounds to warn themselves they used it to find resources or sounds could be used for religious rituals. However, nowadays music is the first thing that comes to our minds when the topic is sound. Music also can contain lots of property inside of it. It is shaped in many different ways; notes, tempos, rhythms, melody, harmony are some of them. These features give direction to music in various ways, and these forms are also called musical genres. According to Spotify, there are over 1,300 music genres in the world[3]. Thus, it is a big and beneficial challenge for data scientists or machine learning enthusiasts for classifying music genres. In this project, it is tried to classify 10 fundamental music genres.

1.1 AIM OF THE PROJECT

Today, music consumption has increased with the increase of the internet and music applications. Within the scope of the project, we tried to classify music genres using artificial neural networks.Within the scope of the project, we tried to classify music genres using artificial neural networks. The music genres to be classified were determined as "blues", "classical", "country", "disco", "hiphop", "jazz", "metal", "pop", "reggae", "rock". The end product can be used as a sub-product in popular music applications.

1.2 LITERATURE RESEARCH

Through searching literature, we found that the GTZAN dataset [1] is used to classify music genres. According to Dong, humans achieve %70 accuracies in the same task [2]. The spectrogram is a 2D representation of speech signals. Since spectrograms have both frequency and time information, it is commonly used in audio related tasks. In the paper[2], the author used CNN to classify 10 different genres and achieved 70%.Rather than directly using spectrograms, the long-term statistical distribution of short-time features MFCC, spectrogram bandwidth, and zero-crossing rate was decided to use [3-4]. Although [3] achieves %88.9 with SVM and %85.6 with neural network, the paper only classifies six genres and one of the most important genre (rock) is not classified within the paper.

2. DATA SET

Finding a music data set was a difficult process due to copyrights. We searched Kaggle for datasets. First, we found the Spotify dataset; however, the dataset does not consist of audio data; it consists of some sound features such as acousticness, danceability, etc. for each song. After that, it was decided to make our own dataset, however, Spotify does not permit downloading songs so we decided to use youtube for getting audio data. We created an algorithm which is searching genre names on Spotify and finds a playlist that consists of chosen genres and searches these songs on youtube again and downloads it. Nevertheless, the problem occurs with the youtube search query, there is a daily limit for searching queries.

Finally, we found a free music data set which is called GTZAN[1] on Kaggle. The dataset consists of 10 genres and it has 100 songs for each genre for 30 seconds. Additionally, each song was sampled at 22050Hz. Moreover, the dataset is used extensively in researches that are made on this topic. The data set both have audio statistical spectrogram features such as variance of chroma stft, mean of mfcc features and zero crossing rate and have mel spectrograms as well as raw audio data. Frequency plot of chosen song shown in Figure 2.1. Spectrogram and Mel spectrogram of the chosen song shown in Figure 2.2 and Figure 2.3 respectively.

