Music Genre Classification using ML

G HARI KRISHNA SAI 22EC01011
K MOHITH PRAKASH 22EC01002
VIKRAM RANJEET PATIL 22EC01008

GROUP – 8

PROJECT OVERVIEW

- DATA EXPLORATION AND VISUALIZATION
- FEATURE EXTRACTION: MELSPECTROGRAM ANALYSIS
- DATA PREPROCESSING
- MODEL BUILDING: ARCHITECTURE AND TRAINING
- MODEL EVALUATION - I: METRICS
- MODEL EVALUATION - II: PRECISION, RECALL, AND CONFUSION MATRIX
- PREDICTION

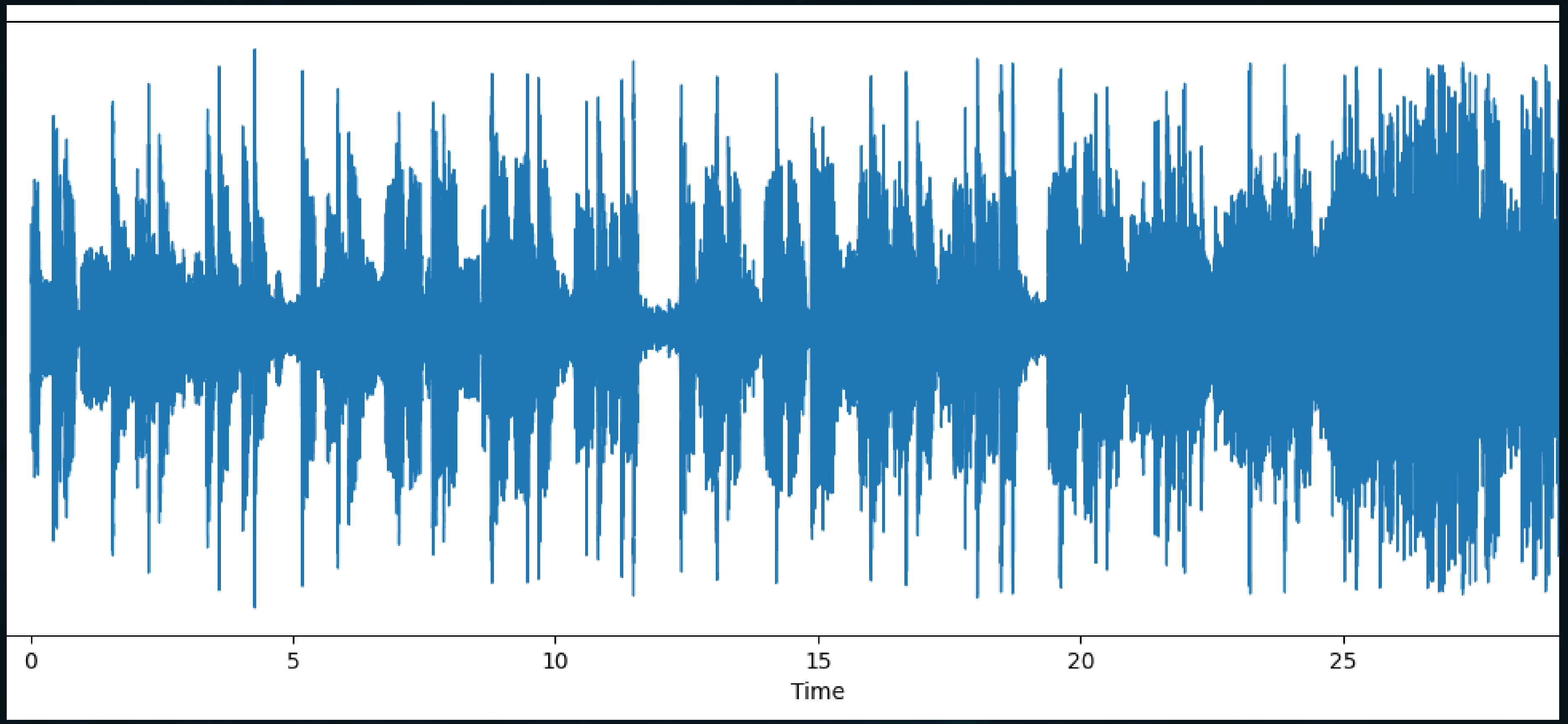
Data Exploration and Visualization

Importing Libraries

Librosa, NumPy, Matplotlib, os, tensorflow Scikit-learn, etc for audio processing and analysis.

