Assessment 1: Research Proposal

Automated Categorization of Music Archives using Machine Learning

|  |  |
| --- | --- |
| Name: | Supervisor: |
| University ID: |  |
| Word Count: 2378 (Without References) |  |
| Date: |  |

Table of Contents

[Abstract 2](#_Toc69506452)

[Introduction 2](#_Toc69506453)

[Background 2](#_Toc69506454)

[Aim/Objective 2](#_Toc69506455)

[Research Questions 3](#_Toc69506456)

[Ethical Considerations 3](#_Toc69506457)

[Literature Review 3](#_Toc69506458)

[CONVOLUTIONAL RECURRENT NEURAL NETWORKS FOR MUSIC CLASSIFICATION 3](#_Toc69506459)

[CROSS-COLLECTION EVALUATION FOR MUSIC CLASSIFICATION TASKS 4](#_Toc69506460)

[An Evaluation of Convolutional Neural Networks for Music Classification Using Spectrograms 6](#_Toc69506461)

[Project Timeline 9](#_Toc69506462)

[Methodology 10](#_Toc69506463)

[Project Evaluation 10](#_Toc69506464)

[Conclusion 10](#_Toc69506465)

[References 10](#_Toc69506466)

# Table of Figures

[Figure 1 - Block Diagrams of the ConvNets being Compared Together 4](#_Toc69506491)

[Figure 2 – Results 5](#_Toc69506492)

[Figure 3 - Cross-validation Results 5](#_Toc69506493)

[Figure 4 - Mapping between Classifier Classes and Ground Truths 6](#_Toc69506494)

[Figure 5 - Cross-collection accuracies for all involved classifiers. 6](#_Toc69506495)

[Figure 6 - Methodology Overview 7](#_Toc69506496)

[Figure 7 - ISMIR 2004 Music Dataset Features 7](#_Toc69506497)

[Figure 8 - Stats of the African Music Collection 8](#_Toc69506498)

[Figure 9 - Working of the ConvNet 8](#_Toc69506499)

[Figure 10 – Results 9](#_Toc69506500)

[Figure 11 - Project Timeline 9](#_Toc69506501)

# Abstract

Music Categorization has been around for many years and many a researcher have taken a crack at it with excellent results. Most of the approaches to Music Categorization are deep-learning oriented that involve the implementation of Convolutional Neural Networks. Music is a dataset in and of itself. The complexity of the data, the huge abode of features like frequency distribution, harmonics, amplitude and wavelength, duration and many more makes it a competitive dataset.

To realize all of these features properly and use them to properly categorize the Music can only be done using the power of deep learning. As of late, CNNs have been joined with recurrent neural networks (RNNs) which are regularly used to show consecutive information, for example, sound signals or word arrangements. This half-breed model is known as a convolutional recurrent neural network (CRNN).

The importance of music classification arises from the intensity with which media is being created and consumed. People today need to hear what they want without being forced to search for it. They want to hear songs that are similar to what they like but something new they’ve never heard before. Machine learning deals with this problem by assigning human music tastes with genres. Classifying music with genre helps software around the world cater to the picky needs of the public. This research aims to provide a machine learning ensemble that can classify music based on Genre.

**Keywords:** Music Categorization, Convolutional Neural Networks, Recurrent Neural Networks, Convolutional Recurrent Neural Network

# Introduction

Machine learning has become an integral part of the world’s attempt to satisfy the need for entertainment through massive amounts of data. With a huge amount of music being created and released on the internet, and different people from different diversities starting to create music due to easier access to technology, music is not the same as it was a decade ago.

A new genre of music comes up almost every fortnight. People have started to consume music at an enormous rate. At the same time, people have become increasingly loyal to the genre they prefer over all other genres. Users on different music providing services want that new music belonging to their favourite genre to be selected automatically and be played for them.

This requires the help of machine learning since computers themselves cannot comprehend choice or music for that matter. This research focuses on the classification of music based on genre using deep-learning with Convolutional Neural Networks.

