ECE 208 - S’2025
Kareem Hassan

Noise Tolerant Music Genre Classification

OBJECTIVE

Design a music genre classification system that remains accurate and reliable, even when audio quality is degraded by noise or corruption.

PROBLEM SOLVED

- Smart Speakers & Assistants
  - Help w/ Music Sorting & Recommendations even with background noise.
- DJ Software or Music Tools
  - Assist by organizing tracks according to genre automatically.
- Audio Archiving
  - Classify old or degraded recordings.
- Music Streaming Platforms
  - Enable genre detection on devices in noisy environments.

DATASET & PREPROCESSING

GTZAN

Features: 1000 Audio Tracks
30 Seconds each
10 Music Genres

Genre: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae, rock

Format: .Wav Files & 22050 Hz Sample Rate

PreProcessing

Convert Audio to **Mel Spectrograms**

Split: 70% Validation, 15% Training, 15% Test

Noise Augmentation

Package: Librosa

White Noise (Static Sound)

Pink Noise (Background Hum)

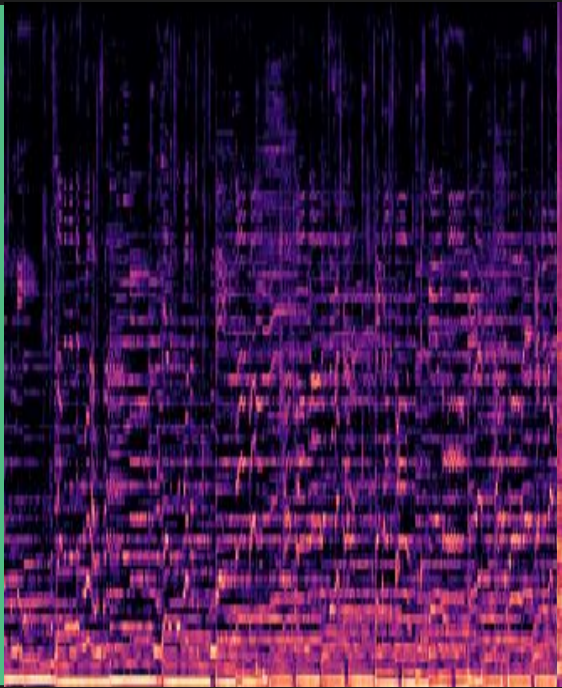
NOISE AUGMENTATION EXAMPLE

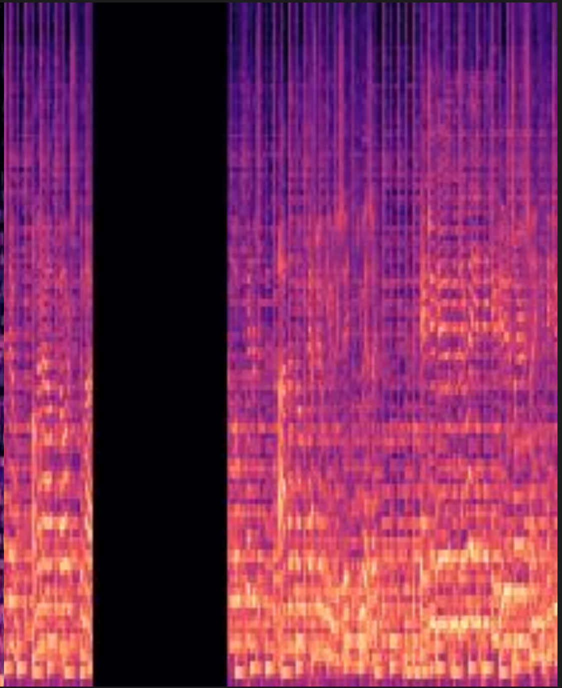
Left Image

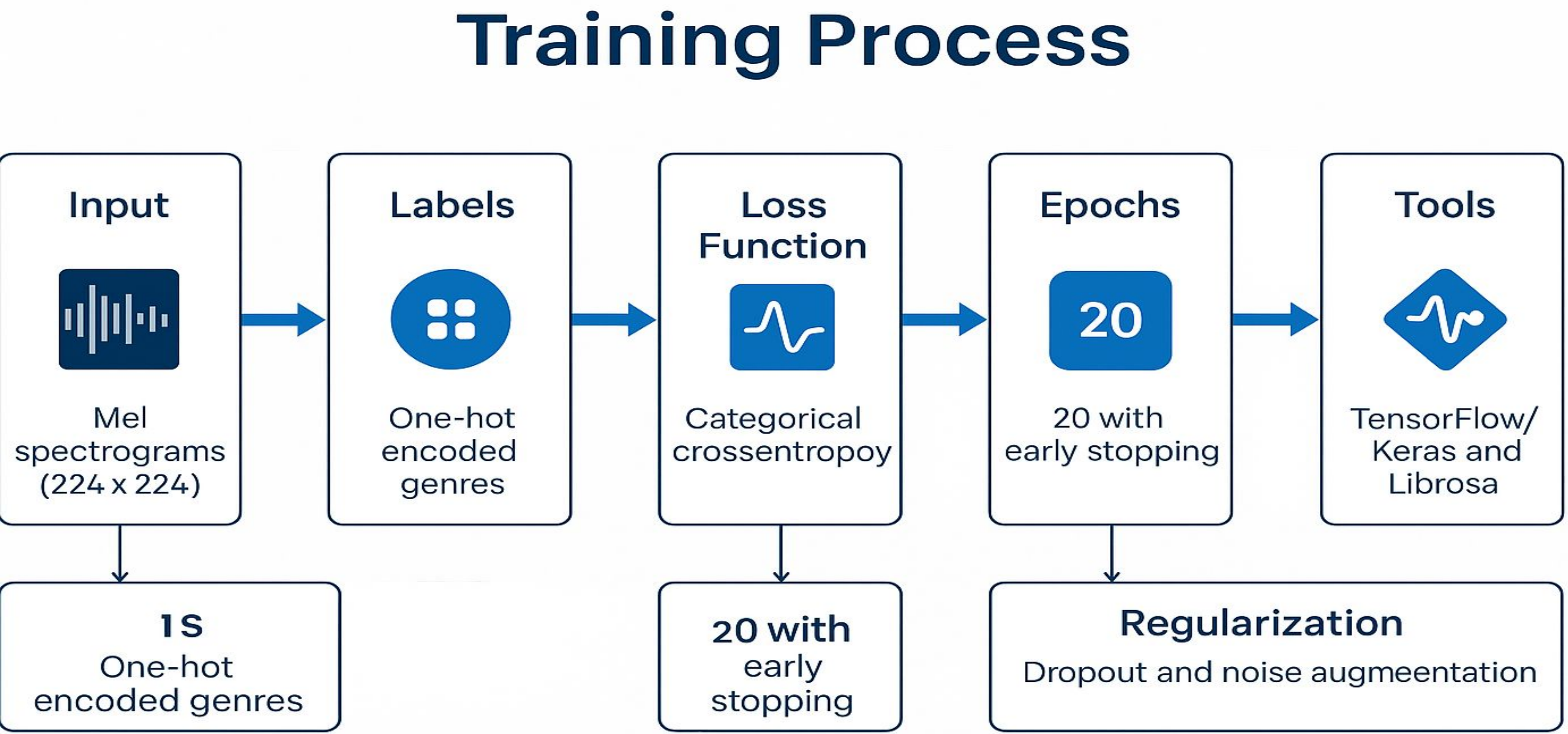
Original Mel Spectrogram for Country 00000

