

BirdSong Sonatas

MIDS 207 Final Project

Amina Alavi
Hamsini Sankaran
Rachel Gao
Andrew Loeber



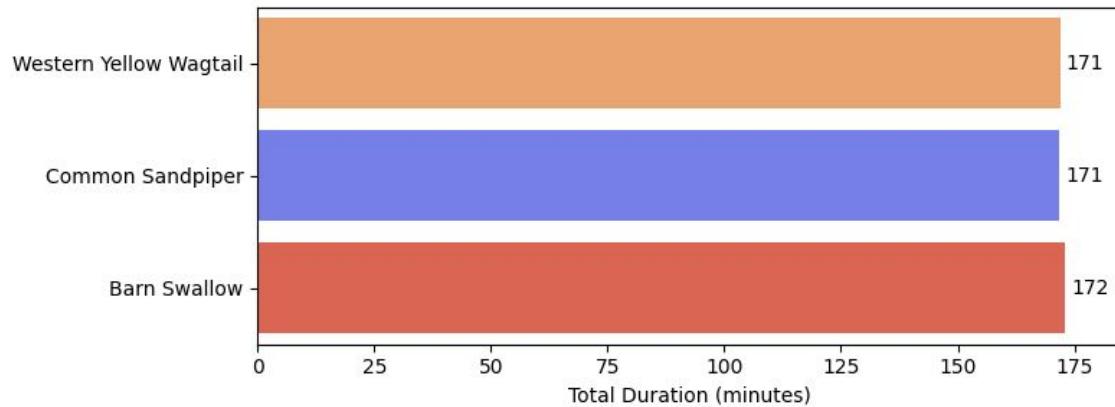
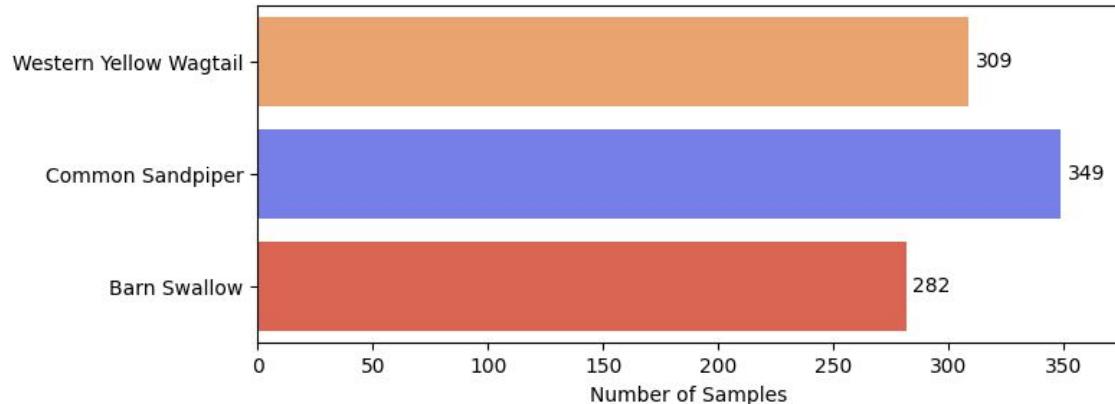
Objective & Dataset

Objective & Dataset

Objective	Bird Species Classifier Based on Bird Calls
Dataset	BirdCLEF 2023 on Kaggle
Labels	Bird Species
Features	Primary: Short recordings of individual bird calls Secondary: Geo location, Quality rating, etc

