Music Genre Classification

CS 4824

Group: G5-CS

Dataset

- GTZAN dataset.

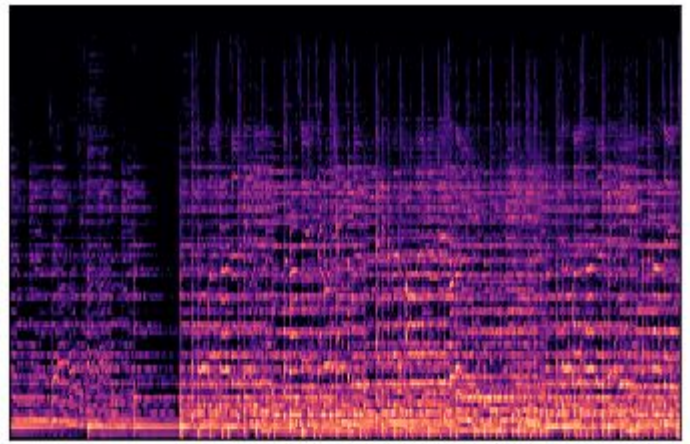
- It has 9990 data points, categorized into 10 different genres.

- It holds audio files, image files, and a CSV file of the features derived from the said audio files.

Dataset

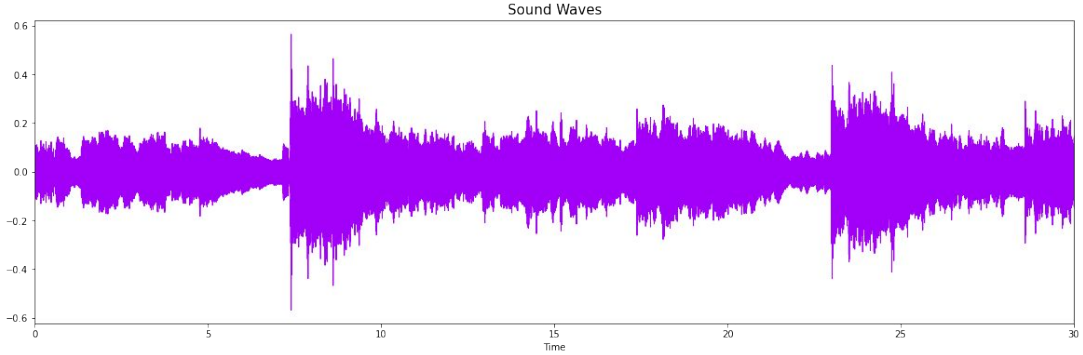
filename	length	chroma_stft_mean	chroma_stft_var	rms_mean	rms_var	spectral_centroid...	spectral_centroid...	spe
9990 unique values	66.1k 66.1k	0.11 0.75	0.02 0.12	0 0.44	0 0.03	473 5.43k	812 4.79m	499
blues.00000.0.wav	66149	0.335406363010406	0.0910482928156852	0.13040502369403	0.003521004226058721	1773.0650319984662	167541.6306866573	1972.
blues.00000.1.wav	66149	0.343065351247787	0.0861465260386467	0.1126992478966713	0.001449685427360236	1816.6937765281948	90525.69086581325	2010.
blues.00000.2.wav	66149	0.346814751625061	0.0922428891062736	0.13200338184833527	0.004628399326086044	1788.539718722745	111407.43761296556	2004.
blues.00000.3.wav	66149	0.363638788461685	0.0868561565876007	0.1325647234916687	0.002447563456371426	1655.289045406881	111952.2845173748	1960.
blues.00000.4.wav	66149	0.335579425096511	0.0881285443902015	0.14328880608881818	0.001708886175967753	1630.6561993510775	79667.26765440499	1948.
blues.00000.5.wav	66149	0.376669734716415	0.0897021070122718	0.1326178014278412	0.003582563484087586	1994.9152192705897	211700.61956933062	2152.
blues.00000.6.wav	66149	0.379908740520477	0.0888273119926452	0.1303347796201706	0.003166271140798926	1962.1500958726688	177443.07004540853	2146.
blues.00000.7.wav	66149	0.331879943609237	0.0921188518404960	0.140600323677063	0.002545942086726427	1701.8909235781698	35678.13061562547	1979.
blues.00000.8.wav	66149	0.347877383232116	0.0942091718316078	0.13312990963459015	0.002538164146244526	1746.4735023276876	138073.93124392346	1887.
blues.00000.9.wav	66149	0.358061254024505	0.0829570963978767	0.11531206220388412	0.001846020924858749	1763.9489423059536	61493.42312137413	1874.
blues.00001.0.wav	66149	0.402400881052017	0.0980397426009178	0.09302367269992828	0.003875555470585823	1279.1822416109314	406513.8167436446	1921.
blues.00001.1.wav	66149	0.345507264137268	0.0910368487238884	0.09465641528367996	0.001494711963459849	1513.7639693508	214768.80444331843	2091.
blues.00001.2.wav	66149	0.338118642565858	0.0836818367242813	0.09777601808309555	0.001385646639391	1308.8697277374274	154209.12244842653	1508.
blues.00001.3.wav	66149	0.330758524997711	0.0939360409975851	0.08536521345376968	0.00264139031860	1479.5303476963734	646868.3731510855	1986.

