Introduction

Music genres allow to categorize musical items that share common characteristics. Modern Music Archives consists of Tens of Thousands of songs which are not ordered properly, hence a proper genre into which these songs can be classified into gives us a proper arrangement of the songs and provides for easier searching of the song name using tune or beats by Apps such as Shazam. During creation of song the song composer does not always make a music specific to one genre, the composer at some point in composing the song may feel like changing the pace of the song for aesthetics. Bohemian Rapsody, a well-known song has four sections piano ballad, Capella, opera and hard rock, classifying such song into any of these genres using mere audio is not so easy.

Literature review 1:

***Improved Music Genre Classification with Convolutional Neural Networks***

Weibin Zhang ∗,Wenkang Lei ∗, Xiangmin Xu, Xiaofeng Xing

In this paper, we proposed two ways to improve music genre classification with convolutional neural networks (CNN): 1) combining max- and average pooling to provide more statistical information to higher level neural networks; 2) using shortcut connections to skip one or more layers, a method inspired by residual learning method. The dataset used for the experimentation and testing is GTZAN dataset consisting of about a thousand songs and 10 major genres.

The input for both the nets is in the form of STFT spectrogram.

*Network with both max and average polling(nnet1)*: This network consists of 10 layers including input and soft max output layer.

*The Residual Network(nnet2):* It is similar to nnet1. The biggest difference between the two networks is the shortcut connections from the output of the first convolutional layer to the output of the third convolutional layer.

In the end the results were very satisfactory with nnet1 giving the accuracy of 84.8% and nnet2 giving the accuracy of 87.4%.

Literature review 2:

***MULTI-LABEL MUSIC GENRE CLASSIFICATION FROM AUDIO, TEXT AND IMAGES USING DEEP FEATURES***

Sergio Oramas, Oriol Nieto, Francesco Barbieri, Xavier Serra

The music genres are not mutually exclusive, in spite of that most related research is traditionally focused on classifying tracks into a single class. Furthermore, these categories (e.g., Pop, Rock) tend to be too broad for certain applications. In this work the aim is to expand this task by categorizing musical items into multiple and fine-grained labels, using three different data modalities: audio, text, and images.

The dataset used is MuMu, a dataset of consisting of more than 31k albums classified into 250 genre classes.

*Audio classification:* To learn the genre labels a CNN with four convolutional layers is designed and experiment with different number of filters, filter sizes, and output configurations is performed.

*Text-based approach:* For this a combination of work done by [1] & [2] is applied to a feed forward network with two dense layers of 2048 neurons and a Rectified Linear Unit (ReLU) after each layer is trained to predict the genre labels in both LOGISTIC and COSINE configurations.

*Image-based approach:* To perform music genre classification from the images, we use Deep Residual Networks (ResNets) [3].

*Multimodal approach:* From each modality network described above, we separately obtain an internal feature representation for every album after training them on the genre classification task. They are then concatenated into a single feature vector, which becomes the input to a simple Multi-Layer Perceptron (MLP), where the input layer is directly connected to the output layer. The output layer may have either a LOGISTIC or a COSINE configuration.

Literature review 3:

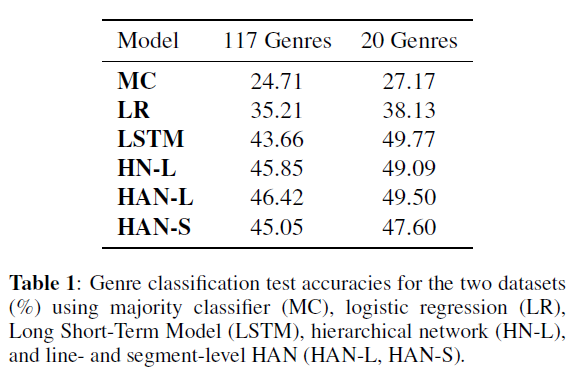
***LYRICS-BASED MUSIC GENRE CLASSIFICATION USING A HIERARCHICAL ATTENTION NETWORK***

Alexandros Tsaptsinos

Music genre classification, especially using lyrics alone, remains a challenging topic in Music Information Retrieval. In this study they apply recurrent neural network models to classify a large dataset of intact song lyrics. As lyrics exhibit a hierarchical layer structure a hierarchical attention network (HAN) is adopted to exploit these layers and in addition learn the importance of the words, lines, and segments.

A dataset of collection of lyrics was obtained through a signed research agreement with LyricFind.

HAN is compared with many baseline models. Table 1 shows the results in percentage.