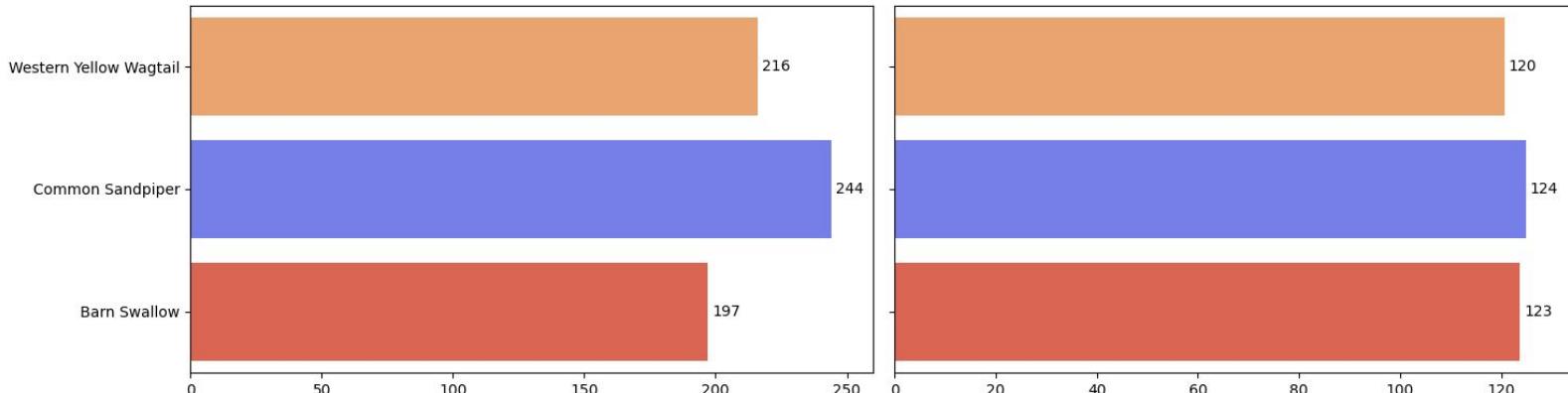
Preprocessing & EDA

Species Selection



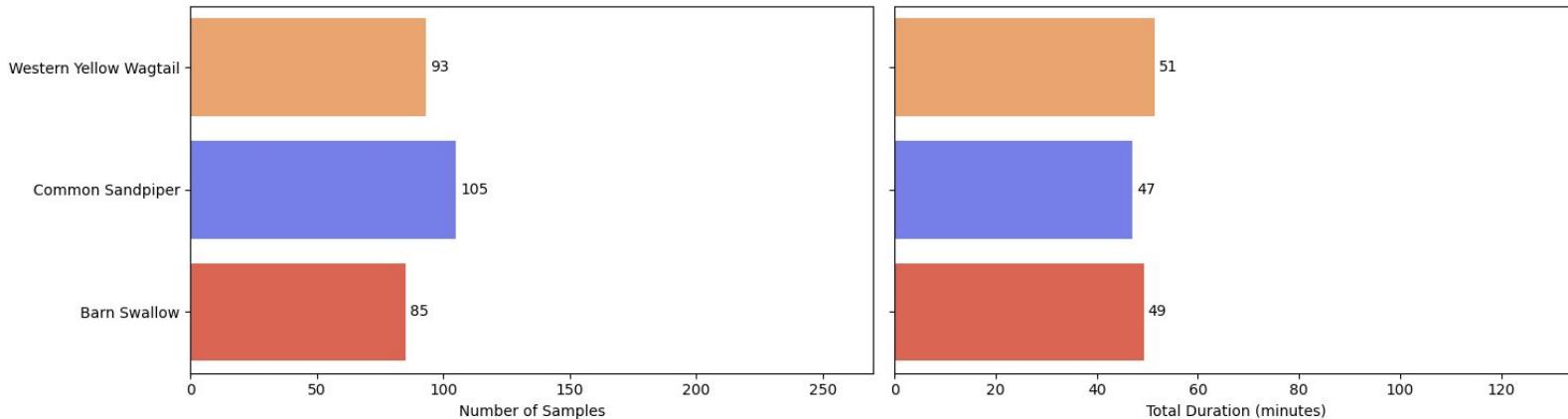
Train/Validation Split

Train



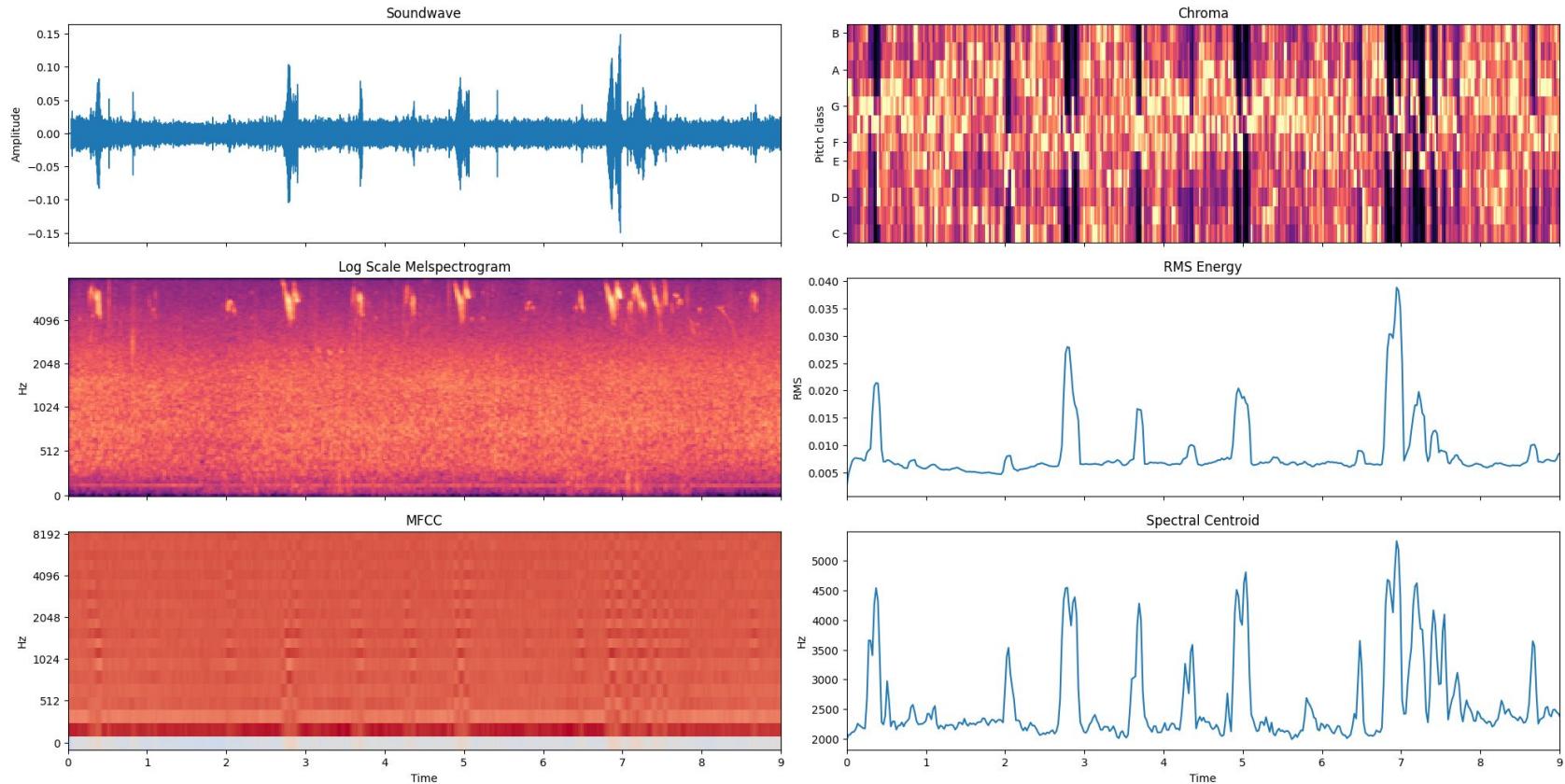
70%

Validation