## Background

The increase in the amount of music produced and the number of people producing it has brought about a lot of varieties of music on the internet. The preference of the people is driven by the genre of music they prefer and they consume music in this way. Without a proper way to classify music based on genre, it would not be easy to fulfil the demand of the consumer market expecting music according to their preferences.

## Aim/Objective

This research aims to classify the data in the Music Archives according to Genre. This can be done by:

* Performing a strong Literature Review to validate the research
* Performing dataset selection, analysis and preprocessing
* Creating the deep-learning logic for digesting and training on the dataset
* Cross-validation for performance metrics

## Research Questions

The question that motivates the project’s progress forward is as follows:

1. Can Deep-learning keep up with the amount of data produced in the music domain for accurate genre classification?

## Ethical Considerations

Ethics is a complicated subject that has only become more prominent during the advent of Big Data. The UK Data Service department also provides guidelines for ethical research with specific relation to Big Data. These guidelines will form the basis for this report’s ethical approach. Some of the concerns that will be addressed are:

* Maintaining confidentiality in line with Birmingham City University (BCU) and DC guidelines.
* Anonymizing information that violates group privacy.
* Ensuring transparency in reasons for data collection.
* Ensuring data is only used for the direct purpose it has been requested
* Referencing sources for all information used within the research project.
* Ensuring all data is stored in the correct location. DC information must remain on DC servers.

As this project encounters any further ethical concerns these will be met within the recommended UK guidelines and with the advice of BCU and DC supervising members.

## Literature Review

A strong Literature Review provides good guidance from the experiences of fellow researchers working in the same domain. It also provides quality and validation to the research being done. The following papers have been finalized for Literature Review in this Research.

### CONVOLUTIONAL RECURRENT NEURAL NETWORKS FOR MUSIC CLASSIFICATION

The authors present a convolutional recurrent neural network (CRNN) for music labelling. CRNNs exploit convolutional neural organizations (CNNs) for the neighbourhood include extraction and intermittent neural organizations for transient summarization of the separated highlights. The authors contrast CRNN and three CNN structures that have been utilized for music labelling while at the same time controlling the number of boundaries regarding their exhibition, what's more, preparing time per test. Generally, the authors found that CRNNs show a solid exhibition regarding the quantity of boundary and preparing time, demonstrating the adequacy of its half breed structure in music include extraction and highlight summarization.

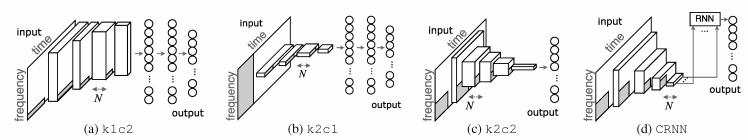


Figure 1 - Block Diagrams of the ConvNets being Compared Together

The authors contrast CRNN and klc2, k2cl, and k2c2, which are outlined in *Figure 1*. The three convolutional networks are named to determine their bit shape also, convolution measurement (for example c2 for 2D convolutions). The details appear in Table 1. For all organizations, the information is thought to be of size 96 x 1366 (Mel-frequency band x time) and single channel. Sigmoid capacities are utilized as enactment at yield hubs since music labelling is a multi-name characterization task. In this paper, all the convolutional and completely associated layers are furnished with indistinguishable enhancement strategies what's more, initiation capacities - group standardization and ELU enactment work. This is for a right correlation since enhancement methods enormously improve the exhibitions of organizations that are having something similar structure. Uncommonly, CRNN has frail dropout (0.1) between convolutional layers to forestall overfitting of the RNN layers.

The authors utilize the Million Song Dataset with last.FM labels. The authors train the organizations to foresee the main 50 tag, which incorporates sorts (e.g., rock, pop), dispositions (e.g., pitiful, upbeat), instruments (e.g., female performer, guitar), and times (the 60s - 00s). 214,284 (201,680 for preparing and 12,605 for approval) and 25,940 clasps are chosen by utilizing the initially gave preparing/test parting and sifting through things with no best 50 labels. The events of labels range from 52,944 (rock) to 1,257 (glad).