Right Image

Augmented Mel Spectrogram for Country 00000

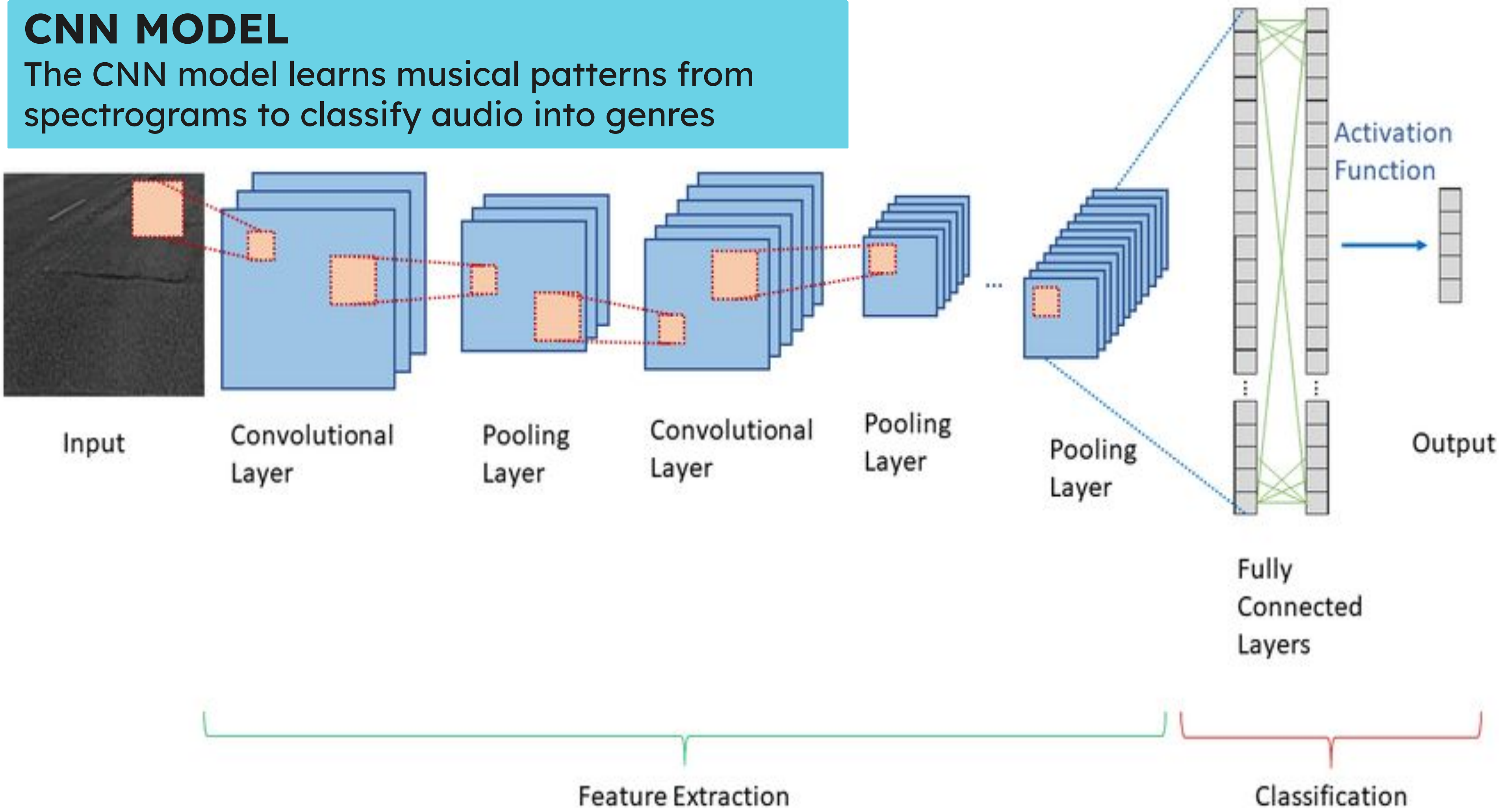





Training Process
Input
Mel spectrograms (224 x 224)
Labels
One-hot encoded genres
Loss Function
Categorical crossentropoy
Epochs
20
20 with early stopping
Tools
TensorFlow/ Keras and Librosa
1S
One-hot encoded genres
20 with early stopping
Regularization
Dropout and noise augmeentation