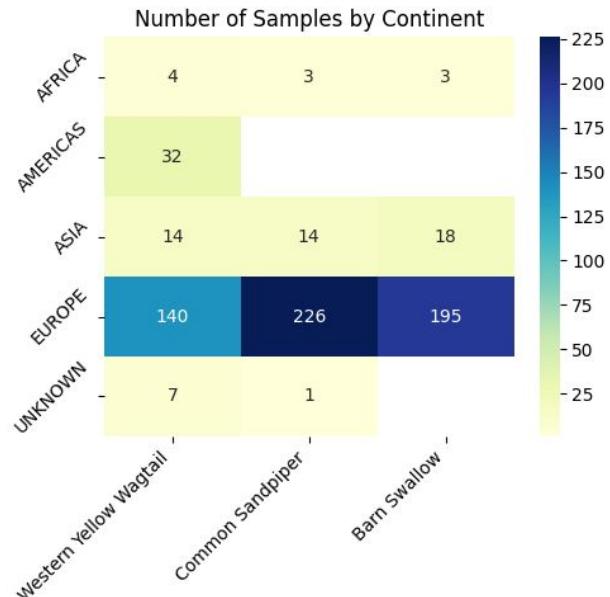
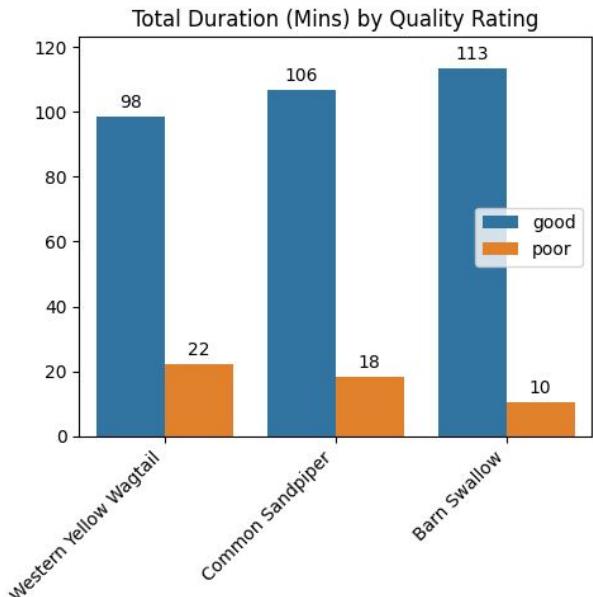
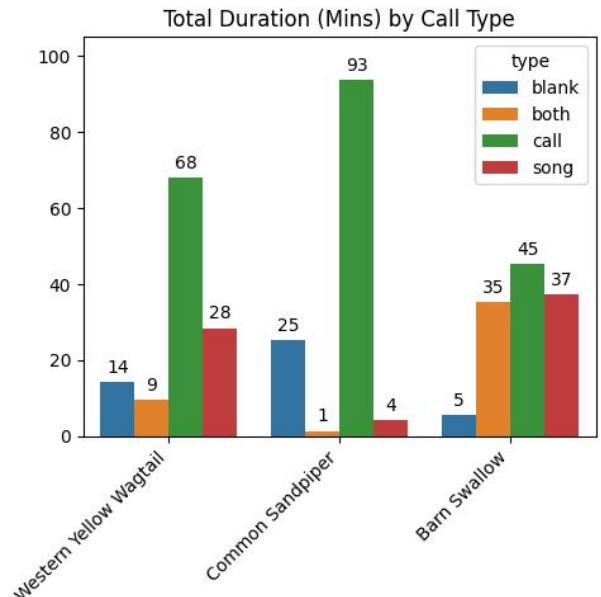


30%

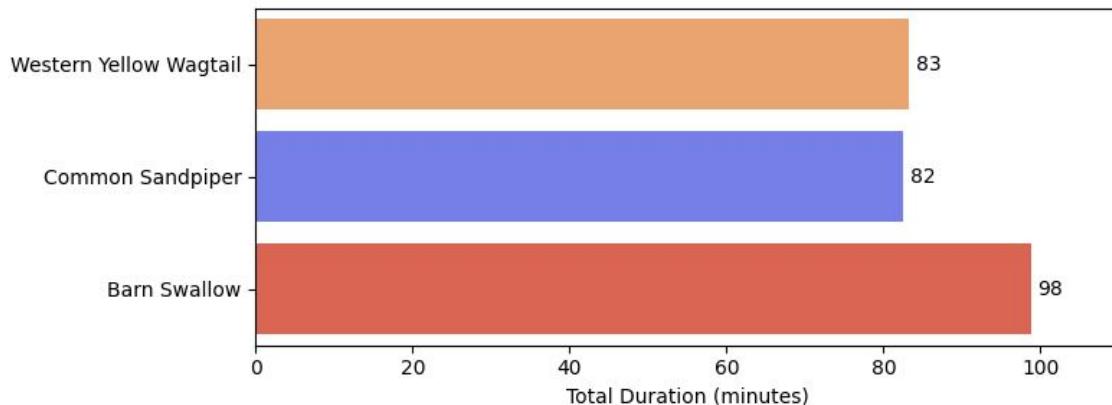
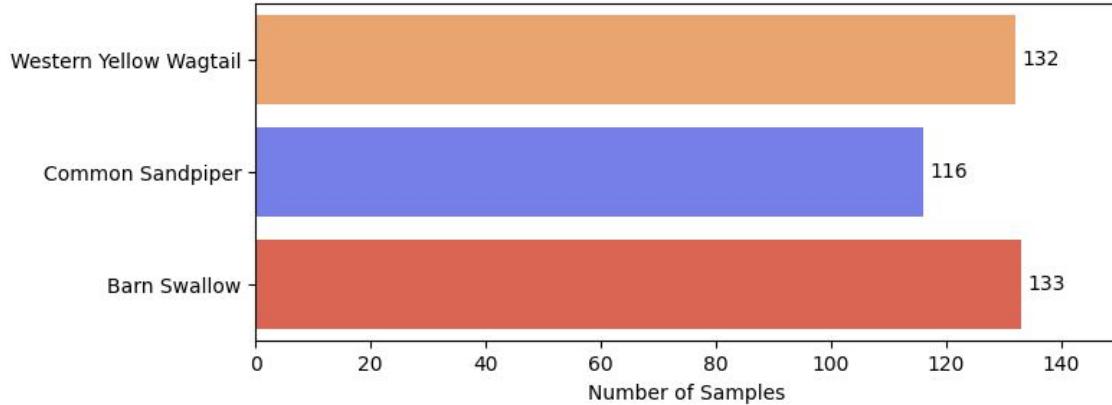
Audio Features