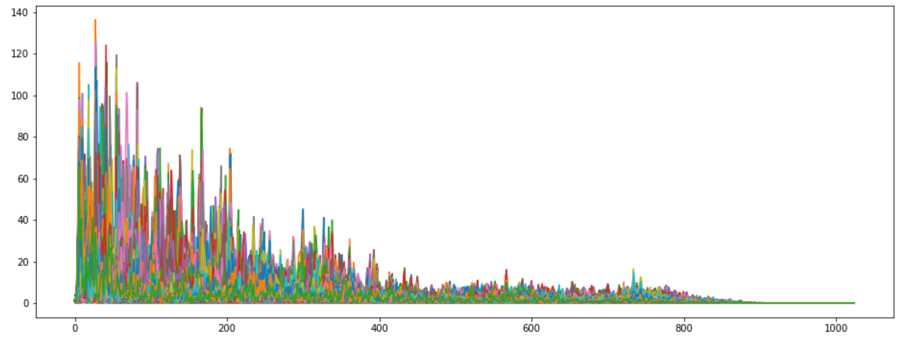


Figure 2.1: Frequency plot of the chosen song from GTZAN Dataset

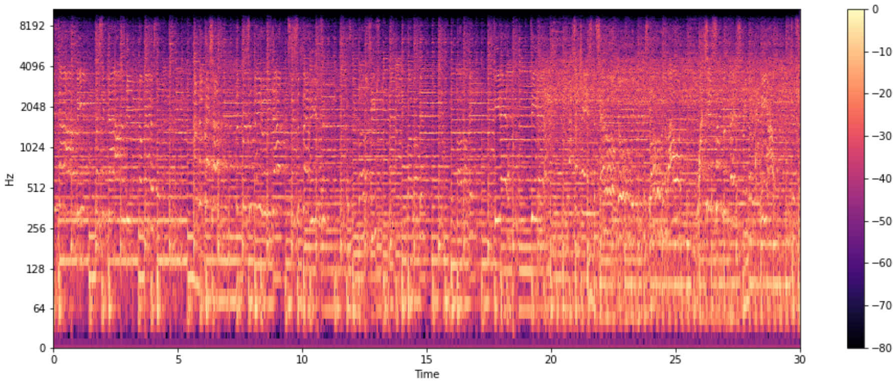


Figure 2.2: Spectrogram of the chosen song from GTZAN Dataset

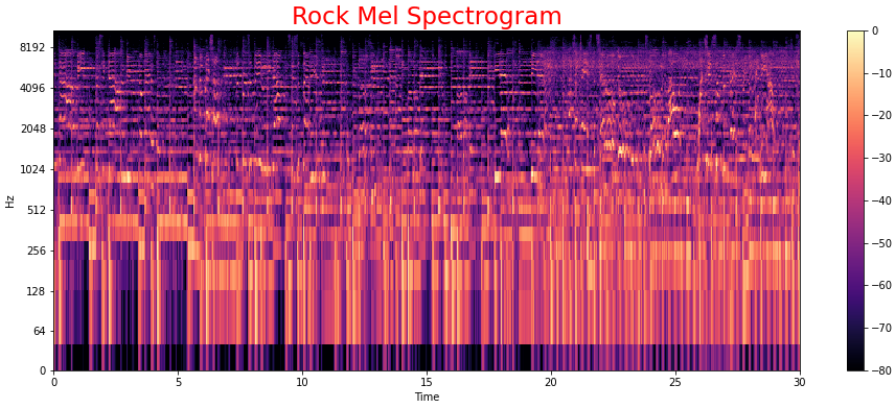
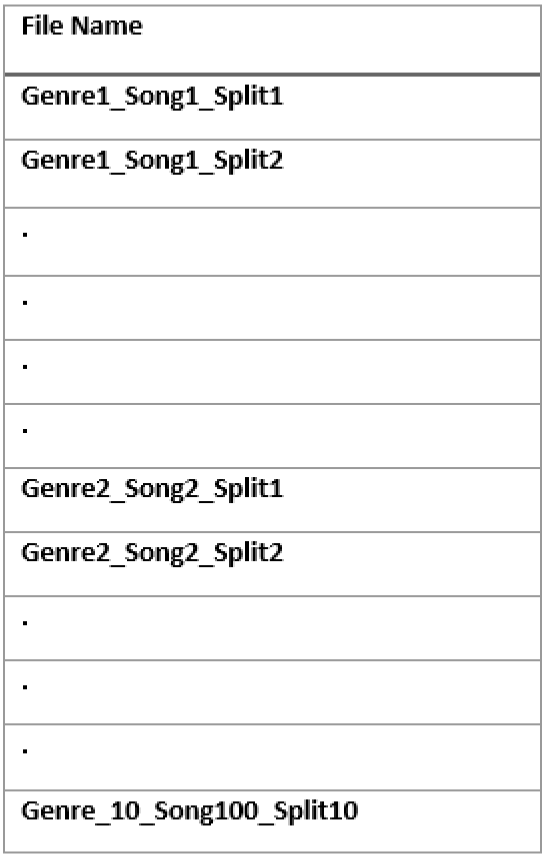


