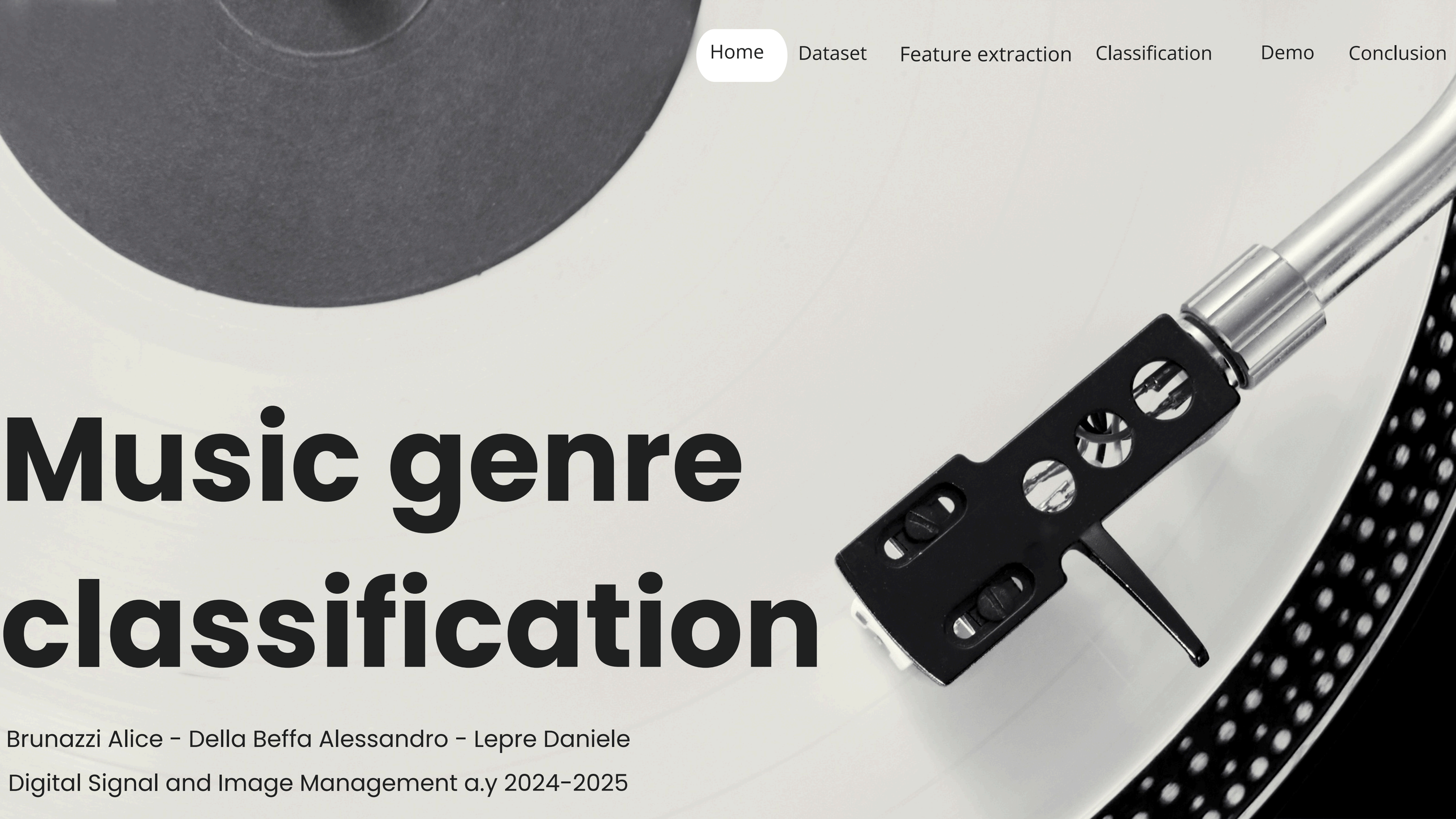


Music genre classification

Brunazzi Alice - Della Beffa Alessandro - Lepre Daniele

Digital Signal and Image Management a.y 2024-2025

Dataset

GTZAN Dataset – Music Genre Classification

It contains 1,000 audio tracks, each 30 seconds long, evenly distributed across 10 genres.

Genres Included

- Blues,
- Classical
- Country
- Disco
- Hip-hop
- Jazz
- Metal
- Pop
- Reggae
- Rock

Audio Format

- WAV files (22.05 kHz, 16-bit, mono).
- 100 tracks per genre.

PreProcess and Feature Extraction

Preprocessing Steps:

- ***Audio Segmentation*:** Each file is divided into 12 overlapping 3-second segments per track.
- ***Resampling & Padding*:** The sampling rate and audio length are standardized to ensure consistency.
- ***Normalization***: use of StandardScaler to standardize the data.
- ***Dataset Split***: 80% for training, 20% for testing.

.