Additional Features

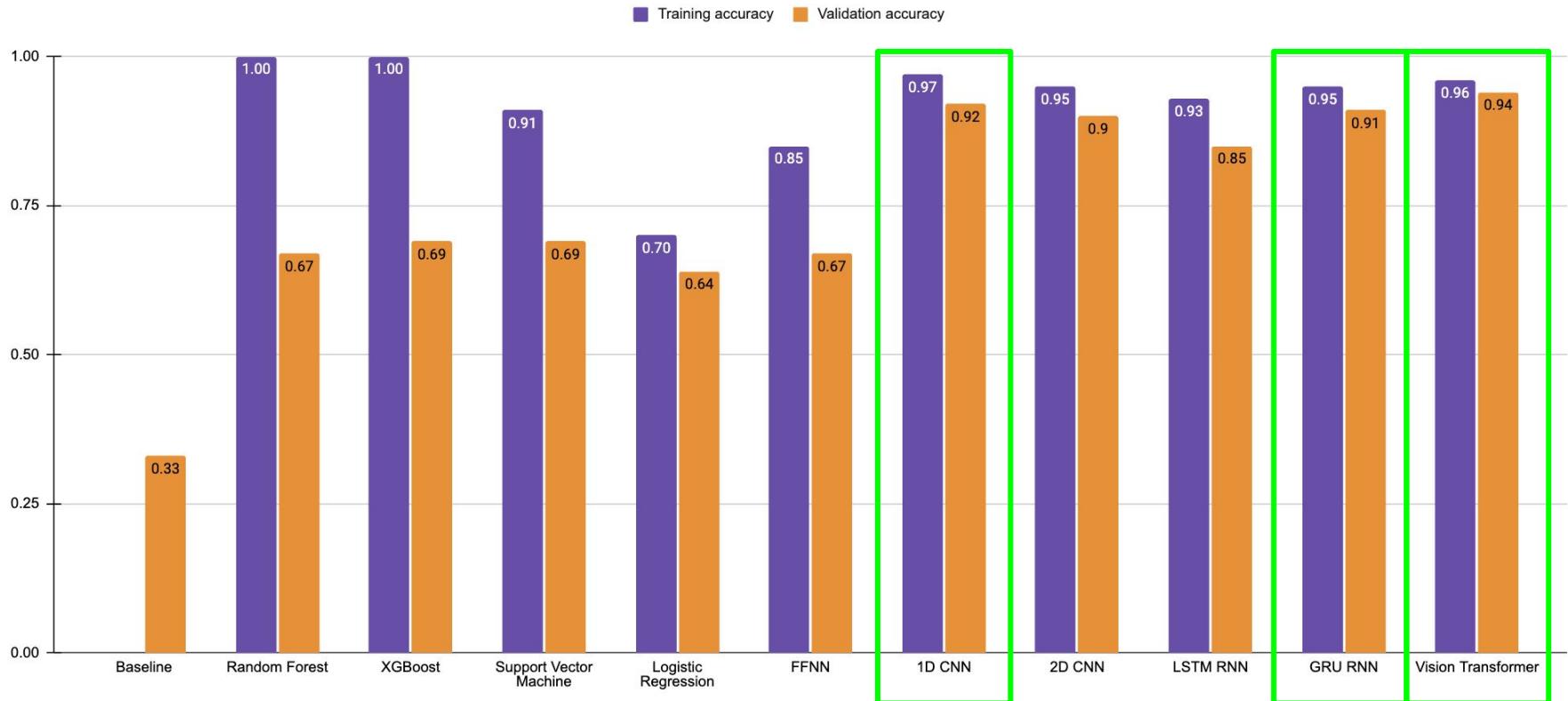


Test data



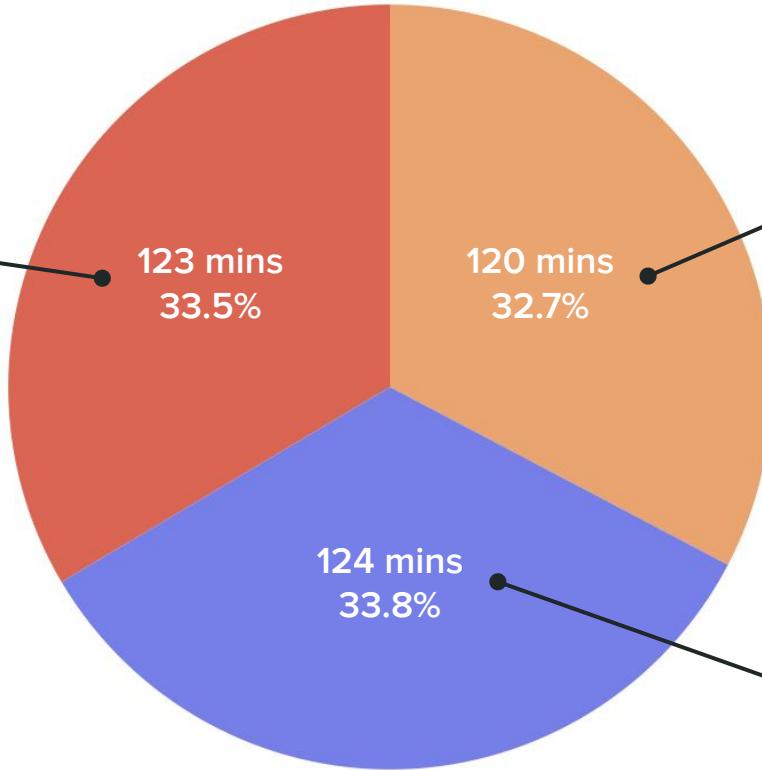
Machine Learning Models