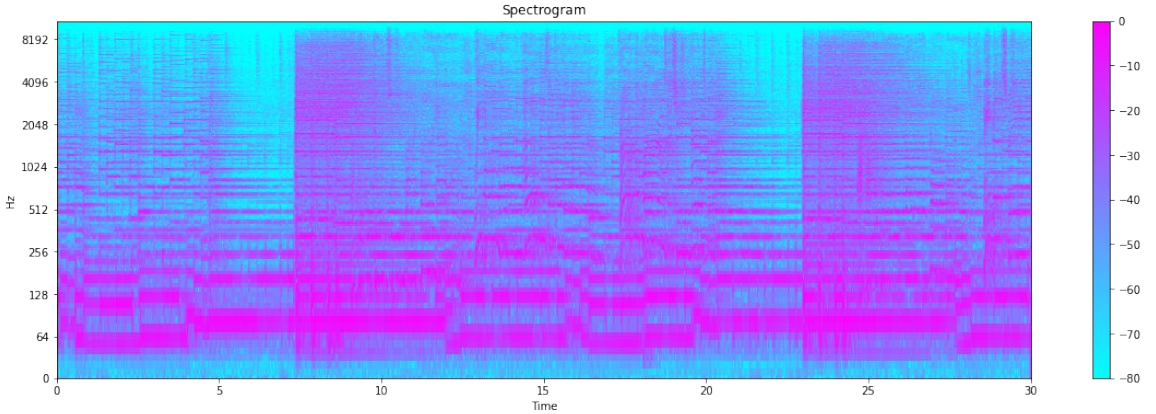
60 features



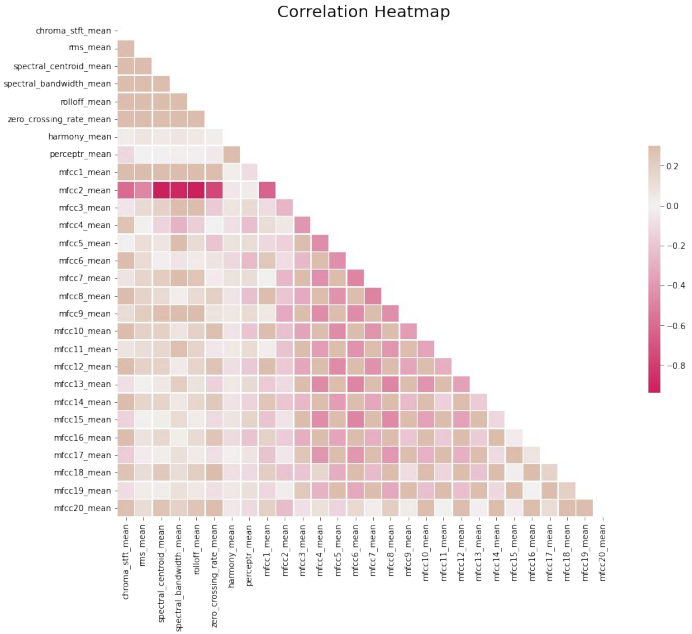
Mel Spectrogram

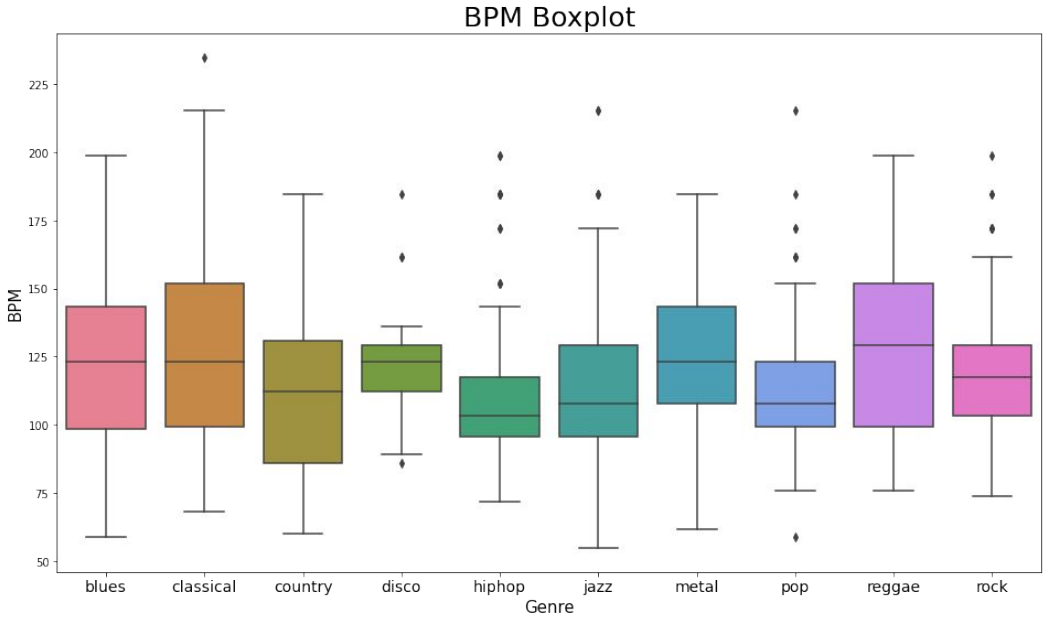
Data Exploration: Raw data

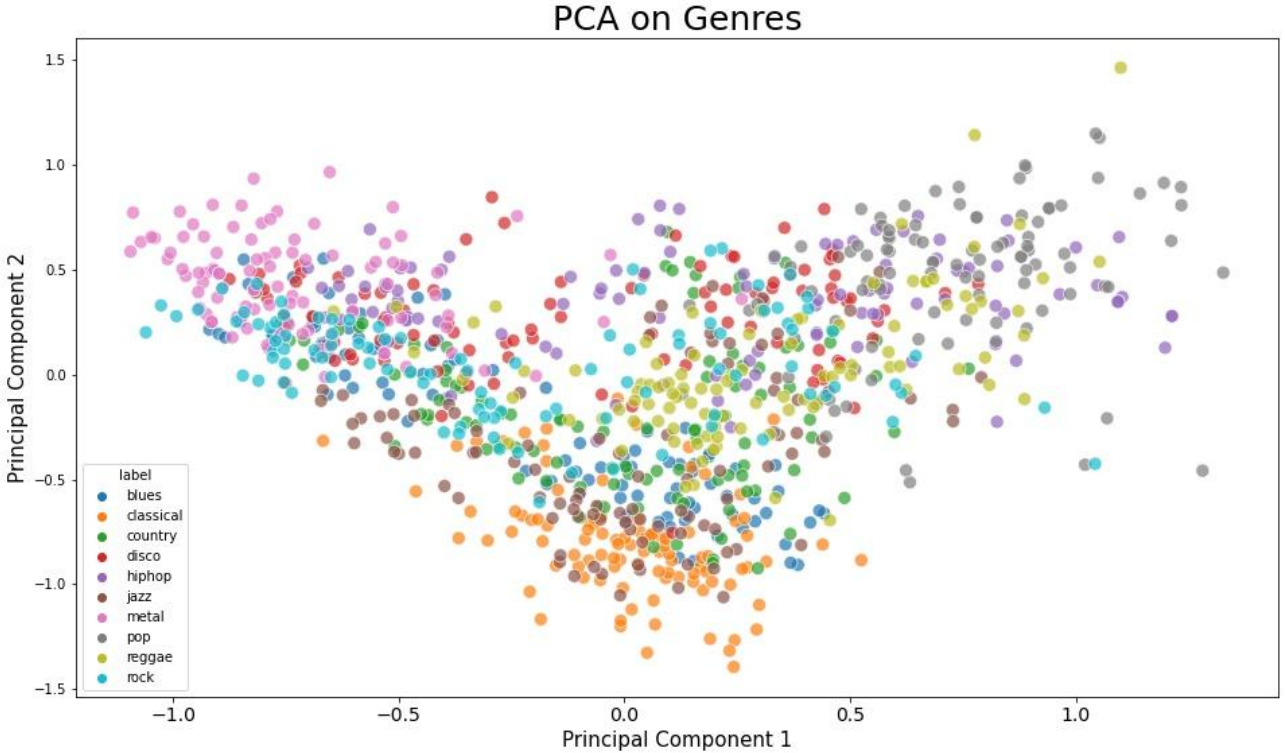




Data Exploration: Features






PCA on Genres
Principal Component 2
Principal Component 1
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
label
blues
classical
country
disco
hiphop
jazz
metal
pop
reggae
rock

Model

- Raw Audio processing
  - SVM
  - Deep Neural Network
- Spectrogram Images
  - Convolutional NN
- Audio features dataset
  - Naive Bayes
  - Random Forest
  - K-NN
  - Support Vector Machine
  - Neural Networks

Raw Audio Processing

Short-time fourier transform

- Split the 30 second audio into 10 slices of 3 sec each
- Apply the STFT on the 3 sec files
- Extract the principal components
- 9990 Samples with 10 principal components
- Split ratio 70 - 15 - 15