Feature Extraction:

- ***MFCCs (Mel-Frequency Cepstral Coefficients)*:**Capturing timbral characteristics.
- Spectral Features:
  - ***Spectral Centroid*** : Represents the center of mass of the spectrum.
  - ***Spectral Bandwidth*:** Measures the spread of frequencies around the centroid.
  - ***Spectral Rolloff*:** Defines the frequency below which a percentage of total spectral energy is contained.
  - ***Spectral Contrast*:** Measures the difference between peaks and valleys in the spectrum.
- ***Chroma Features*:** Represent harmonic content based on pitch class profiles. (12 classes)
- ***Zero-Crossing Rate (ZCR)*:** Detects percussive elements by measuring the rate at which the signal changes sign.
- ***Root Mean Square Energy (RMS)*:** Measures signal power to reflect loudness variations.
- ***Tempo*:** Extracts the estimated beats per minute (BPM) of the audio.

Classification

C=10
gamma=0.1

Accuracy: 0.91

'bootstrap': False, '
criterion': 'gini', '
max_features': 10,
'n_estimators': 300

Accuracy: 0.88

metric': 'manhattan'
'n_neighbors': 10
weights': 'distance'

Accuracy: 0.89.

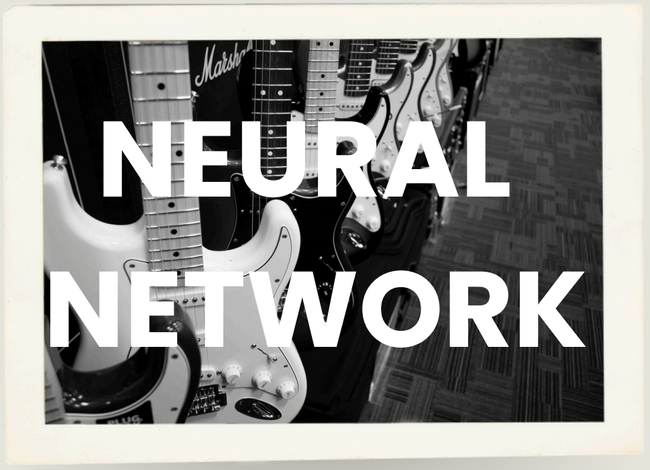
Classification

'n_estimators': 1000,
'learning_rate': 0.05,
'max_depth': 6

Accuracy: 0.89

Combines multiple models:
SVM-RF-KNN
voting='soft'

Accuracy: 0.93.

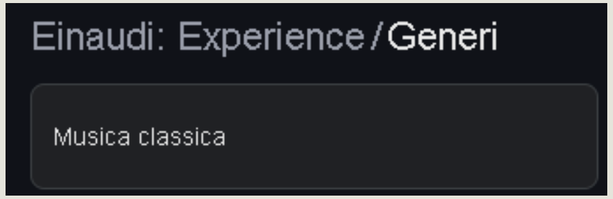


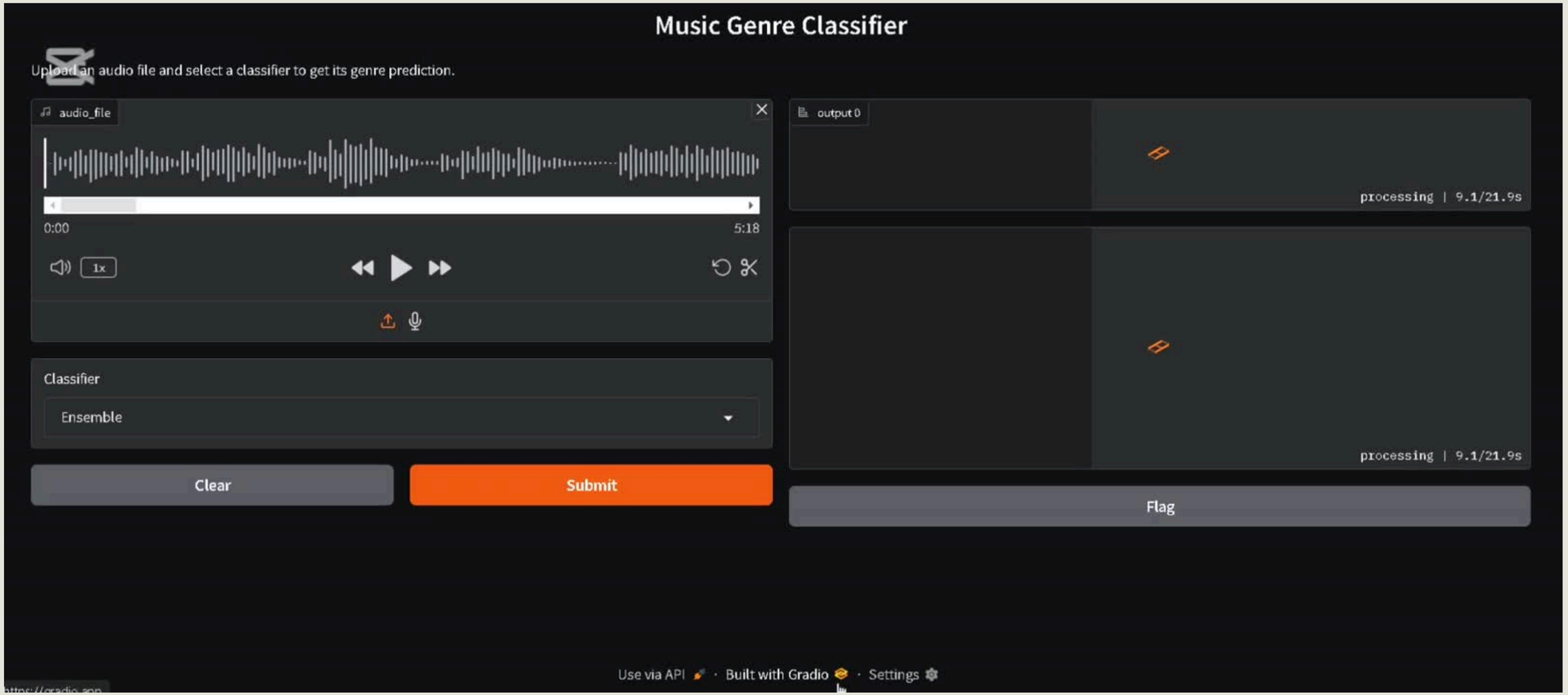
Structure: fully connected network,
4 hidden layers.
Trigger function: ReLU
Dropout: used to mitigate overfitting.
Optimization: Adam algorithm for
updating weights.