The performance of genre classification based on Lyric alone does not provide very good accuracy so performing a multimodal modal with the said algorithm and training the multimodal model might provide better results.

Literature review 4:

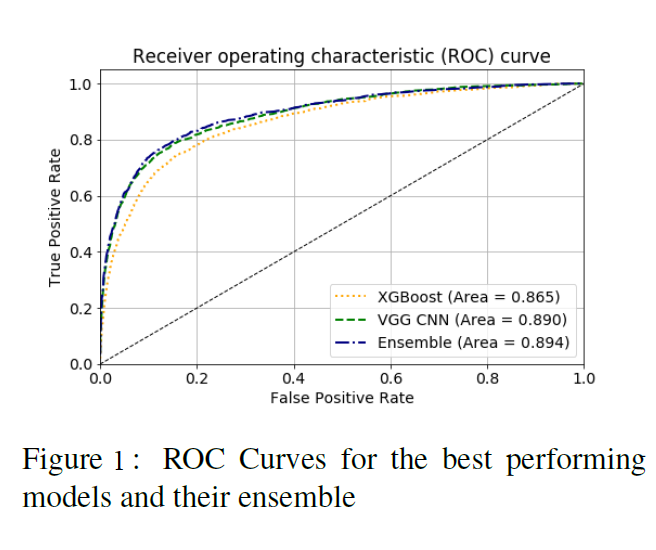
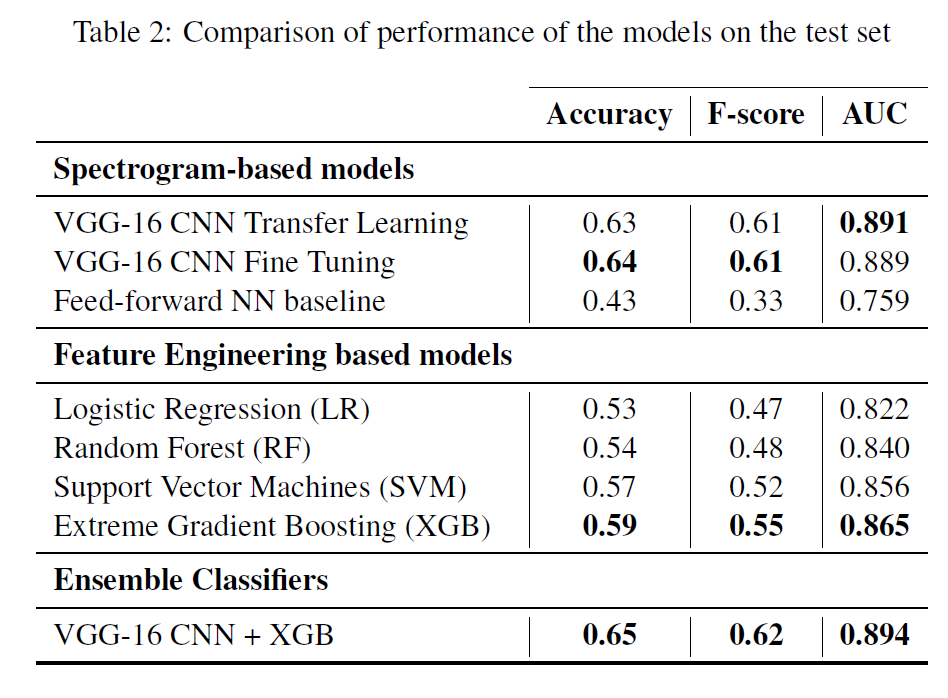
***Music Genre Classification using Machine Learning Techniques***

Hareesh Bahuleyan

In this paper a comparison in the performance of two classes of models is performed first is a CNN model trained end-to-end, to predict genre label of an audio signal only from spectrogram. Second approach utilizes hand-crafted features, both from the time domain and frequency domain. Four traditional machine learning classifiers are also trained to compare the performance. The experiments report an AUC value of .894 for an ensemble classifier which combines the two approaches.

The dataset used is *Audio Set.* It is a large-scale human annotated database of sounds.

The following table 2 shows the comparison of the models on the test data set, fig 1 shows the results for the ensemble approach.



**References:**

[1] Andrea Moro, Alessandro Raganato, and Roberto Navigli. Entity Linking meets Word Sense Disambiguation: A Unified Approach. Transactions of the Association for Computational Linguistics, 2:231–244, 2014.

[2] Sergio Oramas, Luis Espinosa-Anke, Aonghus Lawlor, et al. Exploring customer reviews for music genre classification and evolutionary studies. In ISMIR, 2016.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.