Single Audio Visualization

Plots waveform for time-domain analysis revealing signal amplitude.



Playing Sound

Audio playback embedded using Librosa for verification.

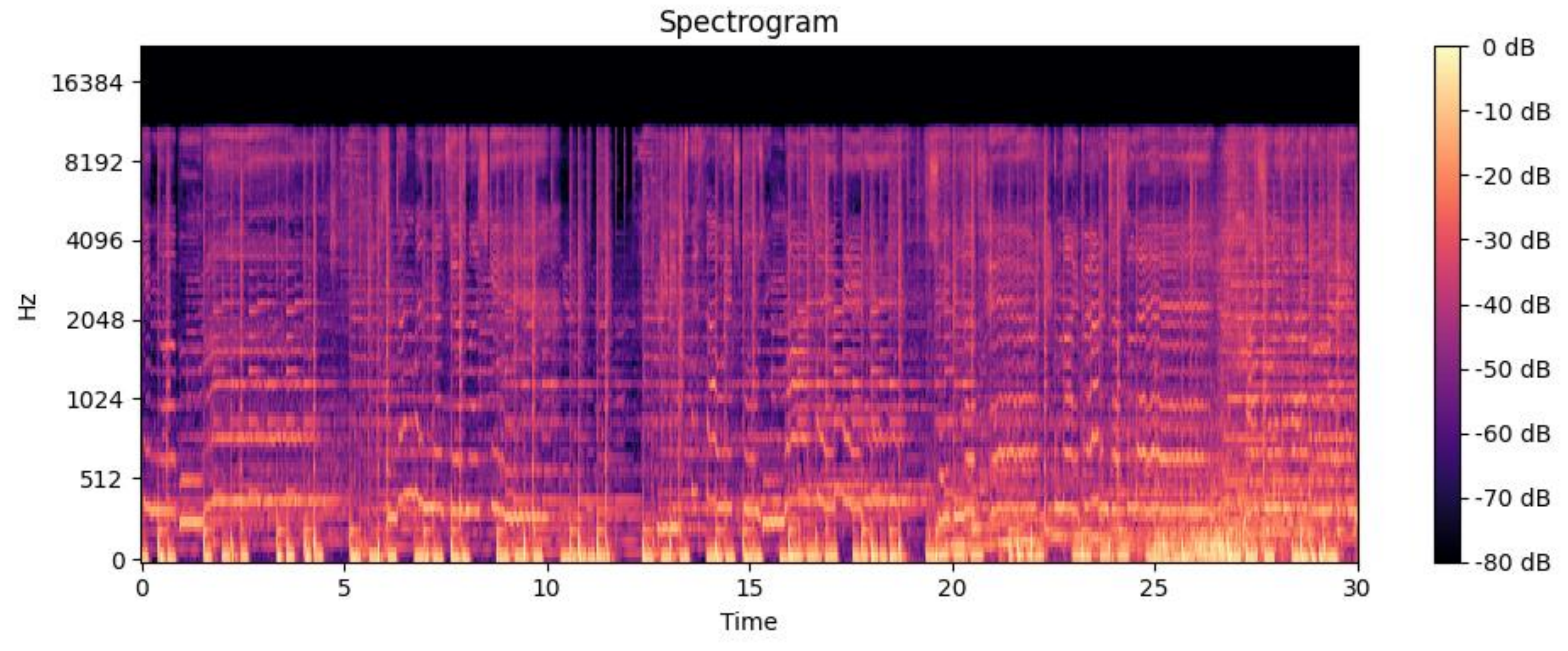
Chunk Visualization

Examines smaller audio segments to detect patterns and features.

