Introduction

Music genre classification is a fundamental task in the field of music information retrieval (MIR) and has numerous applications ranging from music recommendation systems to audio content analysis. In this project, we aim to explore and develop machine learning models for automatically classifying music tracks into genre categories.

Previous Solutions

Many possible methods can be used for music genre classification. Some used traditional ML algorithms, such as XGBoost[1], SVM and k-mean clustering[3]. Others utilized hierarchical taxonomy to look at similarities between genres[2]. Simple Neural Networks were also tried for clustering[3]. The spectogram of the audio could also be given to a CNN model[1].

Dataset

We used the GTZAN dataset, which includes 10 genres (blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock) and 100 samples for each genre. Overall, the dataset consists of 1000 samples, where each sample is a 30 seconds long music sample.
Source: https://huggingface.co/datasets/marsyas/gtzan

Proposed Methods

Data Processing

We used **librosa** for the audio processing tasks.
Our processing steps:

1. **normalizing** the data points in the audio
2. **segmenting** the audio
    a. Note: using 1 segment means, that we didn't split the audio data at all
3. **feature extraction**
    a. we used many features, such as tempo, MFCC, rolloff, etc.
4. **splitting** data
    a. the data was split into train (70%), validation (15%) and test (15%) parts
5. **unrolling** data
    a. some of our models expected non-seuqence inputs (such as Random Forest), to comply with this requirement, the segmented data was unrolled (meaning an audio with 30 segmentes was un-rolled into 30 separate data points, with the same label for each of them)

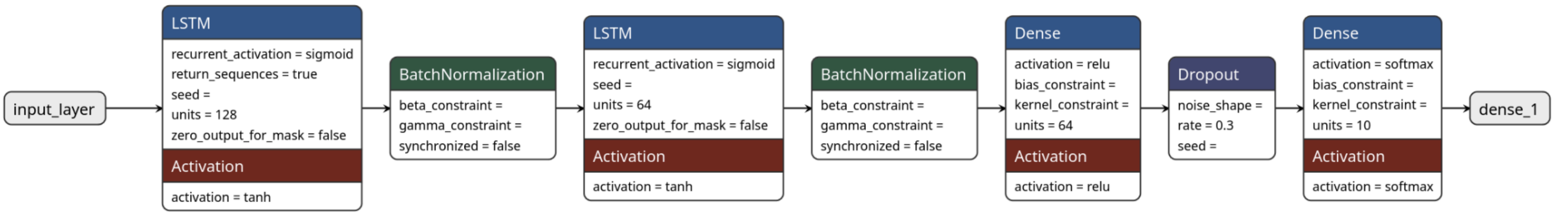
Model Training

Machine Learning models that were used:

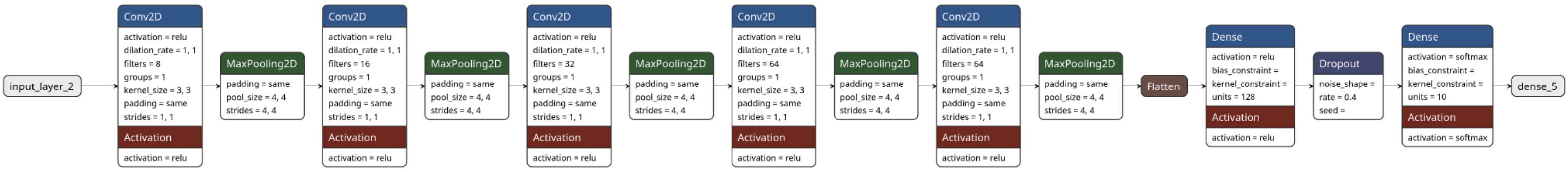
- Logistic Regression (LogReg)
- Random Forest (RF)

Deep Neural Network models that were used:

- LSTM model



- CNN model



The DNN models used Batch Normalization and Dropout layers extensively. Early Stopping was used to evade overfitting on the training set. The architecture and hyperparameters were manually optimized for best overall performance for all models.

Evaluation Method

The following metrics were used for the evaluation of the models:

- **Accuracy:** Measures the overall correctness of the model's predictions.
- **Precision:** Indicates the proportion of correctly predicted positive cases out of all cases predicted as positive.
- **Recall:** Reflects the proportion of correctly predicted positive cases out of all actual positive cases.
- **F1 Score:** Harmonic mean of precision and recall, providing a balanced measure of the model's performance.
- **Confusion Matrix:** Visualizes the model's classification performance by showing the distribution of predicted and actual labels.

Using all these metrics, we can make a strong conclusion on the model’s results in the next section.

Results and Discussion

For 1 segment, the results are:

	Accuracy	Precision	Recall	F1 score
LogReg	21.85%	8.26%	21.03%	0.1153776
RF	65.56%	62.82%	63.15%	0.622867
LSTM	39.07%	47.12%	38.45%	0.3885
CNN	**69.54%**	**68.91%**	**67.17%**	**0.670807**

For 5 segments, the results are:

	Accuracy	Precision	Recall	F1 score
LogReg	21.85%	8.26%	21.03%	0.11538
RF	65.56%	63.66%	63.12%	0.61651
LSTM	47.33%	46.27%	45.19%	0.44437
CNN	**72.40%**	**71.01%**	**71.70%**	**0.704175**

Tha tables show that as we raise the number of segments, the prediction capability of our simpler models worsen, and our more complex models' accuracy improves. The decrease of predicting capability could be explained by the fact that as the segment number rises, the amount of data in the segment decreases, and the genre can not be predicted from such a small part of the audio. While, the improvement can be attributed to the fact that LSTM and CNN require more data to learn well, so the decrease in information is balanced out by the increase in training data.

Overall, the results show that music genre classification has many possible solutions with acceptable accuracy in Machine Learning and Deep Learning.

References

[1] Bahuleyan, H. (2018). Music genre classification using machine learning techniques.
[2] Li, T., & Ogihara, M. (2005, March). Music genre classification with taxonomy. In *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* (Vol. 5, pp. v-197). IEEE.
[3] Haggblade, M., Hong, Y., & Kao, K. (2011). Music genre classification. *Department of Computer Science, Stanford University*.