Accuracy: 0.92

DEMO

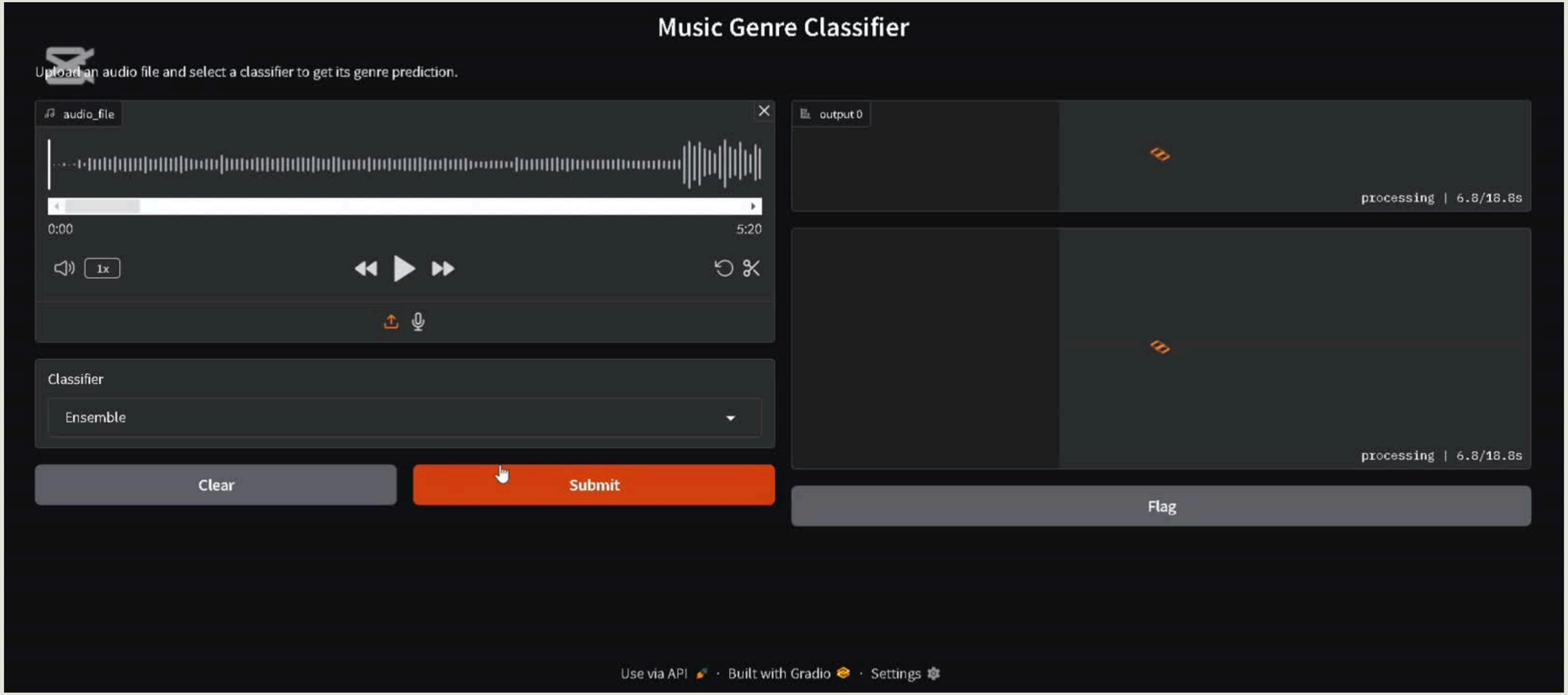
Ludovico Enaudi- Experience





DEMO

Lose Yourself - Eminem

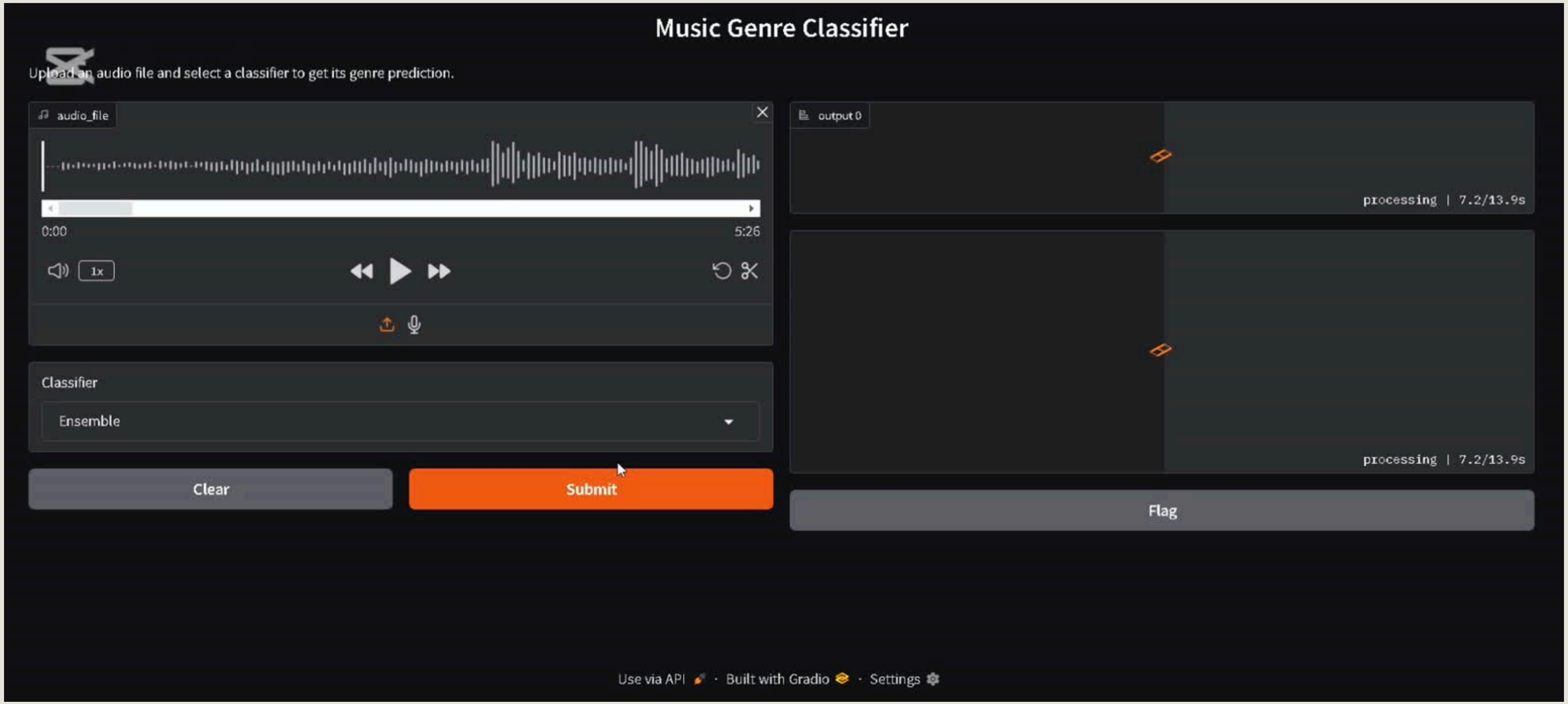


DEMO

Bob Marley - No woman, no cry

No Woman, No Cry / Generi

Reggae



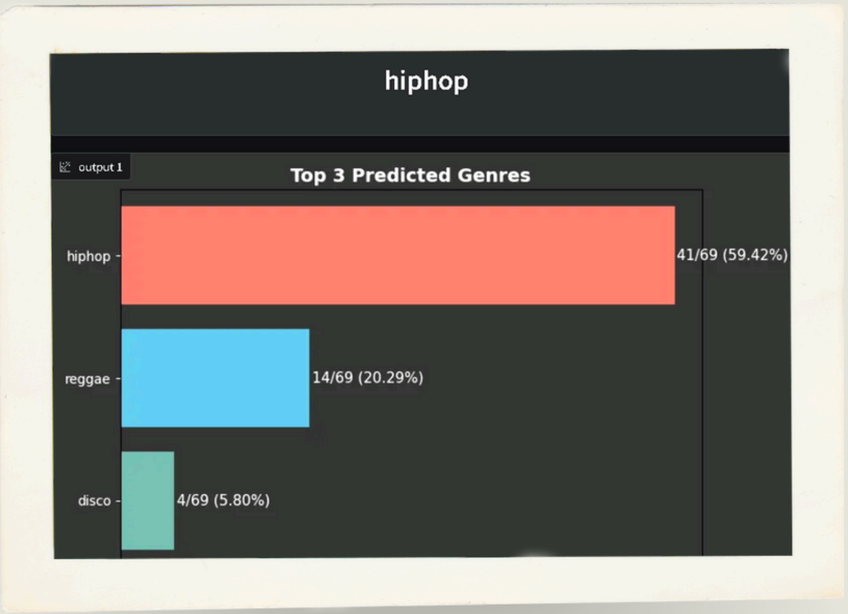
Problems and future implementations

Some genres share similar timbral and rhythmic patterns, making classification challenging.

✓ **Feature Optimization**
Combine existing features (chroma, ZCR, RMS, etc.) with higher-level representations

✓**Data Enrichment**
Enhance the dataset by incorporating additional sources and variations to increase sample diversity and capture more genre-specific characteristics.

✓**Data Augmentation**
Apply advanced data augmentation techniques (e.g., pitch shifting, time stretching) to improve model robustness.