Feature Extraction: Melspectrogram Analysis

Melspectrogram Visualization

Highlights audio frequencies using human auditory scale for clarity.



Data Preprocessing

Standardize Length

Ensures uniform audio length for consistency in feature extraction.

Scaling & Normalization

Applies normalization to reduce bias and improve convergence.

Handling Class Imbalance

Uses techniques like oversampling for balanced datasets.

Model Building: Architecture and Training

Data Splitting

80% for training and 20% for testing to validate model performance.

Model Choice

Convolutional Neural Network chosen for spatial audio feature learning.

Training Details

Trained with Adam optimizer, batch size 32, across multiple epochs.

Layer (type)	Output Shape	Param #
conv2d_10 (Conv2D)	(None, 150, 150, 32)	320
conv2d_11 (Conv2D)	(None, 148, 148, 32)	9,248
max_pooling2d_5 (MaxPooling2D)	(None, 74, 74, 32)	0
conv2d_12 (Conv2D)	(None, 74, 74, 64)	18,496
conv2d_13 (Conv2D)	(None, 72, 72, 64)	36,928
max_pooling2d_6 (MaxPooling2D)	(None, 36, 36, 64)	0
conv2d_14 (Conv2D)	(None, 36, 36, 128)	73,856
conv2d_15 (Conv2D)	(None, 34, 34, 128)	147,584
max_pooling2d_7 (MaxPooling2D)	(None, 17, 17, 128)	0
dropout_3 (Dropout)	(None, 17, 17, 128)	0
conv2d_16 (Conv2D)	(None, 17, 17, 256)	295,168
conv2d_17 (Conv2D)	(None, 15, 15, 256)	590,080
max_pooling2d_8 (MaxPooling2D)	(None, 7, 7, 256)	0
conv2d_18 (Conv2D)	(None, 7, 7, 512)	1,180,160
conv2d_19 (Conv2D)	(None, 5, 5, 512)	2,359,808
max_pooling2d_9 (MaxPooling2D)	(None, 2, 2, 512)	0
dropout_4 (Dropout)	(None, 2, 2, 512)	0
flatten_1 (Flatten)	(None, 2048)	0
dense_2 (Dense)	(None, 1200)	2,458,800
dropout_5 (Dropout)	(None, 1200)	0
dense_3 (Dense)	(None, 10)	12,010