We utilize 30-60s review cuts which are given after managing to address the feature of the tune. We trim sound signs to 29 seconds at the focal point of review cuts and down-sample them from 22.05 kHz to 12 kHz utilizing Librosa. Log-plentifulness Mel-spectrograms are utilized as contribution since they have outperformed STFT and MFCCs, and straight abundancy Mel-spectrograms in prior research. The quantity of Mel-receptacles is 96 and the bounce size is 256 examples, coming about in an info state of 96 x 1366.

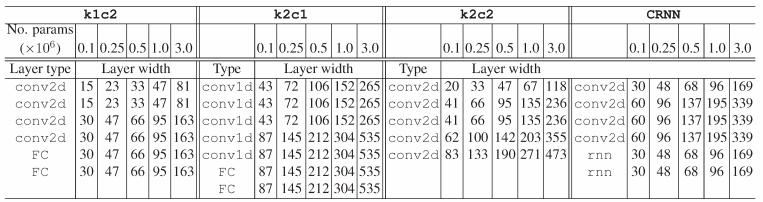


Figure 2 – Results

### CROSS-COLLECTION EVALUATION FOR MUSIC CLASSIFICATION TASKS

In this paper, the authors present a philosophy and programming instruments for cross-assortment assessment for music characterization undertakings. The instruments permit clients to direct huge scope assessments of classifier models prepared inside the AcousticBrainz stage, given a free wellspring of ground-truth comments, and its planning with the classes utilized for model preparing.

The authors propose a cross-assortment assessment measure, that is, an assessment of models on free arrangements of music tracks explained with an autonomous ground-truth source. In this paper, the authors present a technique and programming devices for such assessment for music arrangement undertakings. The authors use AcousticBrainz, a community-based stage for get-together music data from the sound. It contains MIR-related music highlights for over 3 million chronicles including copies. It gives the usefulness to make datasets comprising of accounts and related ground truth, preparing classifier models, and applying them to accounts present in AcousticBrainz.

Various models prepared on type datasets utilized inside MIR are as of now included. The apparatuses permit the AcousticBrainz community group to lead cross-assortment assessments of classifier models prepared on the AcousticBrainz site given any autonomous wellspring of ground-truth explanations for accounts and planning between a model's classes and the classes inside that ground truth. To exhibit the proposed philosophy and devices, the authors assess five type classifier models prepared on MIR sort datasets. The authors assemble a classification ground truth for chronicles in AcousticBrainz utilizing cooperative labels from Last.FM and consider different assessment techniques for planning the classifier models' yields to the ground-truth classes. The authors utilize instruments on approval sets from 240,000 to 1,740,000 chronicles and talk about the got results.

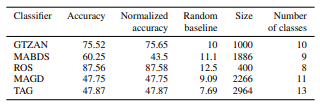


Figure 3 - Cross-validation Results

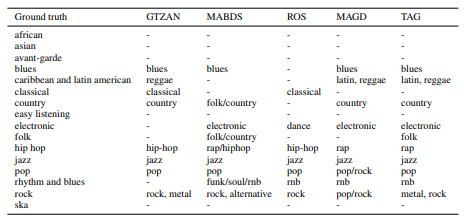


Figure 4 - Mapping between Classifier Classes and Ground Truths

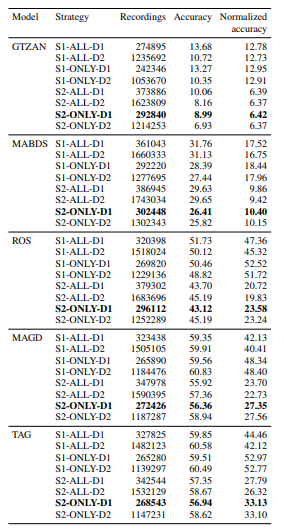


Figure 5 - Cross-collection accuracies for all involved classifiers.

### An Evaluation of Convolutional Neural Networks for Music Classification Using Spectrograms

In this work, the authors contend that they can go further with the time-frequency investigation using portrayal learning. To show that, the authors contrast the outcomes got and a Convolutional Neural Network (CNN) with the outcomes acquired by utilizing high-quality highlights and SVM classifiers. Also, the authors have performed tests intertwining the outcomes acquired with learned highlights and handmade highlights to evaluate the complementarity between these portrayals for the music characterization task. Examinations were conducted on three music information bases with unmistakable qualities, explicitly a western music assortment generally utilized in research benchmarks, an assortment of Latin American music (LMD data set), and an assortment of field accounts of ethnic African music.

