

Music, Mind and Technology (CSE588)

Spring 2021

Music Genre Classification

(20% of the final grade)

Objective: Implement a Genre Classification System based on the Tzanetakis & Cook (2002) paper ([Musical Genre Classification of Audio Signals](#)).

This paper describes the first system developed for this task and has a compilation of several Timbral Texture Features, Rhythmic Content Features, and Tonal (Pitch Content) Features. The dataset can be found here: <https://www.kaggle.com/carlthome/gtzan-genre-collection>

Reference Material: <https://musicinformationretrieval.com/index.html>

Team size: 2 members

Task - 1 (9%)

Every team has to perform the task of Genre Classification by experimenting with:

- 1) Different *Analysis window* sizes (pick from the range: 25ms - 150ms) and *Texture window* sizes (1, 3, 5 seconds)
- 2) Feature Selection
 - a) Try out different subsets of features to see which combination gives a better classification accuracy.
 - b) Note: The list of features provided in the paper is not exhaustive and you are encouraged to refer to more materials.
- 3) Machine Learning Algorithms
 - a) You can use any classifiers you know about (e.g., SVM, kNN, etc.)
 - b) Note: You don't necessarily have to use a single ML classifier, you can also opt for Ensemble Learning (a combination of machine learning models).

This task has to be done separately for two sets of genres from the above dataset (GTZAN), which are as follows:

Set 1 - classical, hip-hop, metal, disco, reggae

Set 2 - pop, rock, country, jazz, blues

Task - 2 (9%)

Following the analysis done for the 2 sets of genres, every team has to compare the following:

(i) Feature Selection: The input feature set used - which features provide greater classification accuracy for each of the 2 genre sets?

(For e.g., are timbral features enough to get a good accuracy for Set 1 while additional rhythmic features are required for Set 2?)

(ii) Window sizes: Which analysis and texture window lengths give the best accuracy?

(iii) ML approach: Does one ML approach consistently perform better for both sets? Or does each set have one approach that works best?

Task - 3 (2%)

Looking at the genres in each of the two sets, which set do you think would have better perceptual separability? Explain why.

Would the set of genres with greater perceptual separability result in higher classification accuracy compared to the other set? What do you observe in your experiments?