Model Evaluation – I: Metrics

1 **Accuracy & Loss**

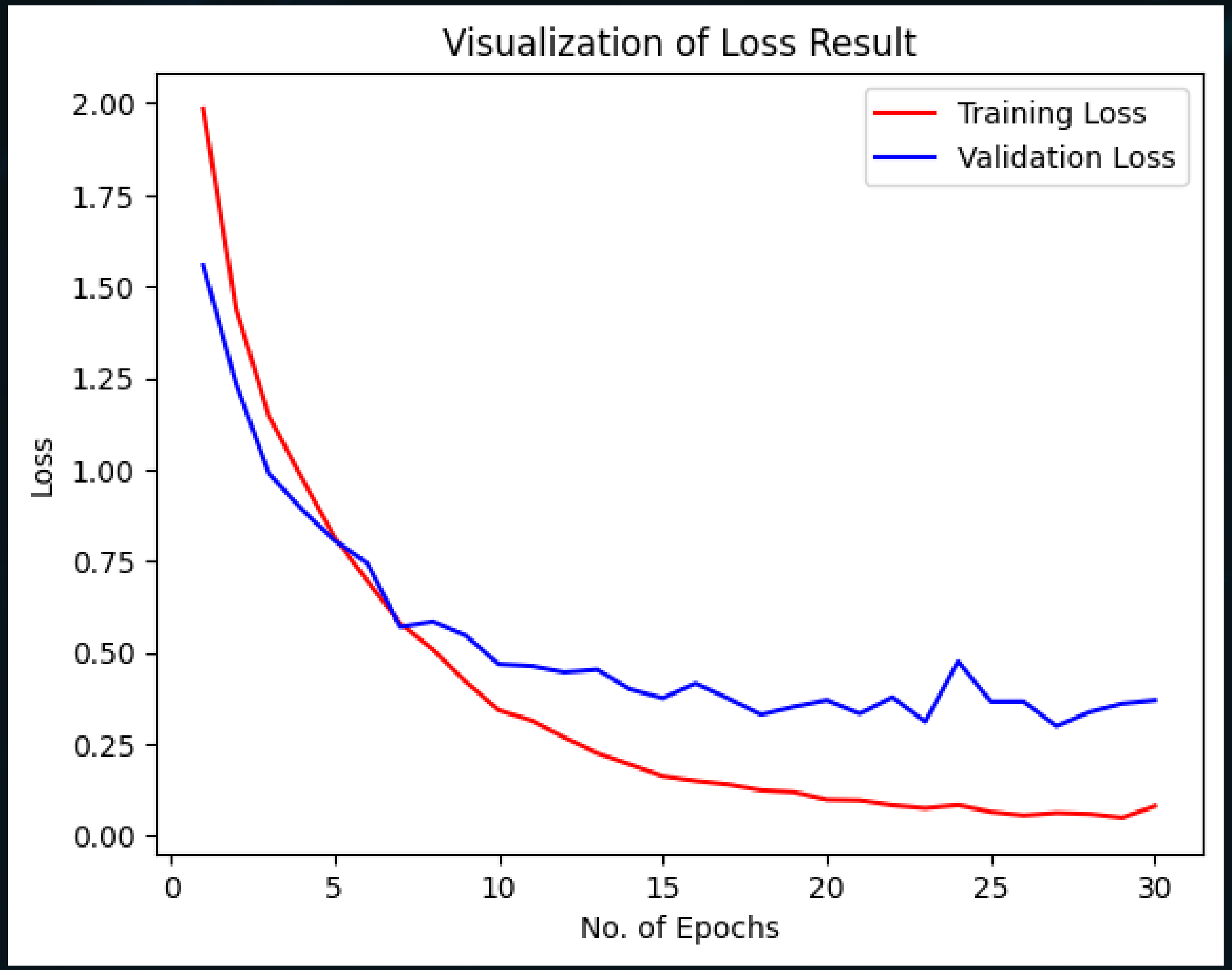
Tracking to detect overfitting and evaluate performance trends.