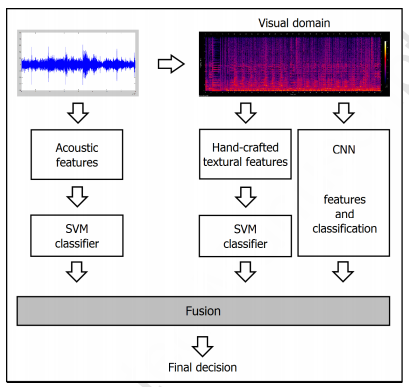


Figure 6 - Methodology Overview

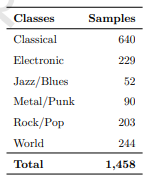


Figure 7 - ISMIR 2004 Music Dataset Features

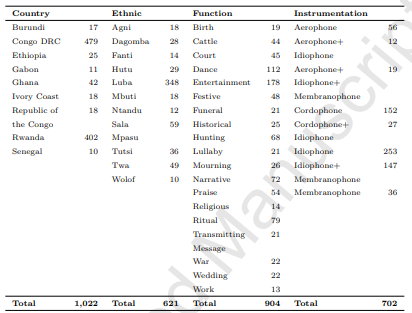


Figure 8 - Stats of the African Music Collection

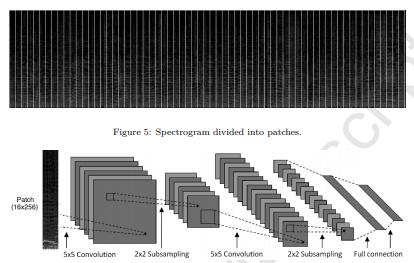


Figure 9 - Working of the ConvNet

The info is a fix of 256 × 16. The convolutional layers have teachable channels that are applied across the whole picture. The meaning of the layers incorporates the channel size and step (distance between the utilization of the channels). If the step is more modest than the channel size, the channel will be applied in covering windows. In this examination, the best outcomes were accomplished utilizing 5 × 5 portions with step 1. The pooling layers carry out a straight down-sampling capacity to lessen dimensionality and catch little interpretation invariances. In our investigations, distinct portions and walks were utilized however the best outcomes consistently were accomplished with window size 2 × 2 and step 2. Also, the completely associated layers are the norm for neural organizations and interface, utilizing unshared loads, every one of the neurons starting with one layer then onto the next one. For this situation, it utilizes SoftMax actuation.

The CNN was prepared to utilize the Stochastic Gradient Descent (SGD) utilizing back-engendering with 80 ages smaller than normal clumps of 128 occasions, force factor of 0.9 and weight rot of 5 × 10-4. The learning rate is set to 10-3 first and foremost to make the loads immediately fit the long gorges in the weight space, at that point it is decreased throughout the time (until 5 × 10-4) to make the loads fit the sharp ebbs and flows. The organization utilizes the notable cross-entropy misfortune work.

The characterization with hand-created highlights was performed utilizing the Support Vector Machine (SVM). SVM was picked dependent on our past chips away at music kind acknowledgement [4, 7, 6, 5]. SVM is the classifier that gives the best outcomes, subsequently, it has been chosen for our investigations in this work. The standardization was done so that each characteristic worth reaches from - 1 to +1. The outcomes introduced here were acquired by utilizing an RBF portion, where the boundaries C and γ were resolved through a network search. Regarding the three acoustic highlights examined in this work, one element vector was separated from every sound example. The RH highlight dimensionality is 60, the RP highlight vector dimensionality is 1,380, and the SSD include vector dimensionality is 161.