Britney Spears - Baby One More Time (POP)

ALICE BRUNAZZI, ALESSANDRO DELLA BEFFA, DANIELE LEPRE

BIDIMENSIONAL SIGNAL

CLASSIFICATION TASK

Digital Signal and Image Management, a.y. 2024/2025

DATASET PRESENTATION

Dataset is composed by 2486 images, divided into 8 categories:

- ALPHA TAURI, 123 images
- FERRARI, 374 images
- MCLAREN, 372 images
- MERCEDES, 324 images
- RACING POINT, 290 images
- RED BULL RACING, 340 images
- RENAULT, 323 images
- WILLIAMS, 340 images

PREPROCESSING

CLASS IMBALANCE Only ALPHA TAURI label was under represented

DATASET CREATION Scanning of the directory, creation of a DF with a list of paths and a label. Each image is associated with a path. Non valid images removed.

DATASET SPLITTING After image loading, NumPy arrays are created and divided into a training set (80%, 1881) and a test set (20%, 471). Data augmentationt applied

LABEL ENCODING Strings are not compatible with ML models, so the choice was to use 1-hot-encoding on the labels

TARGET SIZE Images were converted into a 224x224 size, usefull for all the models and techniques used

AUGMENTATION Rotations, Traslation, zoom, reflection and filling

CLASSIFICATION

END-TO-END MODELS

01 CNN

02 EFFICIENT NET

03 MOBILE NET

04 VGG16

05 RESNET

06 DENSENET

MODEL RESULTS- 1

ACCURACY: 82.90%,

Three convolutional blocks, with:

- Filters: 64,128,256, with 3x3 kernels and ReLu activation
- MaxPooling to reduce dimensionality

Fully connected layer with 128 neurons (dropout to prevent overfitting)
Output layer with softmax activation

MODEL RESULTS- 2

02 EFFICIENTNET

ACCURACY: 11.68 %
ACCURACY : 16%
ACCURACY: 24%

03 MOBILENET

ACCURACY: 9%
ACCURACY: 13.8%

04 RESNET50

ACCURACY: 12.95%,

05 VGG16

ACCURACY: 22,08%,

The dataset contains only images of Formula 1 cars, which differ mainly in visual details such as colors, logos, and text, rather than in fundamentally shapes or structures. These models are TOO COMPLEX for this (small) dataset.

MODEL RESULTS- 3

06 DENSENET121

FROZEN PRE TRAINED LAYERS 64.76 %

Added a GlobalAveragePooling2D, a Dropout layer
20 epochs

FINE TUNING- DATA AUGMENTATION 93.63 %

Unfreeze the last 20 layers (to avoid overfitting).
8 epochs

POST TRAINING FINE TUNING 85.35 %

initial freezing, and only the classifier is trained (fully connected layers)
After 5 epochs, the last 30 layers are unfrozen

CLASSIFICATION

FEATURE ENGINEERING

VGG16 EXTRACTOR

- RANDOM FOREST
- SVM
- FFNN- MLP

HOG EXTRACTOR

- RANDOM FOREST
- SVM
- FFNN- MLP

COLOR HISTOGRAM

- RANDOM FOREST
- SVM
- FFNN- MLP

MODEL RESULTS- 4

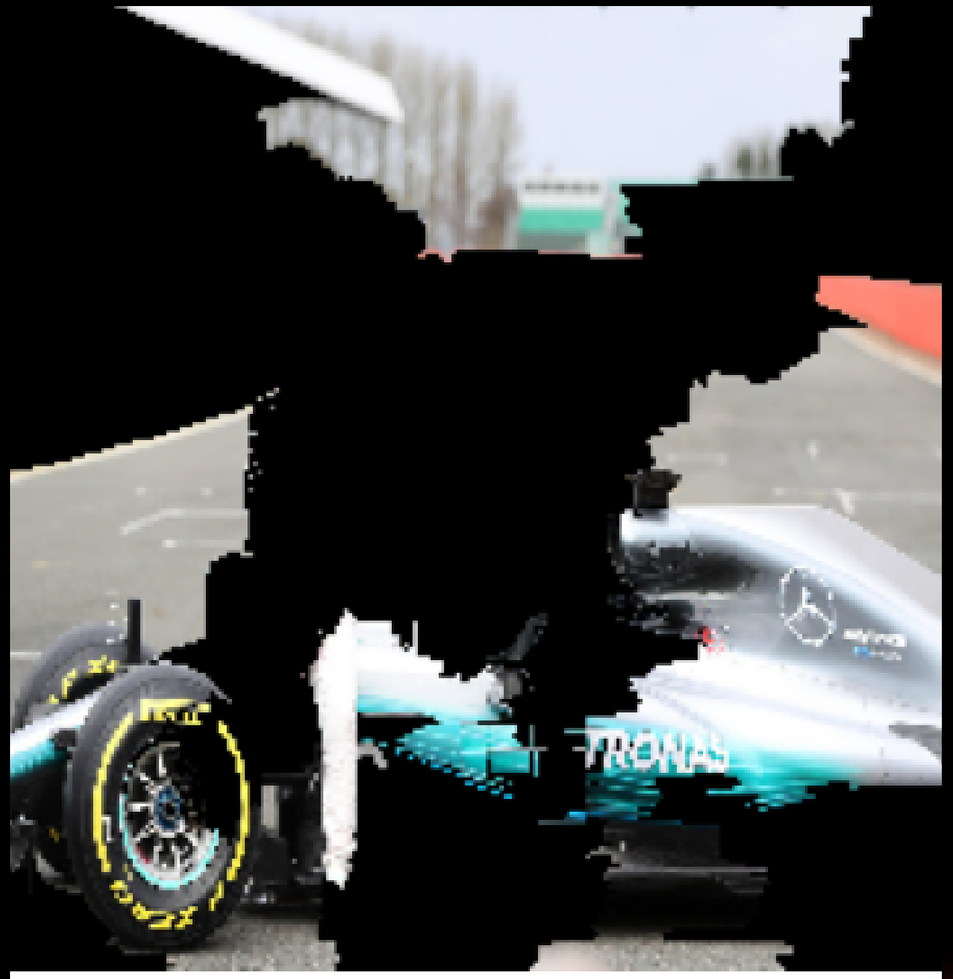
VGG16	ACCURACY: 75%	ACCURACY: 77%	ACCURACY: 80%
HOG	ACCURACY: 24%	ACCURACY: 24%	ACCURACY: 24%
ColorHist	ACCURACY: 47,98%	ACCURACY: 24,20%	

