

MUSIC GENRE CLASSIFICATION

Dániel Veress, Botond Rimmel

DATASET USED FOR OUR PROJECT

GTZAN Dataset - Music Genre Classification

Audio Files | Mel Spectrograms | CSV with extracted features



- most-used public dataset for evaluation in machine listening research for music genre recognition
- collected in 2000-2001 (personal CDs, radio, microphone recordings) ~ 1.5 GB (wav format)

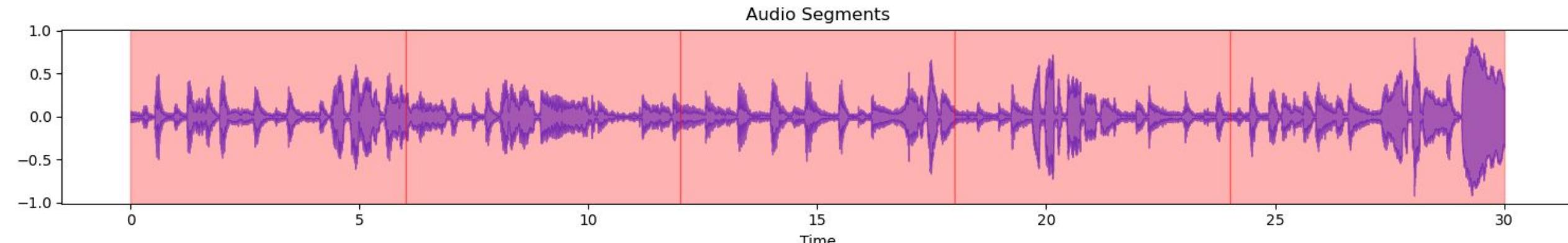
Content of the dataset

- collection of 10 genres with 100 audio files each, all having a length of 30 second
- A visual representation for each audio file
- features of the audio files (2 CSV files)

ABOUT THE CSV FILES

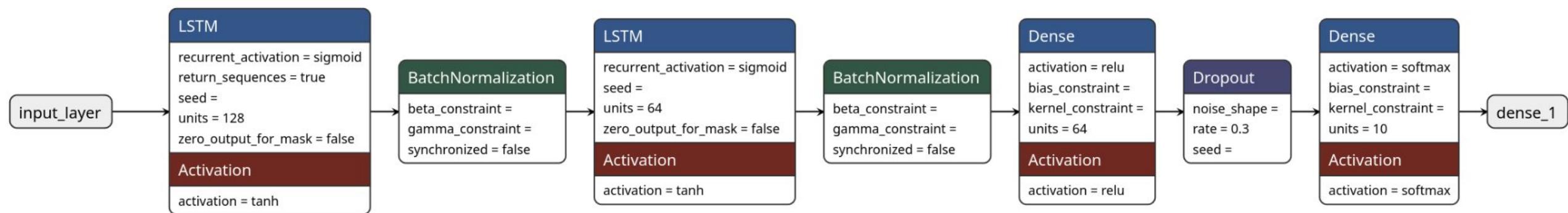
filename	# length	# chroma_stft_mean	# chroma_stft_var	# rms_mean	# rms_var
1000 unique values	660k - 676k	0.17 - 0.66	0.04 - 0.11	0.01 - 0.4	0 - 0.03
blues.00000.wav	661794	0.35008811950683594	0.08875656872987747	0.1302279233932495	0.002826696494594216 3

- One file has for each song (30 seconds long) a mean and variance computed over multiple features
- we will segment the 30 seconds => way more data for our model