Summary of Various Machine Learning Algorithms

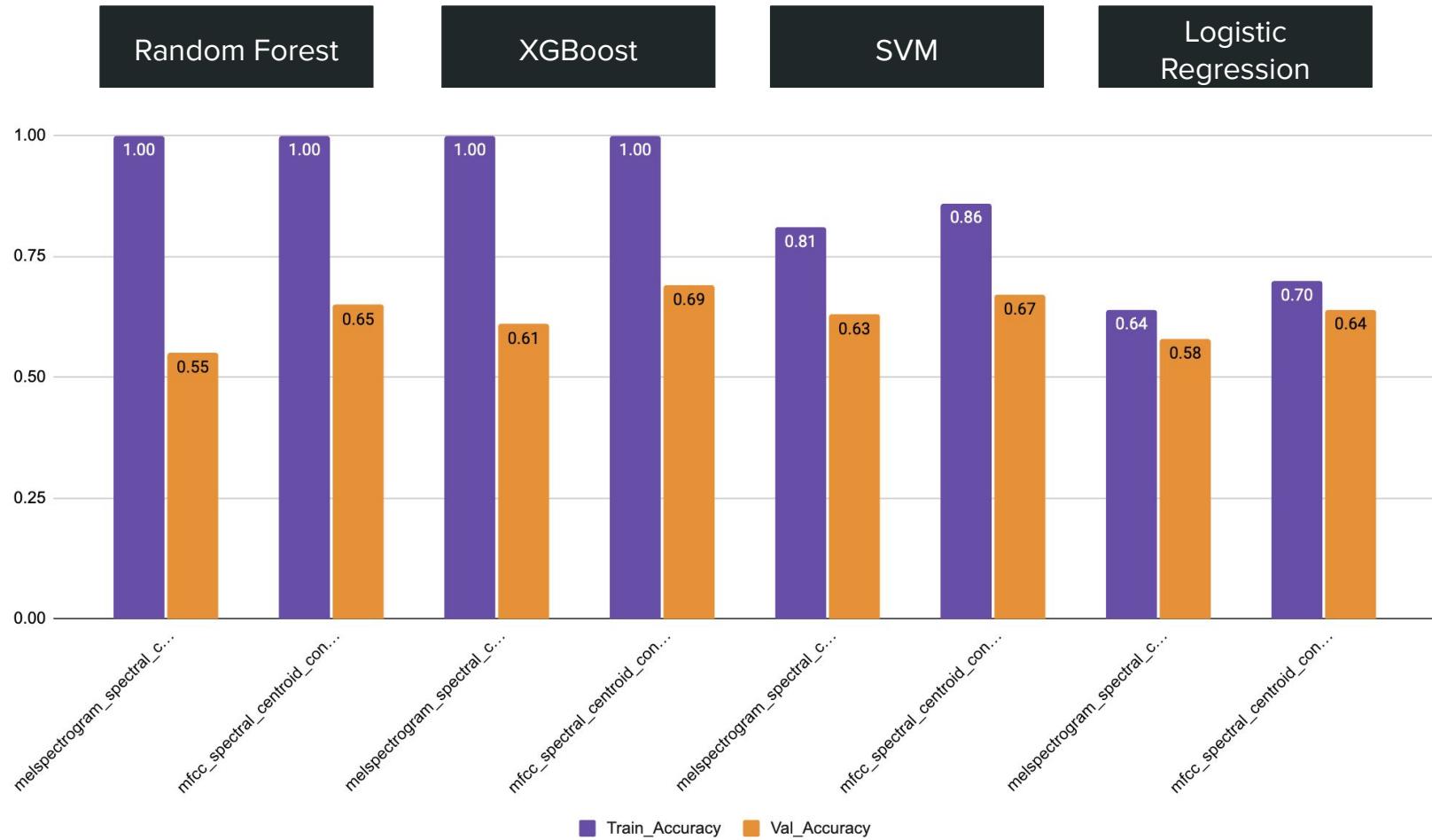


Baseline

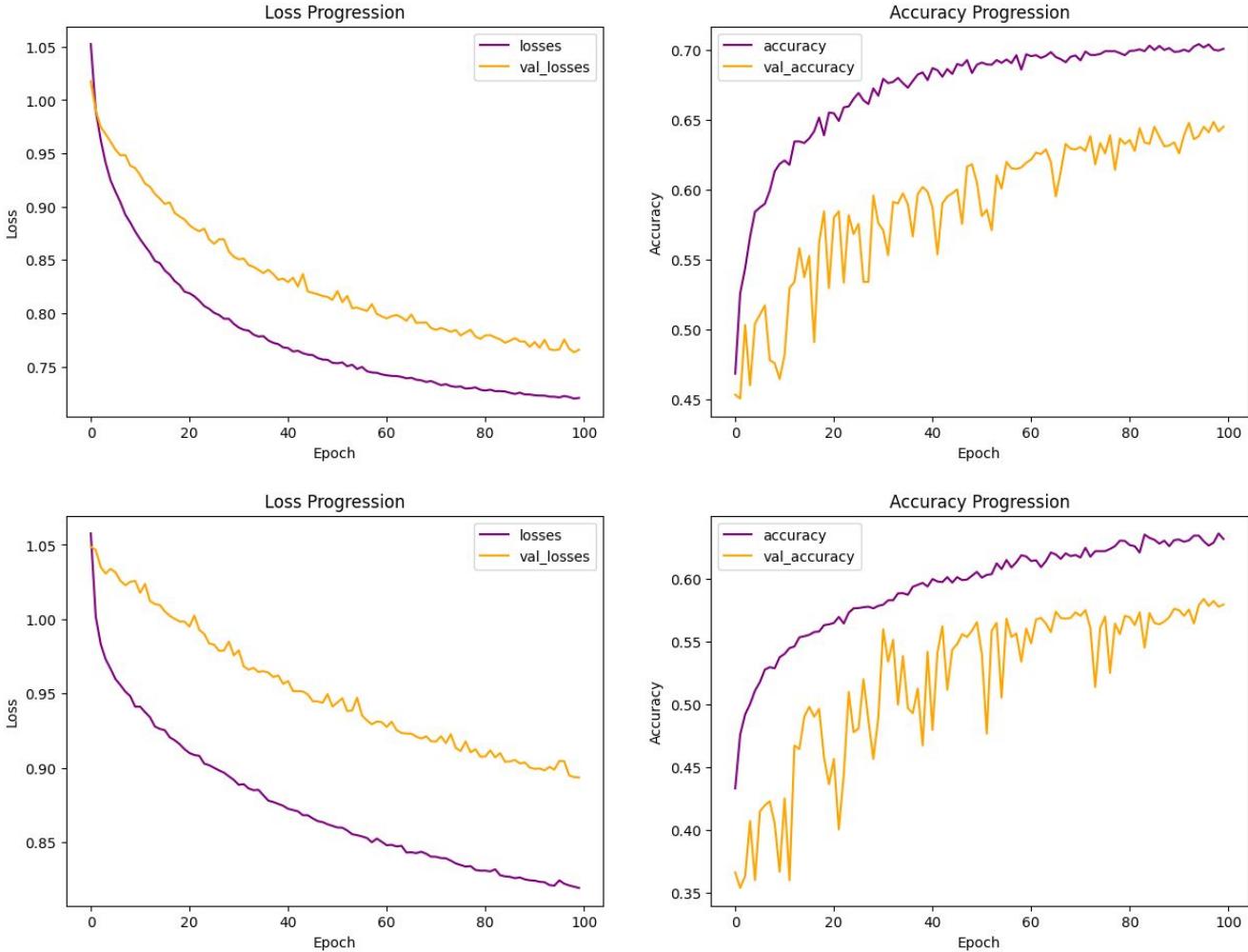
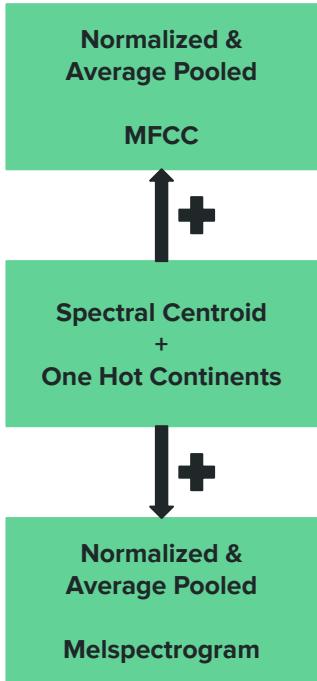