Raw Audio Processing

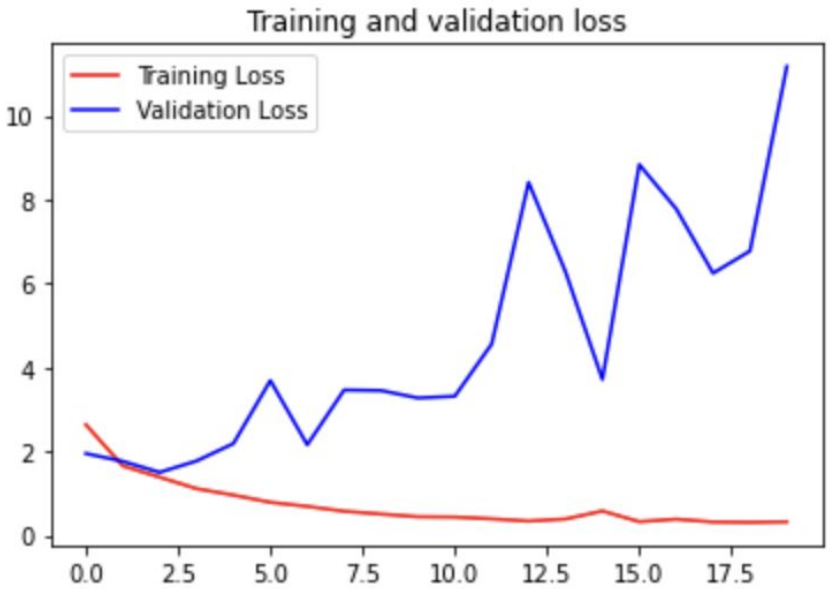
Short-time fourier transform

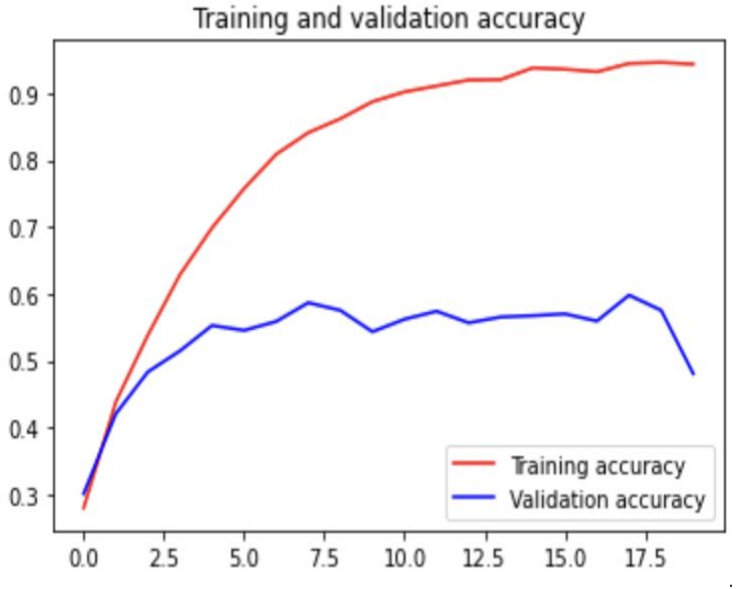
- Neural Network with 6 hidden layers
  - 2570 -> 1024 -> 512 -> 256 -> 128 -> 64 -> 10
  - Testing accuracy 55.2%

- SVM with RBF kernel
  - Testing accuracy 61.1%

Raw Audio Processing