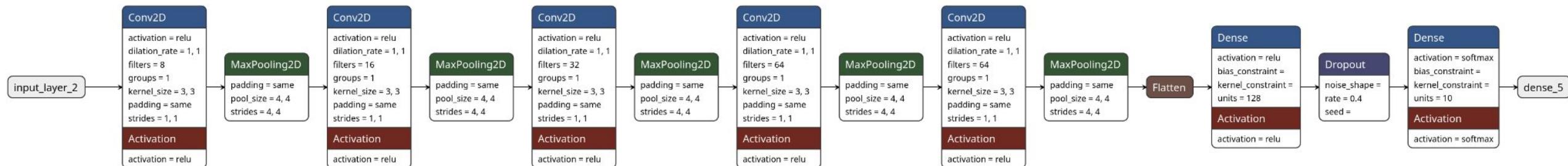


METHODS WE HAVE USED

- Simple ML methods: Random forest, Logistic regression
- LSTM

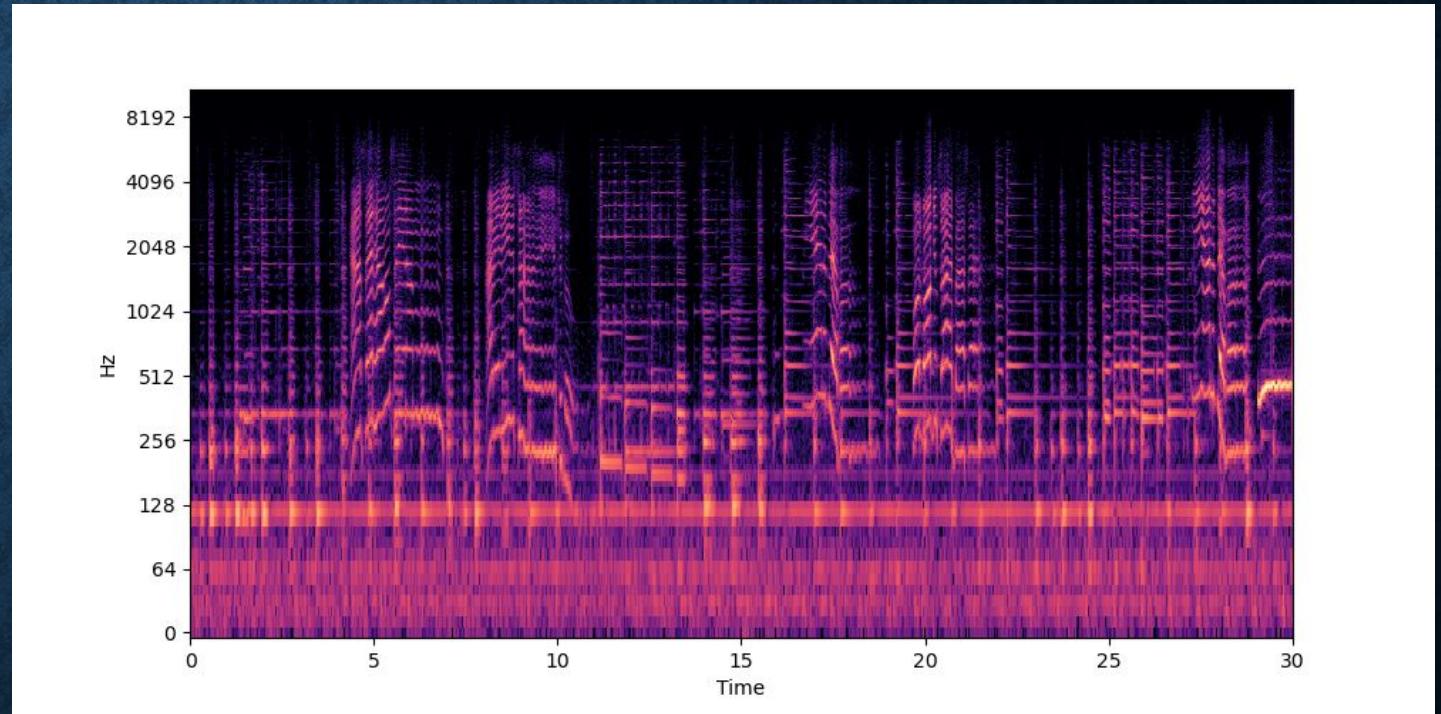


- CNN



Data Preprocessing

- for RF & LogReg - aggregated features
 - for LSTM - mfcc
 - for CNN - mel-spectrogram
-
- segmented all data
 - scaled with maximum value
 - split into train, valid and test



RESULTS

1 segment	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

5 segment	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

Improvements

- different scaling/normalization
- different models (transformer)
- more data & data augmentation
- multiple models - majority vote

THANK YOU FOR YOUR ATTENTION!