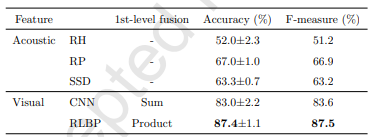


Figure 10 – Results

## Project Timeline

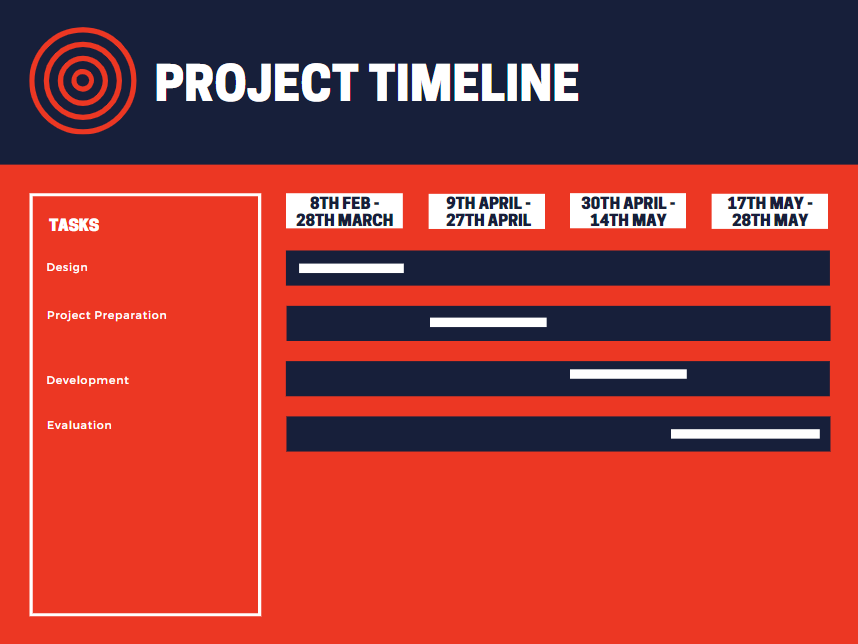


Figure - Project Timeline

# Methodology

The dataset used for this project is available [here](https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification). The dataset contains multiple datatypes. The dataset is divided as follows:

* Genres Original: The dataset contains 100 music files for each of the 10 genres this dataset is focused on. The genres is the dataset are blues, classical, country, disco, hiphop, jazz, metal, pop, reggae and rock. These music files have been collected from the GTZAN music dataset, referred to as the MNIST of sounds.
* Images Original: Each of the 10 genres under focus in this dataset also has a 100 photos each related to it that are used to train for the identification of the music genre.
* There are 2 CSV files in the dataset. Each file contains for each song it mean and variance computed for multiple features extracted from the song. One CSV file has these features for the whole song, the other CSV file has the same features but for each song split into 3 second pieces, i.e., 10 times the data.

The nature of the problem suggests the paradigm of the machine learning approach to be used for this project. The classification approach is to be used in this research, more precisely, the multi-class classification approach. The dataset contains music files, images and CSV data. Due to the presence of a lot of different and complex datatypes, encoding will be key for this project. Label Encoders and Stnadard Scalers will be used to convert the dataset into something more digestible. The Keras Classifier from scikit\_learn will be used for this project. The data visualization will be key for this project. Spectrograms and MEL spectrograms are used to encapsulate the data this research will be using.

# Project Evaluation

Project Evaluation will be done using performance metrics for classification algorithms in Machine Learning. Metrics like Confusion matrices.

# Conclusion

This research focuses on the classification of music provided by the great historical heritage of Britain according to Genre. This is done using the power of Deep Learning. CNN's are used possibly with the combination of LSTM to train and classify the music into genres.

# References

* Bogdanov, D., Porter, A., Herrera, P. and Serra, X., 2016. CROSS-COLLECTION EVALUATION FOR MUSIC CLASSIFICATION TASKS. Proceedings of the 17th ISMIR Conference.
* Choi, K., Fazekas, G., Sandler, M. and Cho, K., 2017. Convolutional recurrent neural networks for music classification. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).
* Costa, Y., Oliveira, L. and Silla, C., 2017. An evaluation of Convolutional Neural Networks for music classification using spectrograms. Applied Soft Computing, 52, pp.28-38.