Figure 2.3 :Mel spectrogram of chosen song from GTZAN Dataset

3. PRE PROCESSING

For the proposed model, we decided to use a mel spectrogram for the input instead of statistical information of spectrograms . In the GTZAN dataset, every music genre has 100 song examples. In order to increase the dataset, we split every song with 3 seconds intervals. The way the data set is stored is shown in Figure 3.1



File Name
Genre1_Song1_Split1
Genre1_Song1_Split2
.
.
.
.
Genre2_Song2_Split1
Genre2_Song2_Split2
.
.
.
Genre_10_Song100_Split10

Figure 3.1: The way data set is stored

In order to prevent mixing the training and validation data sets, the data set was splitted manually . Through literature research, it was found that many projects split the data set wrongly thus it resulted in unrealistic accuracies.

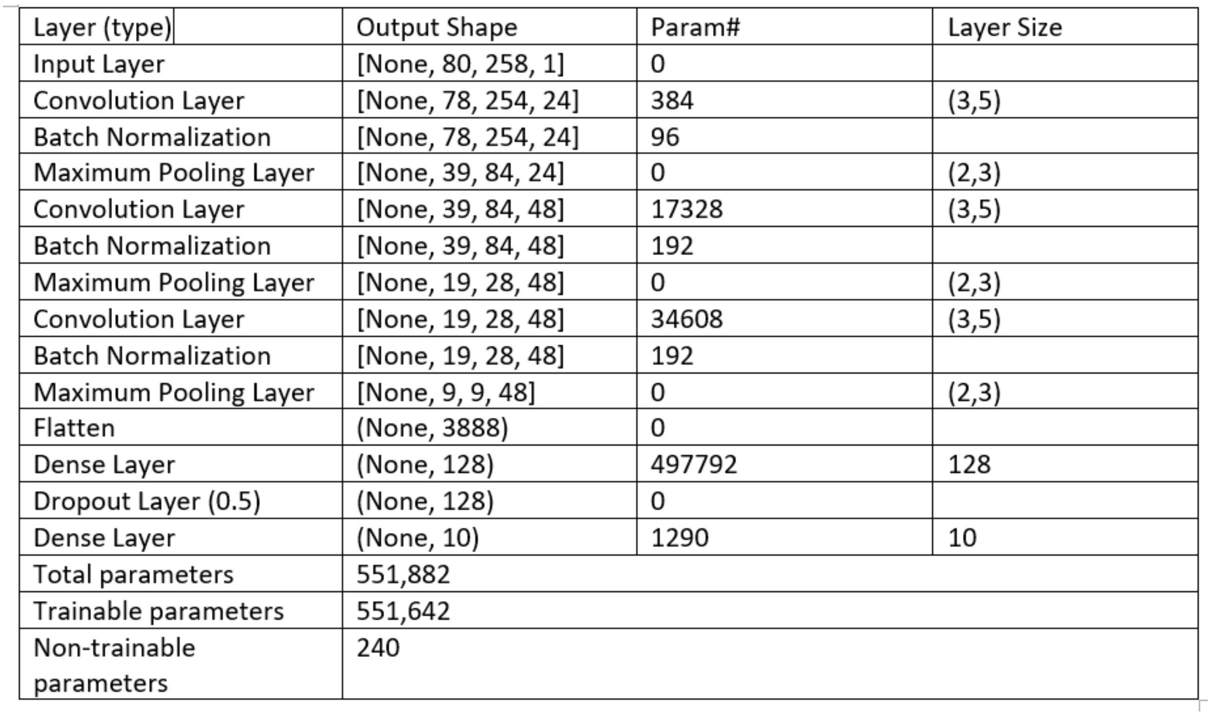
While calculating the Mel spectrograms,512 sample fft size (24 ms) and 256 sample hop length (12 ms) were used. Moreover, 80 mel filters were used. As a result,(80,258) shaped images are obtained.

4. PROPOSED MODEL AND RESULTS

We used CNN (convolutional neural network) to classify music genres. It is known that CNN models are good at pattern recognition problems since they preserve both time and frequency data. With STFT, we can represent audio signals as two dimensional image data. Hence, CNN models can be used in order to classify music genres. Since [2] uses CNN and spectrograms, we decided to use CNN. Our model architecture shown in Table 4.1.