Random Guess - 33% overall accuracy



Traditional Algorithms



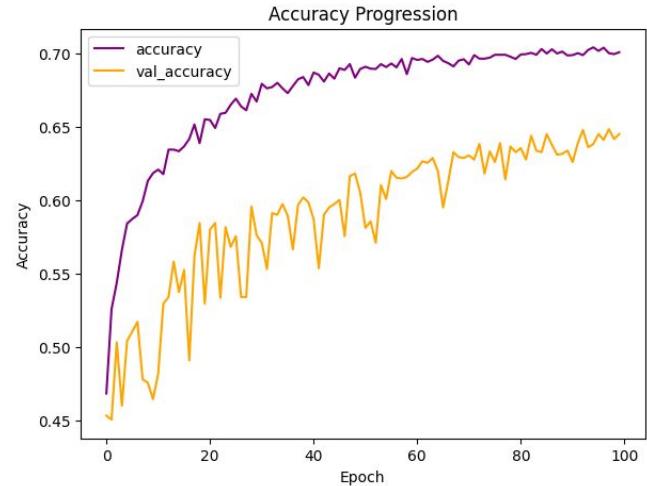
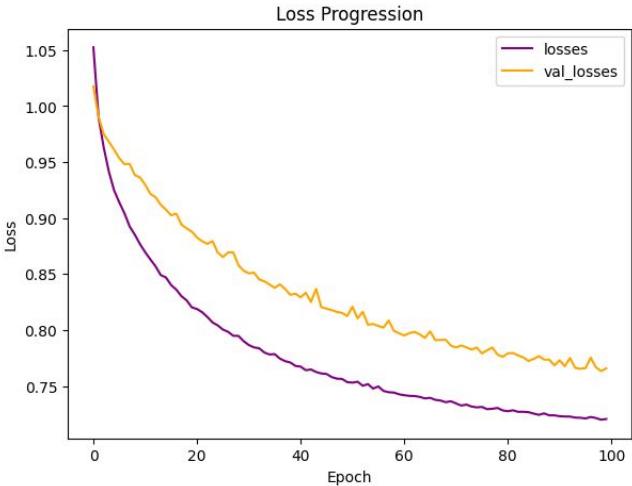
Logistic Regression