XAI CNN

PREDICTION:
Ferrari 100%

XAI CNN

PREDICTION:
Mercedes 98%



XAI CNN

PREDICTION:
Williams 100%

XAI CNN

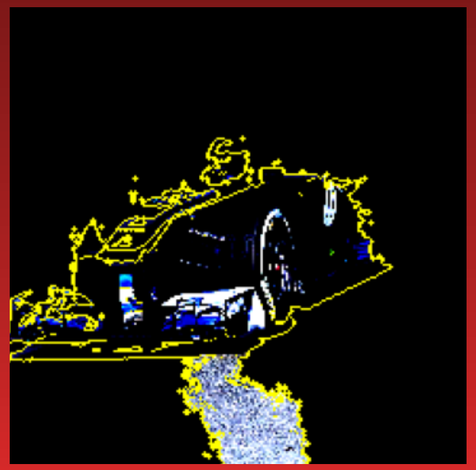
PREDICTION:
Racing Point 97%

XAI CNN

PREDICTION: Alpha Tauri 99%

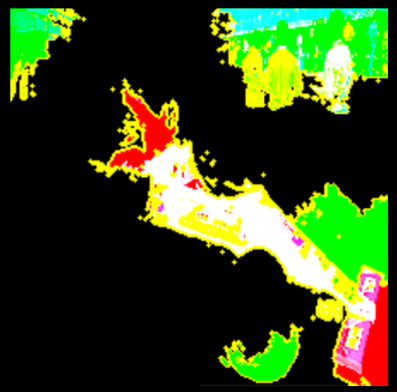
XAI DENSENET

PREDICTION:
Williams 60.17%



PREDICTION:
Ferrari 60.17%

PREDICTION:
RedBull 95%



PREDICTION:
RedBull 48%%

Content-Based Image Retrieval

from a database of paintings using CNN features

by Alice Brunazzi, Alessandro Della Beffa and Daniele Lepre

The fisherman, by Charles Napier Hemy, 1888

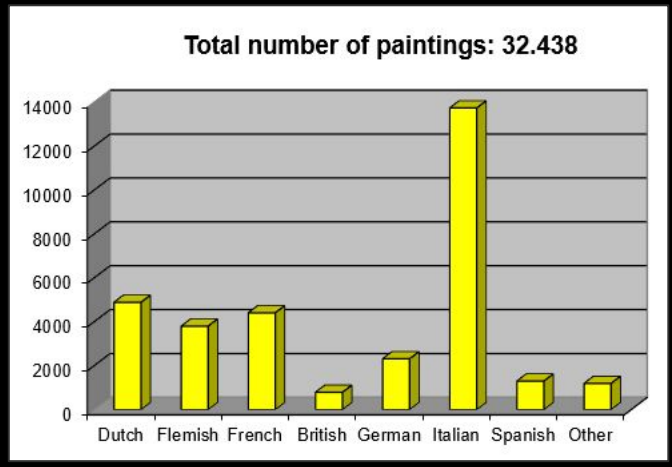
Dataset

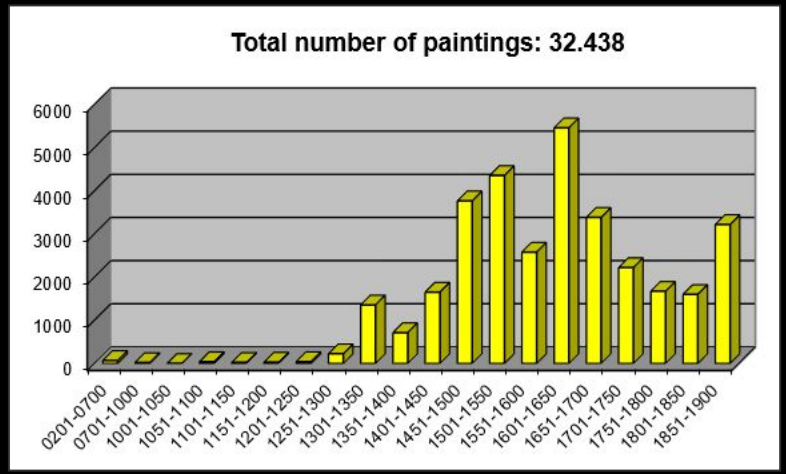
scraped from Web Gallery of Art

> [...] a virtual museum and searchable database of European fine arts, decorative arts and architecture (3rd-19th centuries), currently containing over **52,800 reproductions**.

