**Music Genre Classification Using CNNs**

ST311 Group Project by Ben Steiner and Oliver Mapps

**Dataset:** [GTZAN Dataset - Music Genre Classification](https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification?resource=download)

Our Dataset contains audio recordings of 1000 songs, each 30 seconds long and grouped evenly into 10 genres. For each song, there is also associated meta-data and a mel-spectrogram. A mel-spectrogram is a way of visualising what frequencies are more intense on the melodic scale at a given time. Two examples are given below:

|  |  |
| --- | --- |
| A picture containing text, curtain  Description automatically generated | A picture containing text  Description automatically generated |
| Metal Spectrogram | Classical Spectrogram |

We observe that the metal spectrogram is significantly busier than the classical spectrogram. We plan to learn these differences, along with the meta-data, to classify each song into one of the ten genres.

To accomplish our goal, we plan to use a CNN to learn the visual characteristics of the mel-spectrogram. We will then compare this to an MLP model that takes the meta-data as input. We will then join the two models to create an ensemble model, to maximise the usage of our data.

It is important to note that our dataset is relatively small, and as such we will take measures to increase the amount of data we can feed into our models. Firstly, the dataset provides meta-data on each 3 second segment of each 30 second recording. This dramatically increases the quantity of data we can feed into our MLP. Secondly, we plan to alter our spectrograms in ways such as splitting them vertically into sections