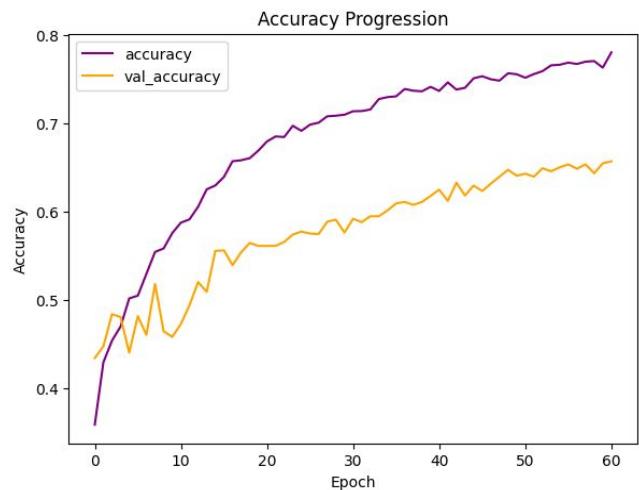
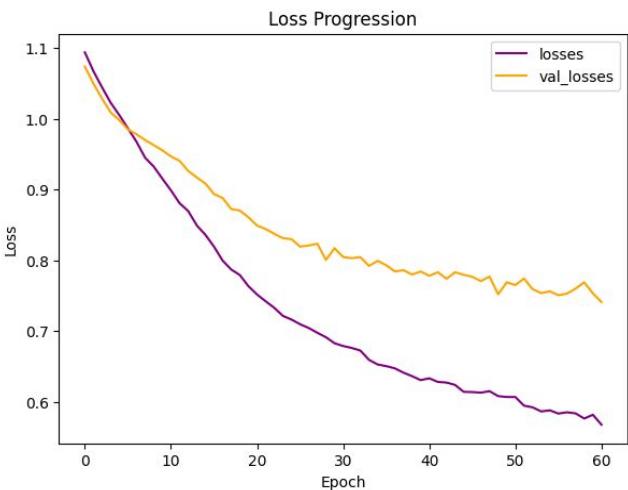
Deep Learning Algorithms

Logistic Regression

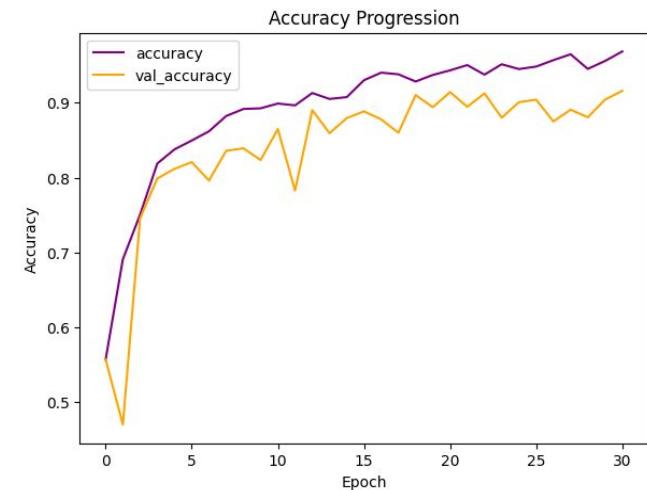
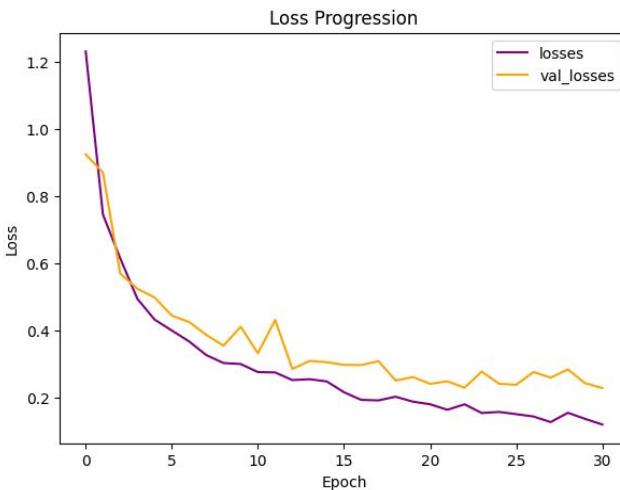
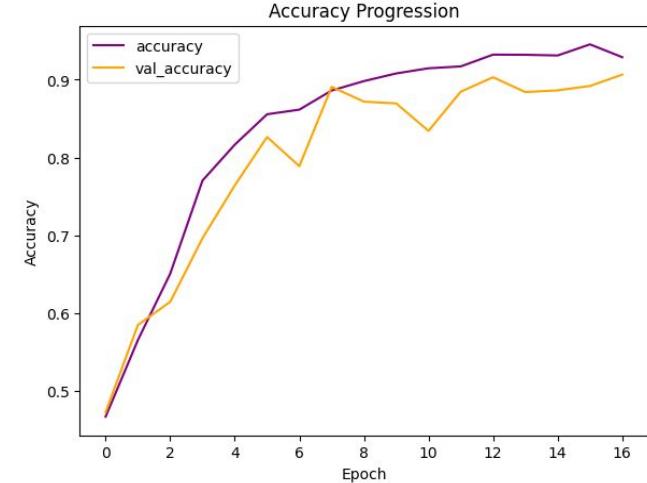
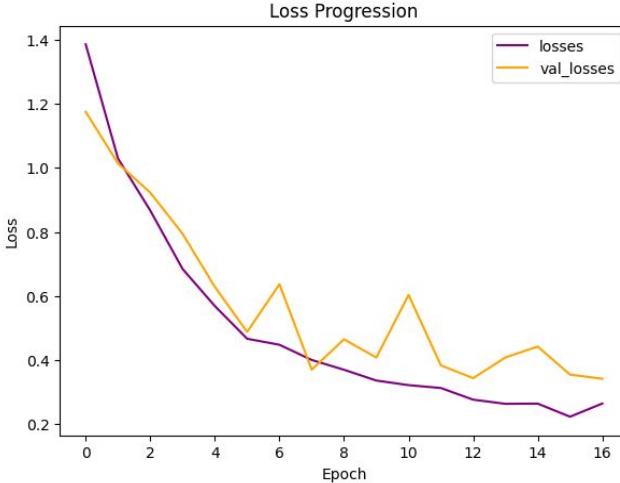
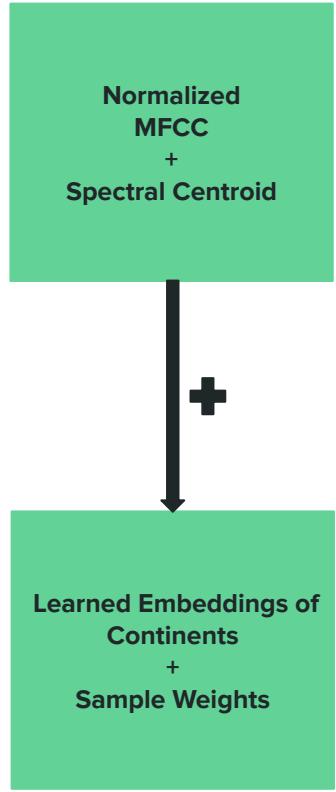
Normalized &
Average Pooled
MFCC
+
Spectral Centroid
+
One Hot Continents