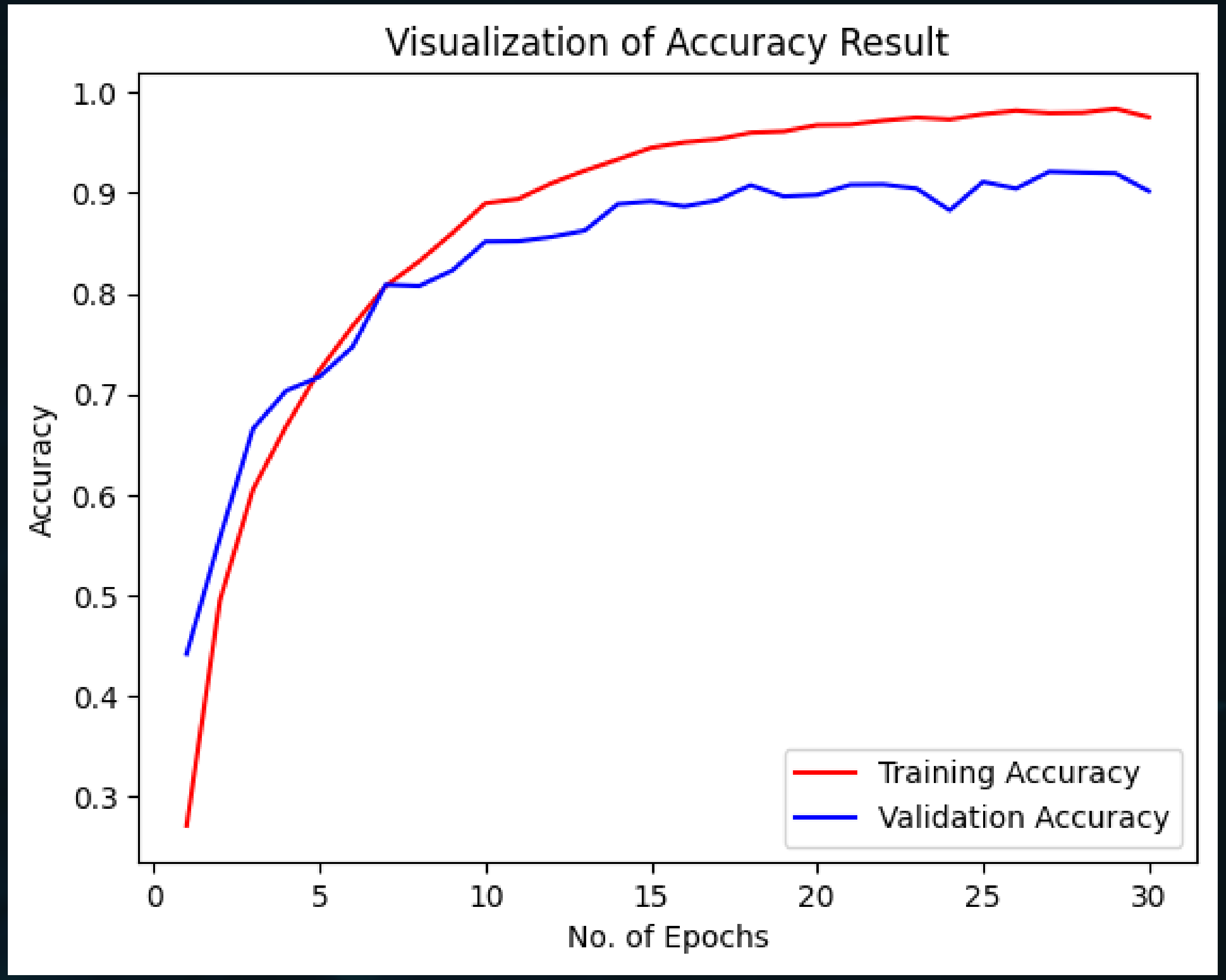
2 **Metrics Used**

Accuracy, precision, recall, and F1-score for comprehensive assessment.

3 **Metric Explanation**

Each metric highlights different capabilities of the model.





Classification report

	precision	recall	f1-score	support
blues	0.81	0.92	0.86	299
classical	0.94	0.97	0.95	299
country	0.90	0.81	0.85	329
disco	0.93	0.88	0.91	304
hiphop	0.89	0.97	0.93	290
jazz	0.90	0.93	0.91	308
metal	0.97	0.96	0.96	319
pop	0.94	0.84	0.89	283
reggae	0.92	0.88	0.90	283
rock	0.83	0.86	0.85	278
accuracy			0.90	2992
macro avg	0.90	0.90	0.90	2992
weighted avg	0.90	0.90	0.90	2992