In our first attempt, we split the training, validation and test data with %60,%20,%20 respectively. Since we don't have a big dataset, %96.8 training and %61.45 validation accuracies were achieved. It is clear that our model has overfitting issues. Although we used methods such as dropout, kernel regularizers and batch normalization in order to prevent overfitting we could not avoid overfitting as a result of small data set especially for music genre classification problems. Thus we decided to split the data set with %80,%20 in order to increase training data size. Training and validation accuracy shown in Figure 4.1. Training and validation loss is shown in Figure 4.2. Confusion matrices are shown in Figure 4.3 and Figure 4.4, as expected the model has difficulty to classify. The model classifies “metal” music with %91 accuracy, it makes sense since metal songs have a much bigger power spectrum compared to other genres. The model also classifies “rap” music with quality mostly because rap music has a high tempo rate compared to other genres. As it is seen in Figure 4.3, %19 of pop and %26 portion of reggae music classified as hiphop music. It makes sense since these two genres have close tempo rates like rap music.
Moreover, while classifying rock songs our model reaches %38 accuracies with rock and %37 accuracies with blues, so as it is understood there is confusion between them. The reason for that is rock kinds of music generally composed with blues music scale.

Table 4.1: Model Architecture



Layer (type)	Output Shape	Param#	Layer Size
Input Layer	[None, 80, 258, 1]	0	
Convolution Layer	[None, 78, 254, 24]	384	(3,5)
Batch Normalization	[None, 78, 254, 24]	96	
Maximum Pooling Layer	[None, 39, 84, 24]	0	(2,3)
Convolution Layer	[None, 39, 84, 48]	17328	(3,5)
Batch Normalization	[None, 39, 84, 48]	192	
Maximum Pooling Layer	[None, 19, 28, 48]	0	(2,3)
Convolution Layer	[None, 19, 28, 48]	34608	(3,5)
Batch Normalization	[None, 19, 28, 48]	192	
Maximum Pooling Layer	[None, 9, 9, 48]	0	(2,3)
Flatten	(None, 3888)	0	
Dense Layer	(None, 128)	497792	128
Dropout Layer (0.5)	(None, 128)	0	
Dense Layer	(None, 10)	1290	10
Total parameters	551,882		
Trainable parameters	551,642		
Non-trainable parameters	240		