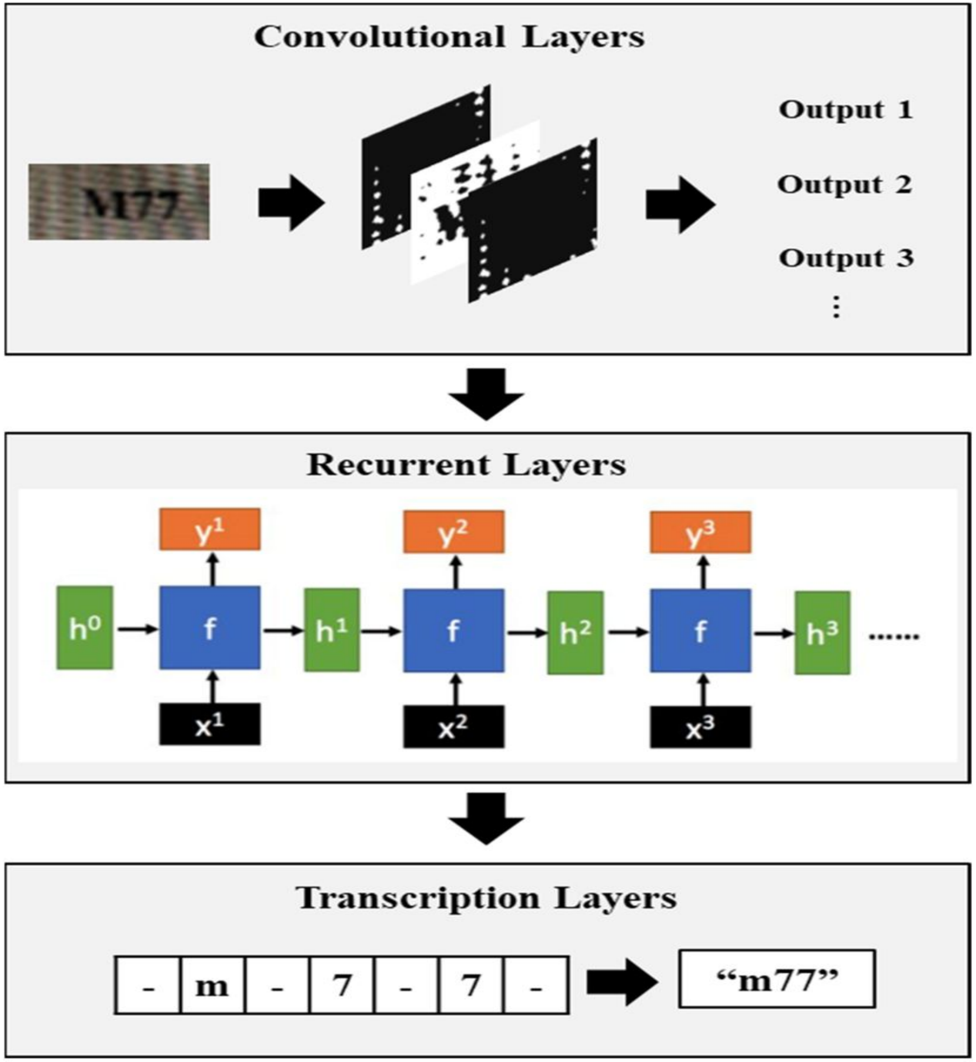
CNN MODEL

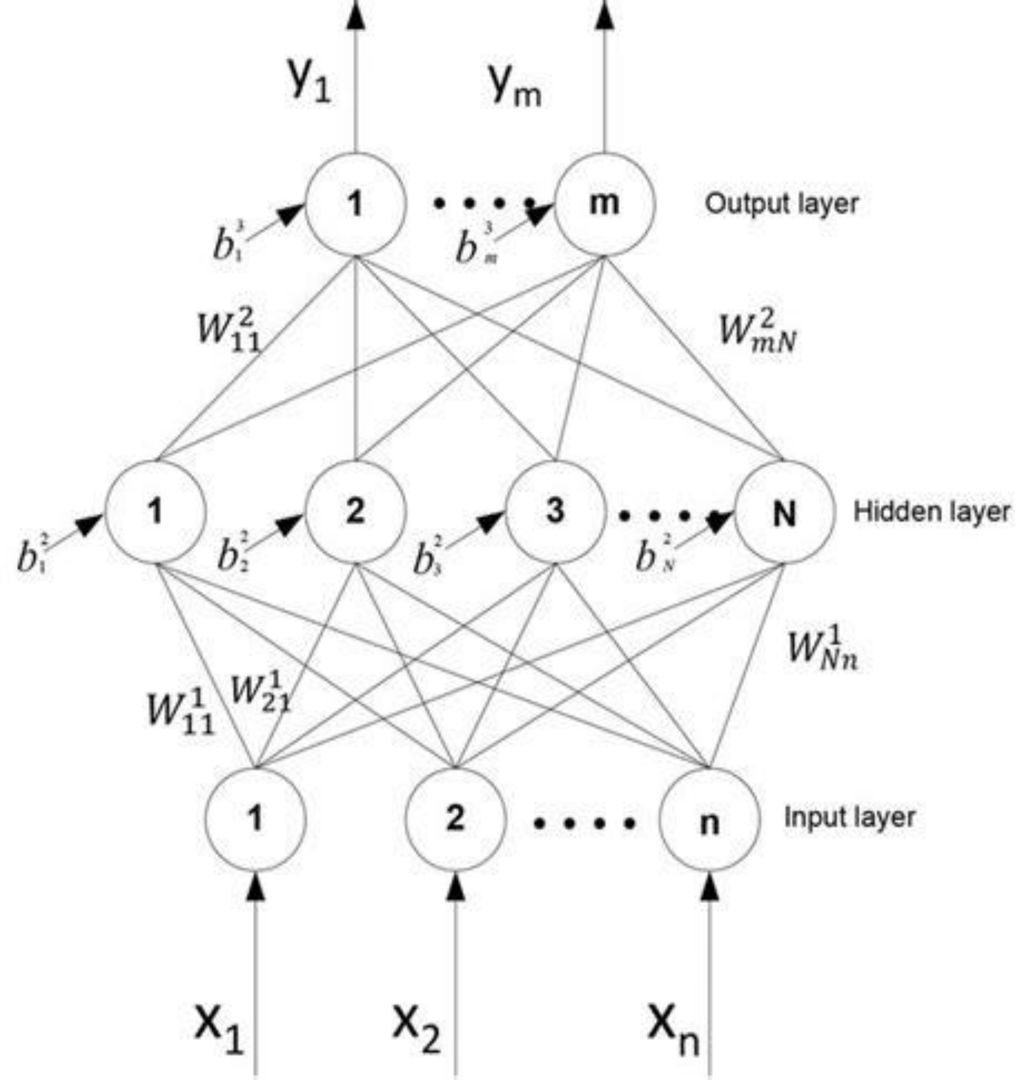
The CNN model learns musical patterns from spectrograms to classify audio into genres



CRNN MODEL

The CRNN model combine
convolutional layers for
feature extraction with
recurrent layers to capture
the temporal flow of music
making it well-suited for
genre classification



