University of Plymouth

School of Computing, Electronics and Mathematics

PRCO304

Final Stage Computing Project

2018/2019

*Classifying Music into Genres Using Machine Learning*

William Andrew Bilsdon

10558769

BSc (Hons) Computer Science

# **All headings and main content is very rough and is subject to heavy editing before deadline**

# Abstract

This Report will describe and evaluate a final year project that aims to explore the accuracy of a convolutional neural network and implement a trained network into a GUI in order to aid users in the classification of music based on a 30 second track.

To start the report, we discuss the methods of project management that will be used to keep the project on track. Discussing technologies used to ensure backups of the project are maintained and project management tools used to organise each week’s sprints. In this section an analysis of the initial project management is performed followed by an evaluation of the project management’s effectiveness considering the changes that had to be made to the project as time progressed.

Once the project had been planned it was important to perform initial research into the viability of the project idea. Have people used machine learning to classify music into genres before, what methods have they used and to what effectiveness. This initial research allowed us to gain an entry level understanding on Short Time Fourier Transfers (STFT for short) as well as how this data is used to generate a spectrogram. On top of this the research also entailed research into convolutional neural networks, a new network time that we had not used before.

Before the model entered development it was important to choose how many and which genres we would be training the network to classify to, we ended up choosing ten genres, some more popular and defined than others. The development of the model utilised Python, Keras and Tensorflow as a backend. These tools allowed us to efficiently design models that could be trained against our dataset. The Dataset was collected from two sources, the GTZAN audio data set and the FMA dataset, because of this a few genres have fewer tracks to be trained on than others. We finish the Model development section by evaluating the effectiveness of our chosen model when compared to other models trained during the process.

The GUI went through multiple changes during development, these changes are discussed and evaluated in order to determine why we had short comings in our GUI planning. We also analyse our HCI testing results as well as how and why we implemented the changes suggested.

The report is concluded with an evaluation of the product as a whole, suggest for improvement as well as how to avoid the same short comings in the future are also presented.

# Acknowledgements

Table of Contents

[Introduction 3](#_Toc7639488)

[Project Management 3](#_Toc7639489)

[Initial Project Plan 3](#_Toc7639490)

[Project Planning Methods 3](#_Toc7639491)

[Changes to the plan over the course of the project 3](#_Toc7639492)

[Evaluation of Project Management 3](#_Toc7639493)

[Initial Research 3](#_Toc7639494)

[Existing attempts and Their Results 3](#_Toc7639495)

[Analysis of these attempts 3](#_Toc7639496)

[Model Development Process 3](#_Toc7639497)

[Finding an appropriate method to classify 3](#_Toc7639498)

[Reason behind this chosen method 3](#_Toc7639499)

[Data Set Generation 3](#_Toc7639500)

[Reasoning behind data set size 3](#_Toc7639501)

[How I collected the dataset 3](#_Toc7639502)

[Problems faced when creating the dataset 3](#_Toc7639503)

[Model Training 3](#_Toc7639504)

[Analysis and Evaluation of each models results 3](#_Toc7639505)

[GUI Implementation 3](#_Toc7639506)

[Initial Mobile GUI Idea 3](#_Toc7639507)

[Technologies used to develop the mobile application 3](#_Toc7639508)

[Issues faced with mobile technologies 3](#_Toc7639509)

[Discuss oversights 3](#_Toc7639510)

[Revised Desktop GUI Idea 3](#_Toc7639511)

[Technologies used to develop GUI 3](#_Toc7639512)

[HCI Testing and results evaluation 3](#_Toc7639513)

[Evaluation of GUI design when compared to initial idea 3](#_Toc7639514)

[Conclusions 3](#_Toc7639515)

# 

# Introduction

Streaming music is becoming a more and more popular method of listening to music every year. Services like Apple Music make it very easy for artists to share their songs and for music fans to find new songs however all this music needs to be classified somehow and it can be expensive and time consuming for the streaming companies to higher people to do that job. We wanted to see if we could train a neural network that could classify a song with more speed, accuracy and certainty. If we could accomplish this process or at the least prove that it’s possible these companies could save a lot of money in the long run and it would also make it cheaper for new competitors to start up since they would not have to spend as much money in hiring people to perform the classification.

In order to tackle this problem we performed initial research into existing attempts to train a network to classify music and decided that using spectrograms plotted using data from a Fourier transform would be best route to take. We found several attempts that have used convolutional neural networks before however results varies **[provide references]** and we say a chance to try and improve on previous results. Others methods used included manually identifying features in songs after a fourier transform and passing the data through a Multi layer perception. These provided *worse/better* results than using CNNs[**reread and include reference links].** Another approach used a neural network to identify features then process these features in a multi layer perception in order to classify music**[include paper results and add reference]**. In the end we decided that if we took a 30 second track, we could split it into ten separate 3 second images and use a convolutional neural network (CNN for short) to detect patterns and features in the spectrogram image slices. The network would then predict the genre for each slice and then figure out the most commonly predicted genre and assign the full track to that genre. We chose to use a CNN because it was a new type of network that we had not yet used before and believed the network to be an interesting challenge and solution to the aforementioned problem.

For this project it was important that we set realistic goals for the final product. Our first goal was to have a CNN model at the very least perform better than a coin flip, 50/50. Based on our initial research we believe this goal to be a reasonable initial goal with some models in other papers reaching 70% accuracy. In the case of this objective we would consider ourselves successful if we achieved 60% accuracy per slice. With a voting system in place the accuracy of our model would actually increase since we would discard the results of any slices if the result was not the majority.

A key objective we decided upon was to allow the user to select and classify multiple songs at once. This was important as it allowed for a much smoother user experience in the final product, we felt this goal would not add too much additional time as a solution would be simple; Allow the user to select multiple files then classify them one at a time.

Finally we wanted to be able to automatically trim songs down to 30 seconds. This meant that the user would not need to perform any preprocessing themselves. Sticking to 30 second songs was important in order to maintain the accuracy of the network. By training the network on a 3 second clips of a 30 second song and then passing a 3 minute song through it there would be less information for the CNN to detect features and patterns on since the spectrogram would be compressed down to a set size thus causing results to be inaccurate.

In the following chapters we will take a look at each element of the project, analysing the process undertaken, the reasoning behind our design choices as well as an objective look at our performance, breaking down and evaluating our mistakes and difficulties faced during the project.

# Project Management

## Initial Project Plan

## Project Planning Methods

## Changes to the plan over the course of the project

## Evaluation of Project Management

# Initial Research

## Existing attempts and Their Results

## Analysis of these attempts

# Model Development Process

## Finding an appropriate method to classify

### Reason behind this chosen method

## Data Set Generation

### Reasoning behind data set size

### How I collected the dataset

### Problems faced when creating the dataset

## Model Training

### Analysis and Evaluation of each models results

# GUI Implementation

## Initial Mobile GUI Idea

### Technologies used to develop the mobile application

### Issues faced with mobile technologies

### Discuss oversights

## Revised Desktop GUI Idea

### Technologies used to develop GUI

### HCI Testing and results evaluation

### Evaluation of GUI design when compared to initial idea

# Conclusions