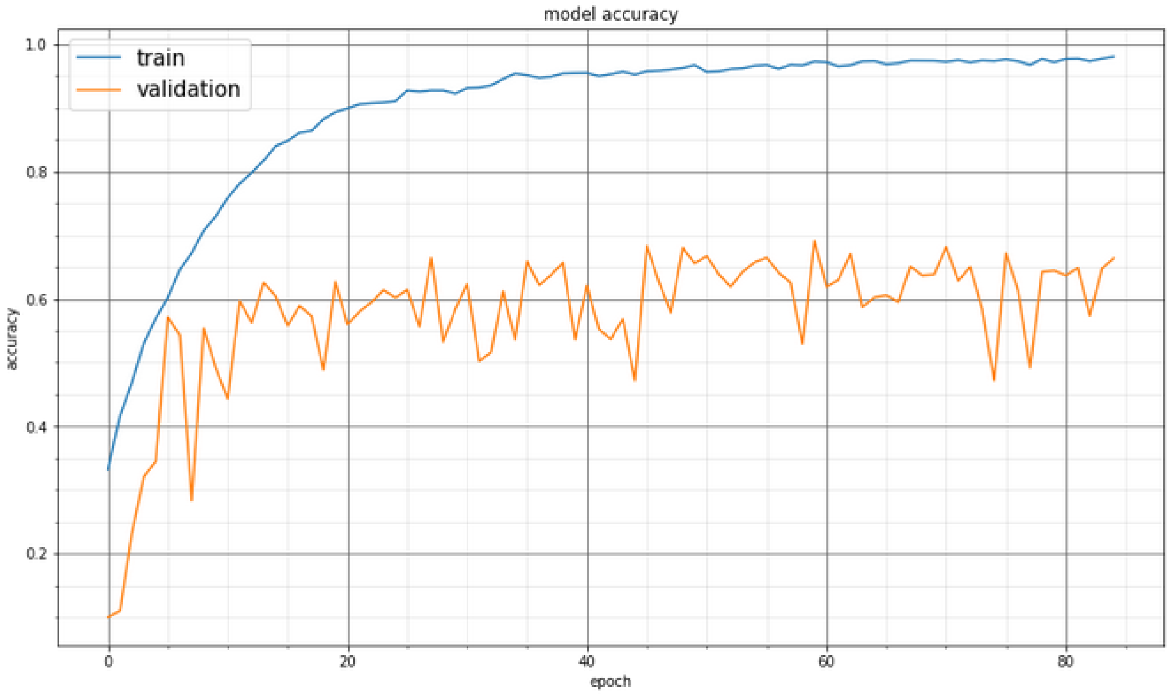


Figure 4.1: Training and validation accuracy for Model-A

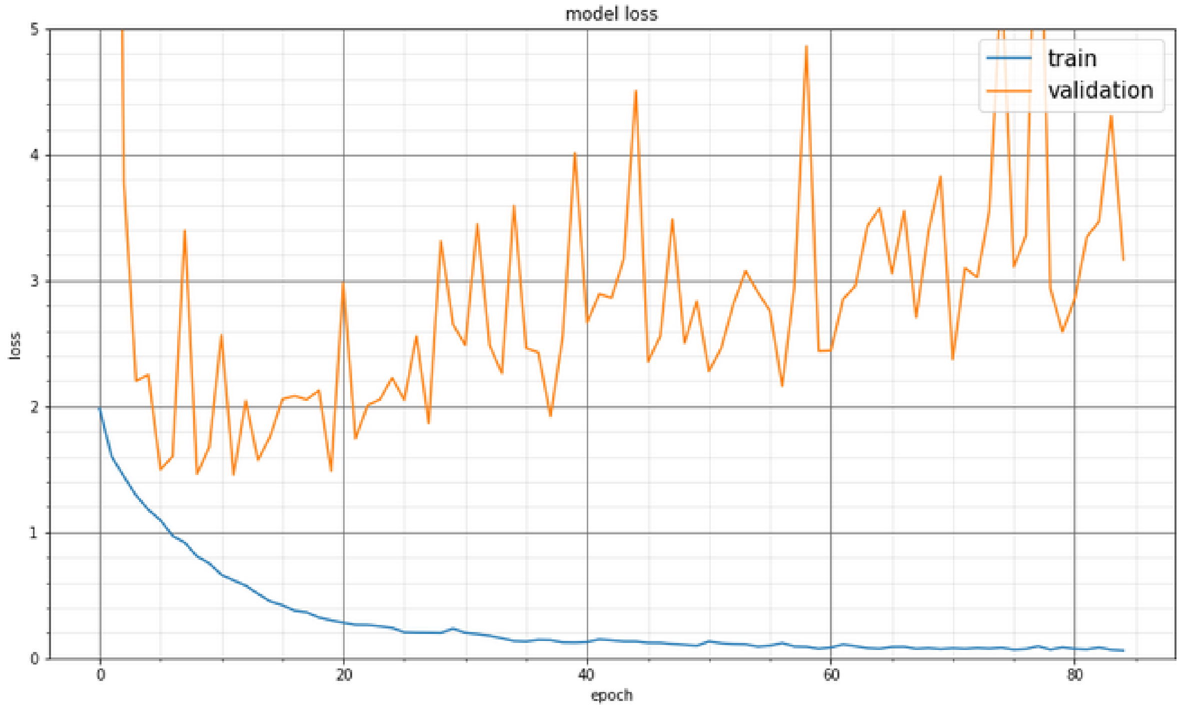
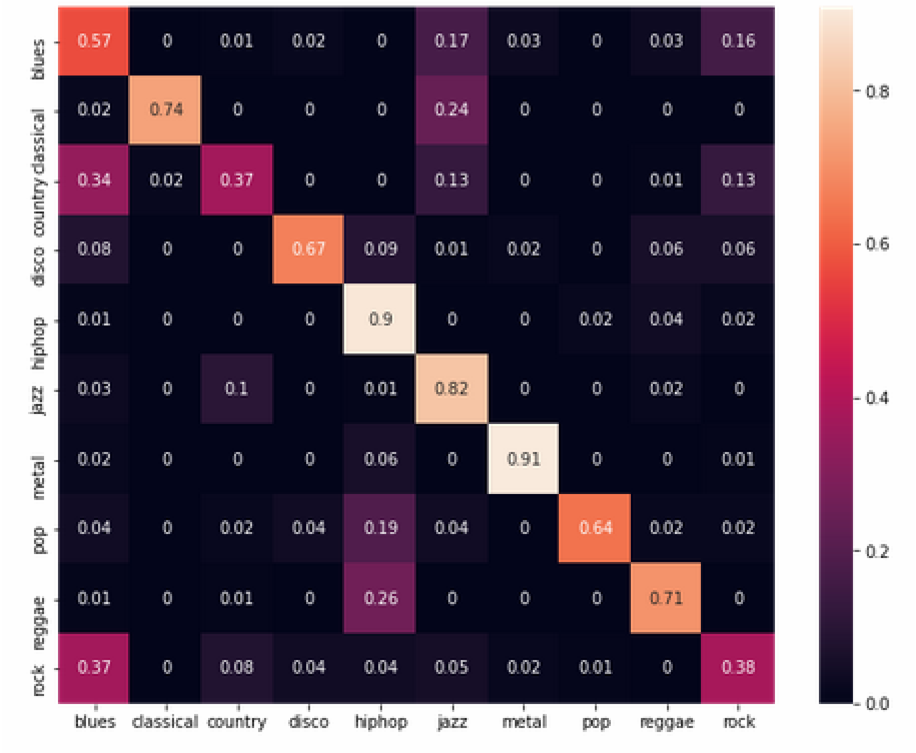