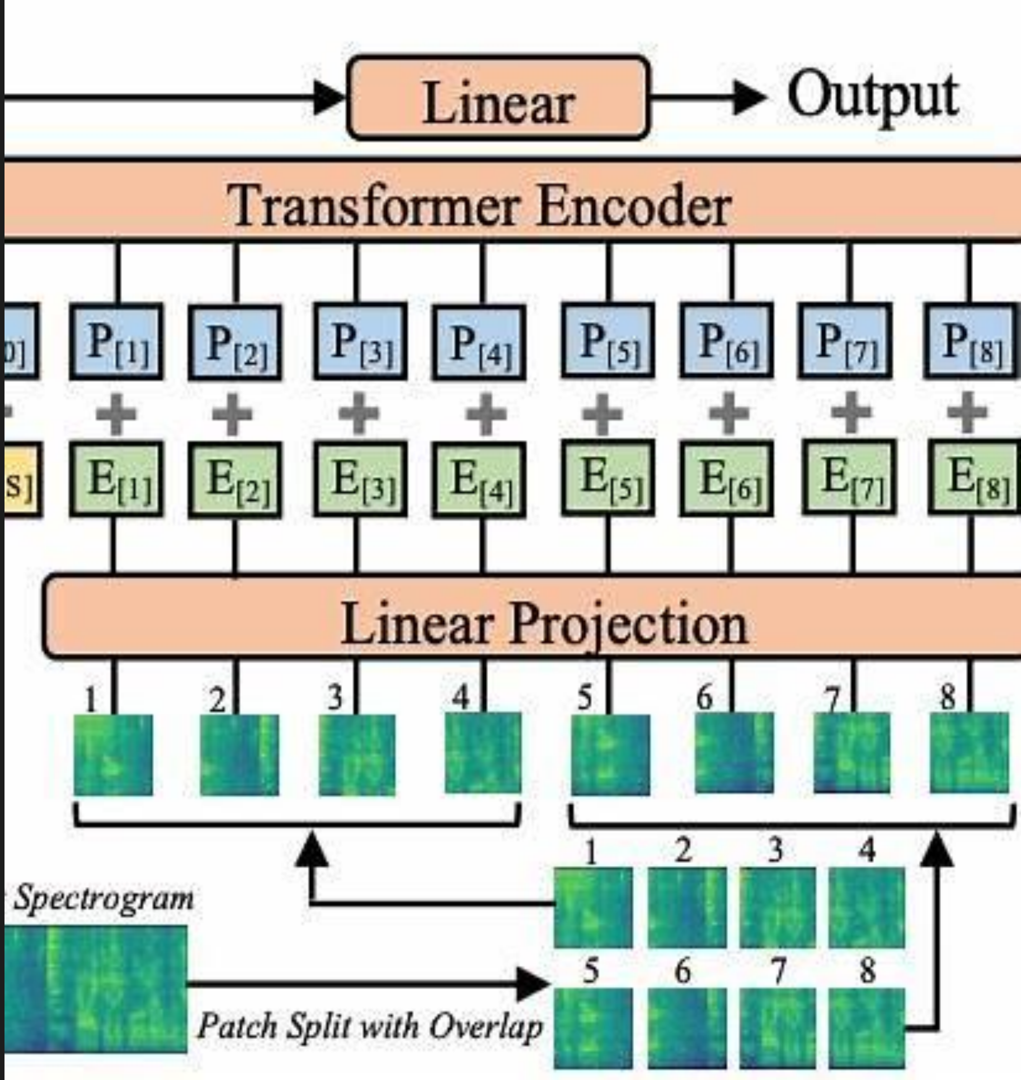


MLP MODEL

Our MLP model learns from flattened spectrogram features using fully connected layers to classify music into genres.

AST MODEL

The AST model leverages self-attention mechanisms to analyze spectrogram patches, capturing both local and global audio patterns for accurate genre classification