Short-time fourier transform





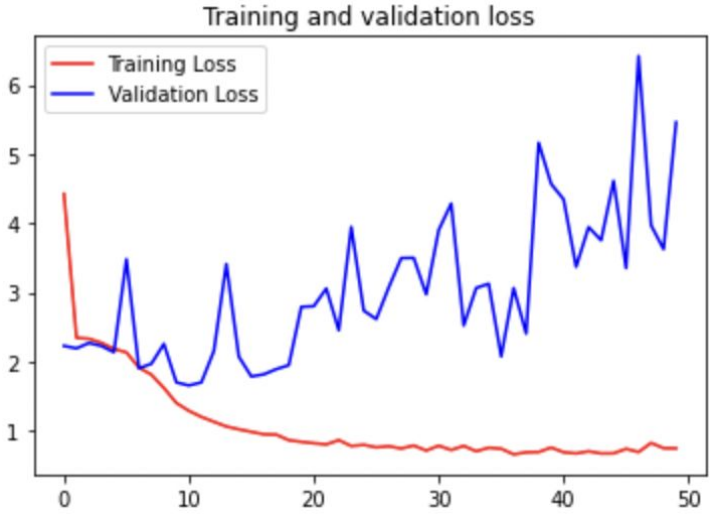
Raw Audio Processing

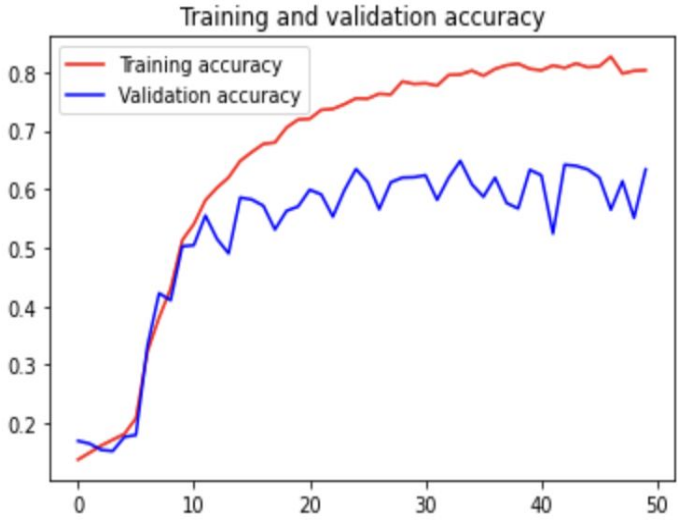
MFCC features

- Neural Network with 6 hidden layers
  - ReLU activation for hidden layers
  - 2570 -> 1024 -> 512 -> 256 -> 128 -> 64 -> 10
  - Testing accuracy 62.1%
- SVM with RBF kernel
  - Testing accuracy 56.7%