Figure 4.2: Training and validation loss for Model-A



	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	0.57	0	0.01	0.02	0	0.17	0.03	0	0.03	0.16
classical	0.02	0.74	0	0	0	0.24	0	0	0	0
country	0.34	0.02	0.37	0	0	0.13	0	0	0.01	0.13
disco	0.08	0	0	0.67	0.09	0.01	0.02	0	0.06	0.06
hiphop	0.01	0	0	0	0.9	0	0	0.02	0.04	0.02
jazz	0.03	0	0.1	0	0.01	0.82	0	0	0.02	0
metal	0.02	0	0	0	0.06	0	0.91	0	0	0.01
pop	0.04	0	0.02	0.04	0.19	0.04	0	0.64	0.02	0.02
reggae	0.01	0	0.01	0	0.26	0	0	0	0.71	0
rock	0.37	0	0.08	0.04	0.04	0.05	0.02	0.01	0	0.38

Figure 4.3: Confusion matrix with ratios

5.

6. SUGGESTIONS

Through our project, data set size is one of the biggest problems that is dealt with and for this kind of problem more songs with bigger window size is needed. In our first attempt, we split the data set with (0.6,0.2,0.2) ratio and achieved %61 accuracy; in the second attempt, we split the data set with (.8,.2) ratio and achieved %69 accuracy.

The model is small enough to run in mobile devices, one can build an application to classify music genres with the trained model.

RESOURCES

[1] Data Sets. (n.d.). Retrieved December 20, 2020, from http://marsyas.info/downloads/datasets.html

[2] Dong, M. (2018, February 27). *Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification*. ArXiv.Org. https://arxiv.org/abs/1802.09697

[3] Yildiz, Oktay & Karatana, Ali. (2017). Music genre classification with machine learning techniques. 10.1109/SIU.2017.7960694.