Model Evaluation – II: Precision, Recall, and Confusion Matrix

Precision and Recall

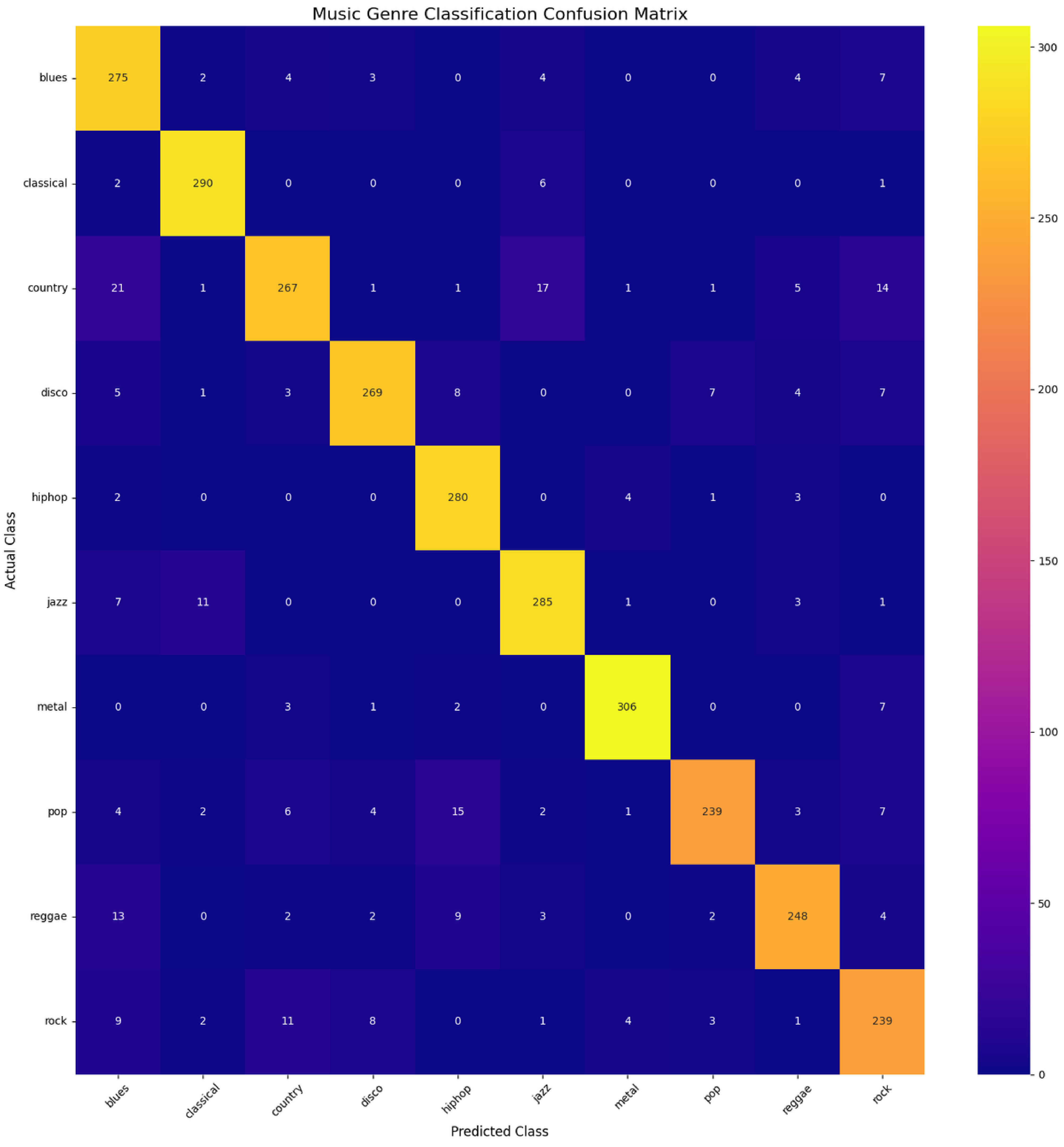
Measure correctness and completeness per genre classification.

Confusion Matrix

Visualizes true vs predicted labels, highlighting errors.

Insightful Discussion

Analyzes misclassifications to improve model and focus on weak genres.



Prediction

- Model Loading and Setup
- Audio Preprocessing
- Genre Prediction Per Chunk
- Final model prediction

Future Directions:

1. Model Improvements

Use Pretrained CNN Models (Transfer Learning), Data Augmentation for Audio,Use MFCCs or Chromagrams

2. Multi-Label Classification

Train the model to allow multiple genres per song using:

- sigmoid activation instead of softmax
- binary_crossentropy loss

3. Use Recurrent Layers or Transformers

Add LSTM/GRU layers after CNNs

4. Deploy as an App or Web App

5. Expand Dataset and Classes

Conclusion

Summary

Achieved strong classification accuracy demonstrating model effectiveness.

Limitations

Dataset size and feature scope limit full model potential.

Future Directions

Explore deeper networks, larger datasets, and advanced audio features.