Raw Audio Processing

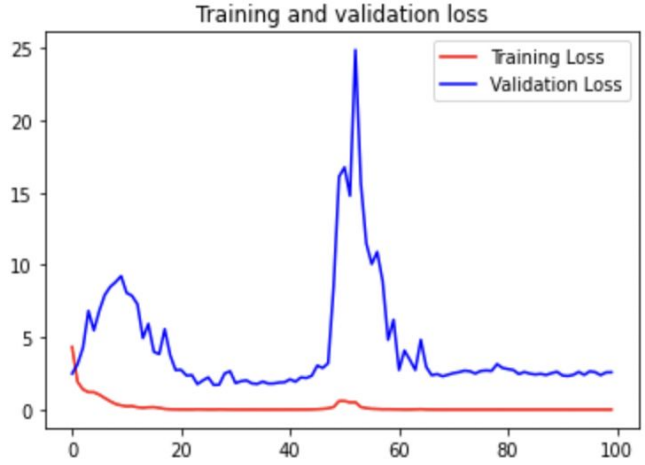
MFCC features

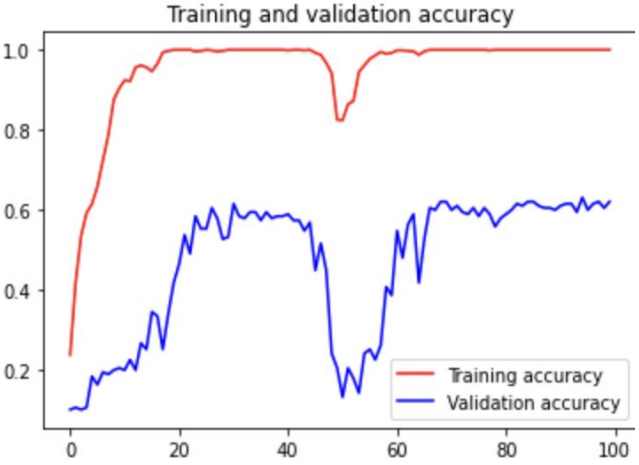




Spectrogram Images

- Convolutional NN
  - Batch normalization
  - Max pooling
  - Relu activation
- Dataset size 999 samples
- Testing Accuracy 61.9%





Audio features dataset

- We normalize all the features.
- We use a 60-20-20 train-validation-test split.
- We tune our hyperparameters (if any) using the validation set.
- We finally test on our set-out test set and report the results as follows

Results

Naive Bayes	Normalized audio features	9990	0.523
1-NN	Normalized audio features	9990	0.914
9-NN	Normalized audio features	9990	0.85
Decision tree	Normalized audio features	9990	0.640
Random Forest	Normalized audio features	9990	0.868
SVM	Normalized audio features	9990	0.747
Logistic Regression	Normalized audio features	9990	0.690
Neural Network	Normalized audio features	9990	0.672

Understanding our best models

KNN

	precision	recall	f1-score	support
blues	0.95	0.94	0.94	208
classical	0.90	0.93	0.92	203
country	0.85	0.84	0.85	186
disco	0.90	0.93	0.91	199
hiphop	0.95	0.89	0.92	218
jazz	0.87	0.90	0.88	192
metal	0.96	0.98	0.97	204
pop	0.95	0.93	0.94	180
reggae	0.92	0.93	0.92	211
rock	0.89	0.86	0.87	197
accuracy			0.91	1998
macro avg	0.91	0.91	0.91	1998
weighted avg	0.91	0.91	0.91	1998

Random Forest

	precision	recall	f1-score	support
blues	0.90	0.86	0.88	208
classical	0.92	0.99	0.95	203
country	0.72	0.81	0.76	186
disco	0.84	0.85	0.84	199
hiphop	0.94	0.85	0.89	218
jazz	0.85	0.90	0.87	192
metal	0.87	0.96	0.91	204
pop	0.92	0.91	0.91	180
reggae	0.89	0.86	0.87	211
rock	0.87	0.69	0.77	197
accuracy			0.87	1998
macro avg	0.87	0.87	0.87	1998
weighted avg	0.87	0.87	0.87	1998

KNN can pick up on metal, but not on country.

Random Forest is good with classical but not with rock.

Questions?