as chosen in other works [1] related to CBIR of paintings

Paintings by *type* (subject): ***religious*** (*ca.* 13500), ***historical*** (*ca.* 900), *mythological*, ***landscape*** (*ca.* 4100), ***portrait*** (*ca.* 4800), still-life, *interior*, *genre*, *study*, *other*

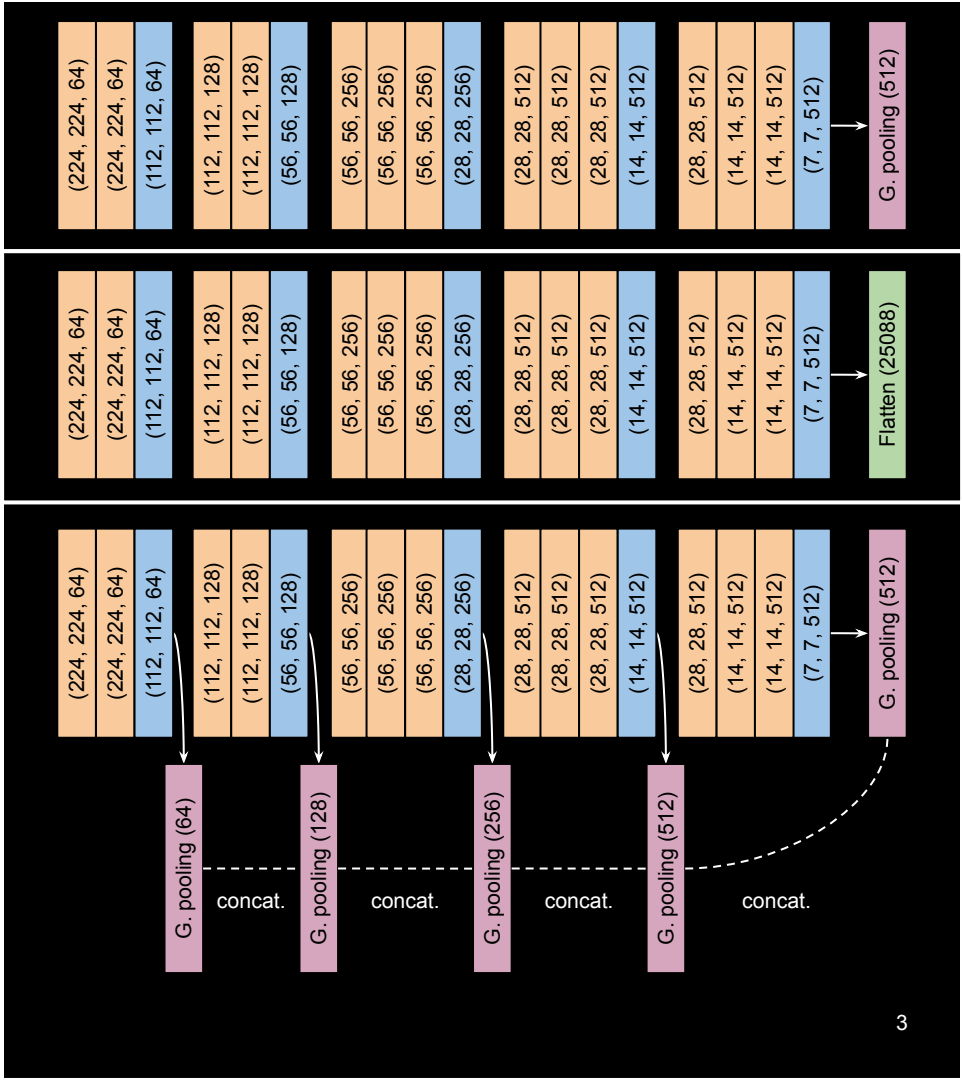




Feature extraction

inspired by [1], [2] and others, and implemented through Keras

1. **VGG16** with **global pooling layers** (512 features)
2. **VGG16** with **PCA** (25088 to 512 features)
3. **VGG16** with multiple **global pooling layers** (1472 features)
4. **VGG19** with **global pooling layers** (512 features)
5. **MobileNetV2** with **global pooling layers** (1280 features)