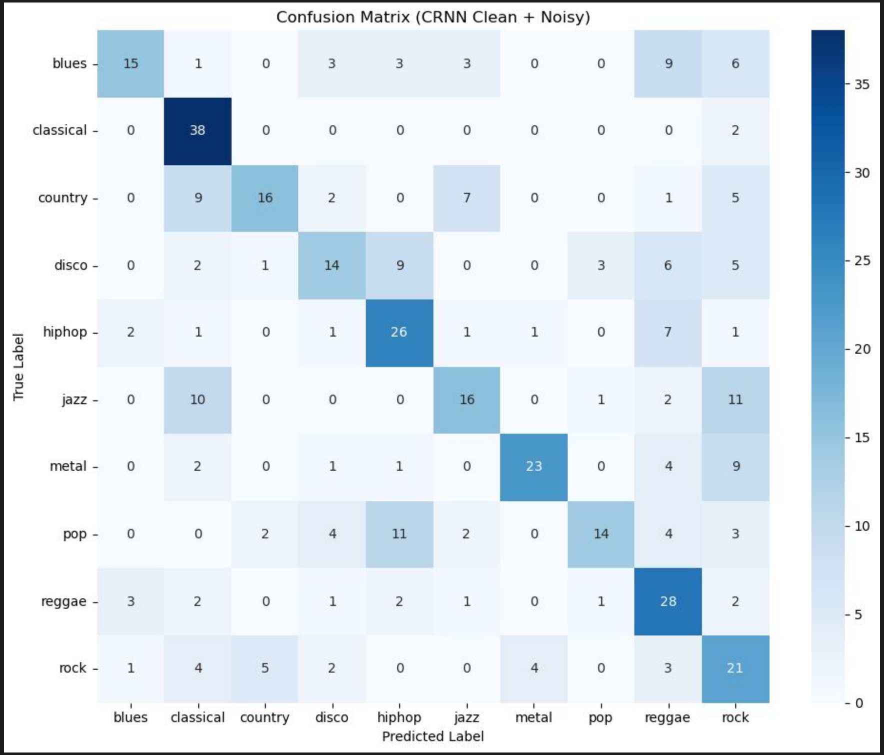


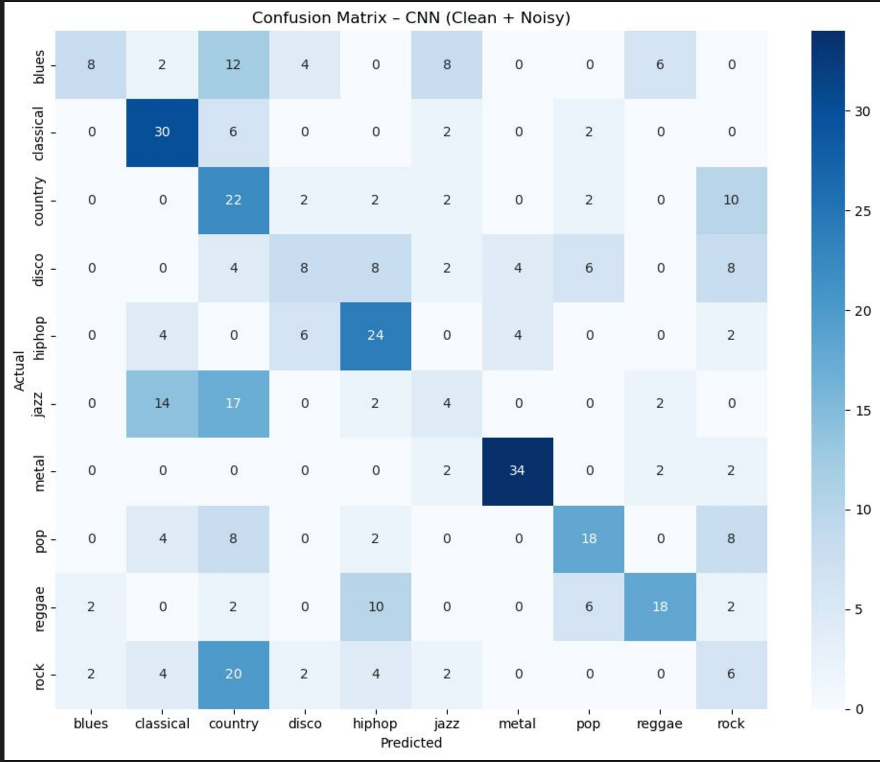
PERFORMANCE

$$\frac{2 \text{ Precision x Recall}}{\text{Precision + Recall}}$$

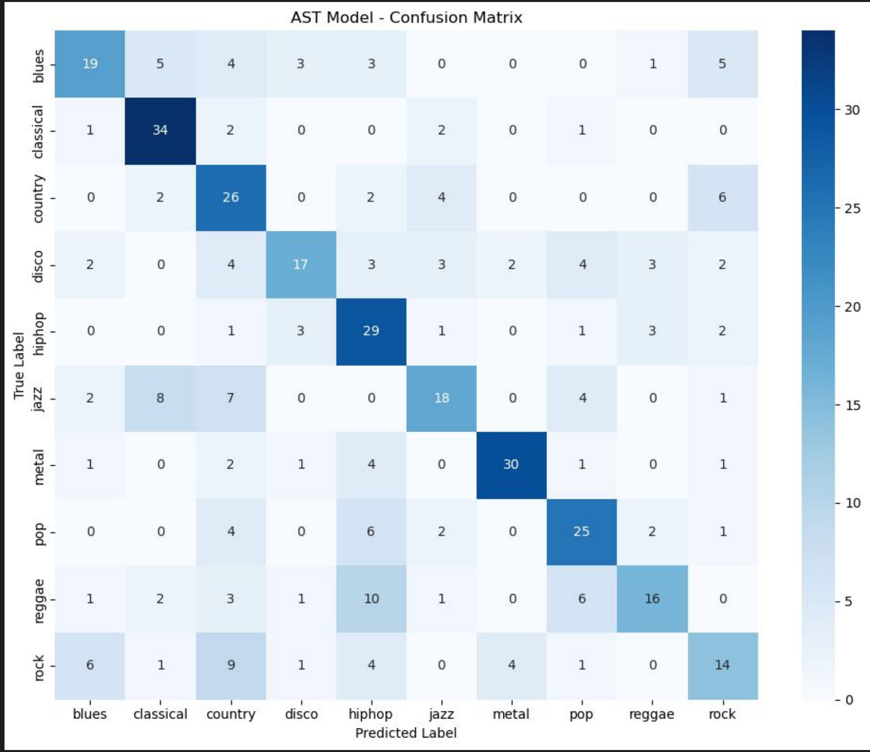
Model	Accuracy	Macro F1	Weighted F1
CNN	0.53	0.52	0.52
CRNN	0.67	0.64	0.65
MLP	0.63	0.58	0.59
AST	0.72	0.7	0.7

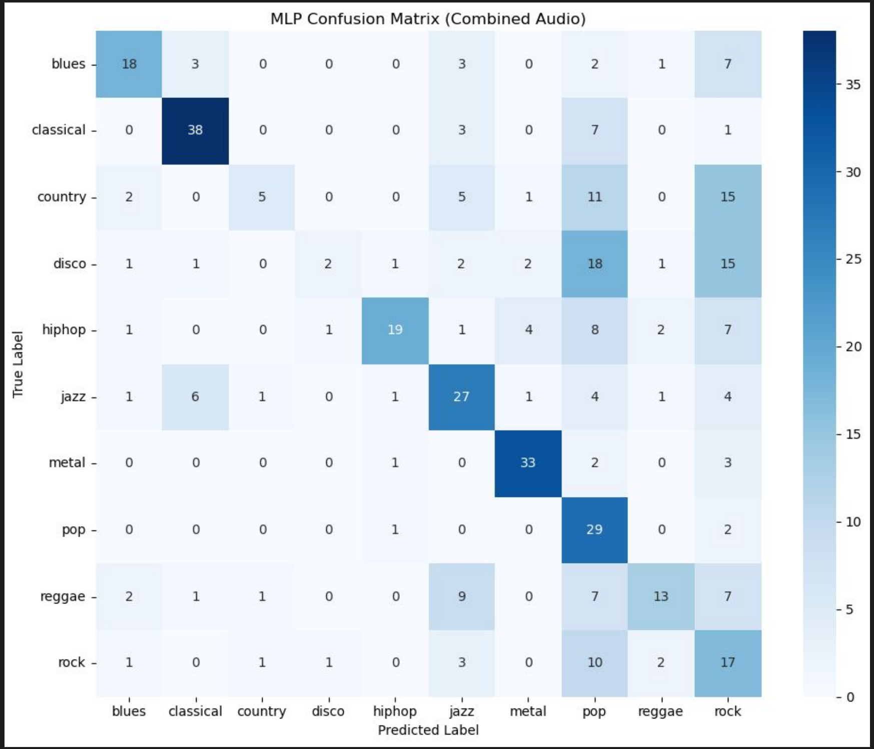
CNN & CRNN CONFUSION MATRIX





AST & MLP CONFUSION MATRIX





ISSUES

Genre Overlap & Ambiguity

Similar Sounding Genres were often Misclassified

- Disco & Pop
- Rock & Country

Data Imbalance

Some Genres had more consistent learning patterns, biasing other genres

Sensitivity to Noise in Simpler Models

CNN & MLP struggled with noisy inputs due to lack of temporal or attention mechanisms

Limited Dataset Size

Models such as AST were prone to overfitting due to lack of audio samples.

CONCLUSION

Temporal modeling improves performance

Accuracy was limited by overlapping genre characteristics

Self-attention enhances pattern recognition

Future work includes ensemble models and data enhancement

Q&A

?