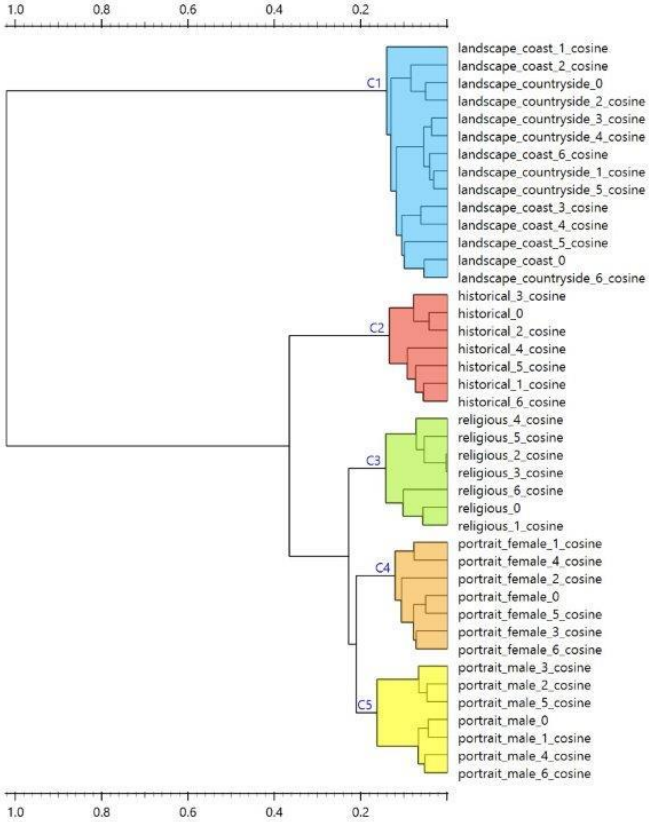
Assessment

Input: **3 couples of** potentially similar **images** selected by *type* (*religious* and *historical*; coastal and country *landscape*; female and male *portrait*)

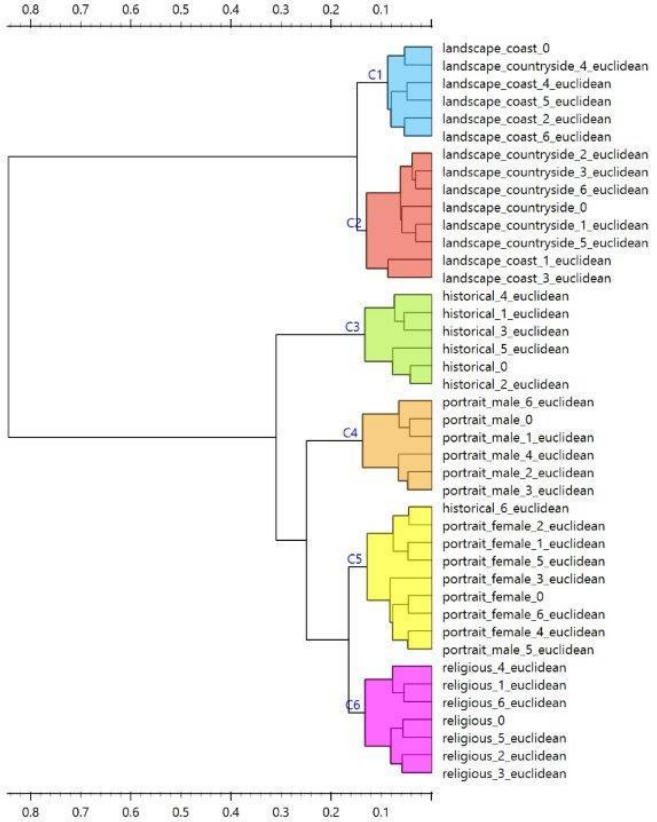
Output: for each input, the **6** most similar **images** based on **cosine similarity**, **Euclidean distance** or both

Clustering of the **6 + 36** images represented as embeddings of **different CNNs** (SqueezeNet [3], Painters), implemented via Orange [4]

	Metric	Clusters	Acc*	Prec*	Rec*
MobileNetV2	cosine, Euclidean	5	0,938	0,938	0,938
VGG16 (1)	cosine	5	**1**	**1**	**1**
	Euclidean	6	**0,881**	**0,881**	**0,881**
VGG16 (2)	cosine	5	0,786	0,786	0,786
VGG16 (3)	cosine, Euclidean	6	0,667	0,792	0,667
	Euclidean	5	0,905	0,905	0,905
VGG19	cosine	6	0,833	0,833	0,833
	cosine	5	0,905	0,905	0,905



VGG16 (1), cosine similarity



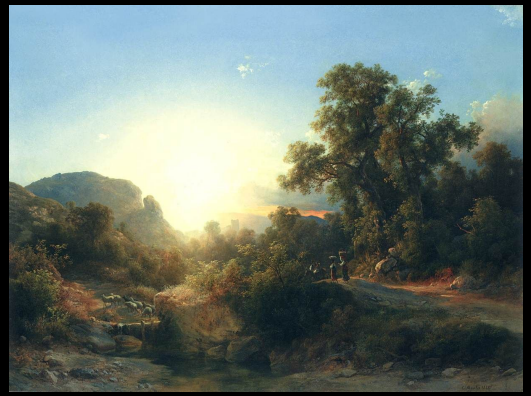
VGG16 (1), Euclidean distance

Massa, Bay of Naples, by John Brett, 1864

Shipping before a Mediterranean Coast, by Reinier Nooms, *ca.* 1623/1664

Estuary at Day's End, by Simon de Vlieger, *ca.* 1640/1645

Italian Valley, by Albert Bierstadt, 1860



La vita di Pisa, by Karoli Marko the Elder, 1791-1860

Schatacook Mountain, Housatonic Valley, Connecticut, by Jasper Francis Cropsey, 1845

Thank you

Farewell, by Edmund Blair Leighton, 1922