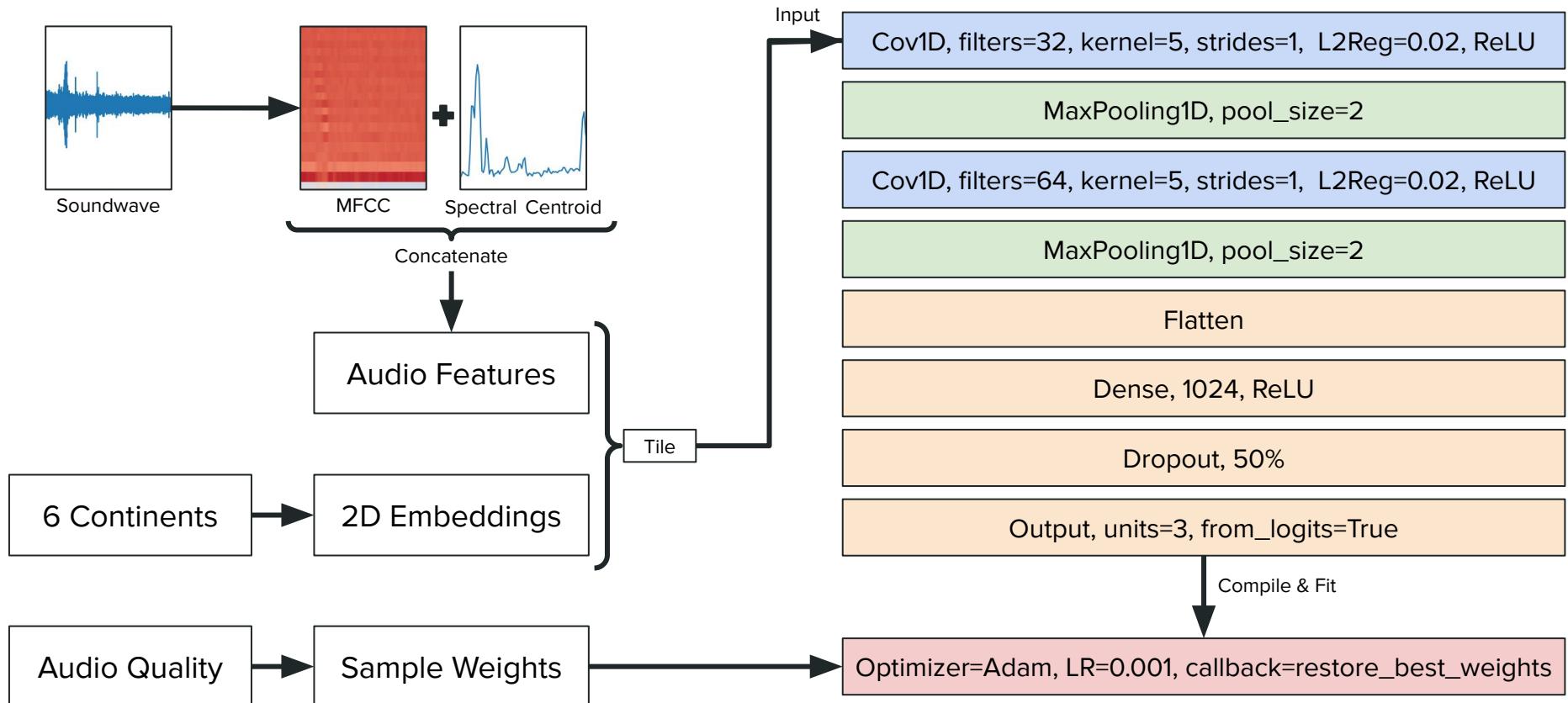
FFNN



1D CNN

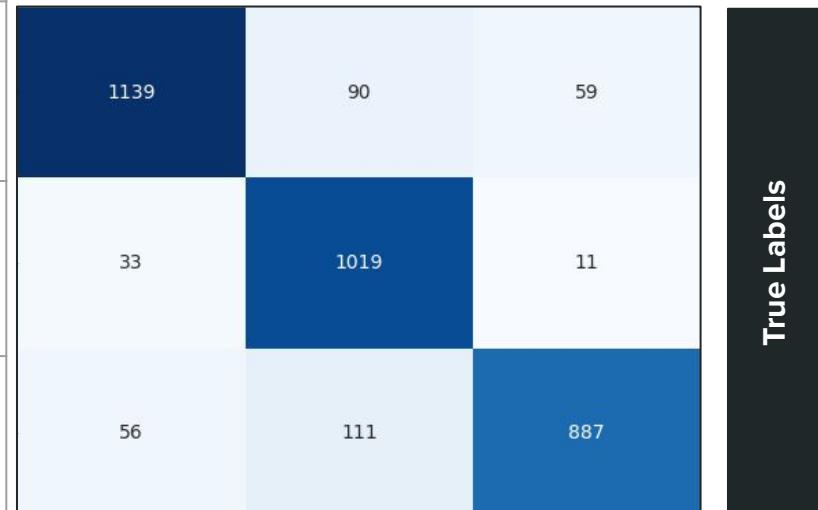


Architecture - 1D CNN



Inference results - 1D CNN

	Precision	Recall	F1-score	Support
Barn Swallow	0.93	0.88	0.91	1288
Common Sandpiper	0.84	0.96	0.89	1063
Western Yellow Wagtail	0.93	0.84	0.88	1054
Overall Accuracy	0.89			



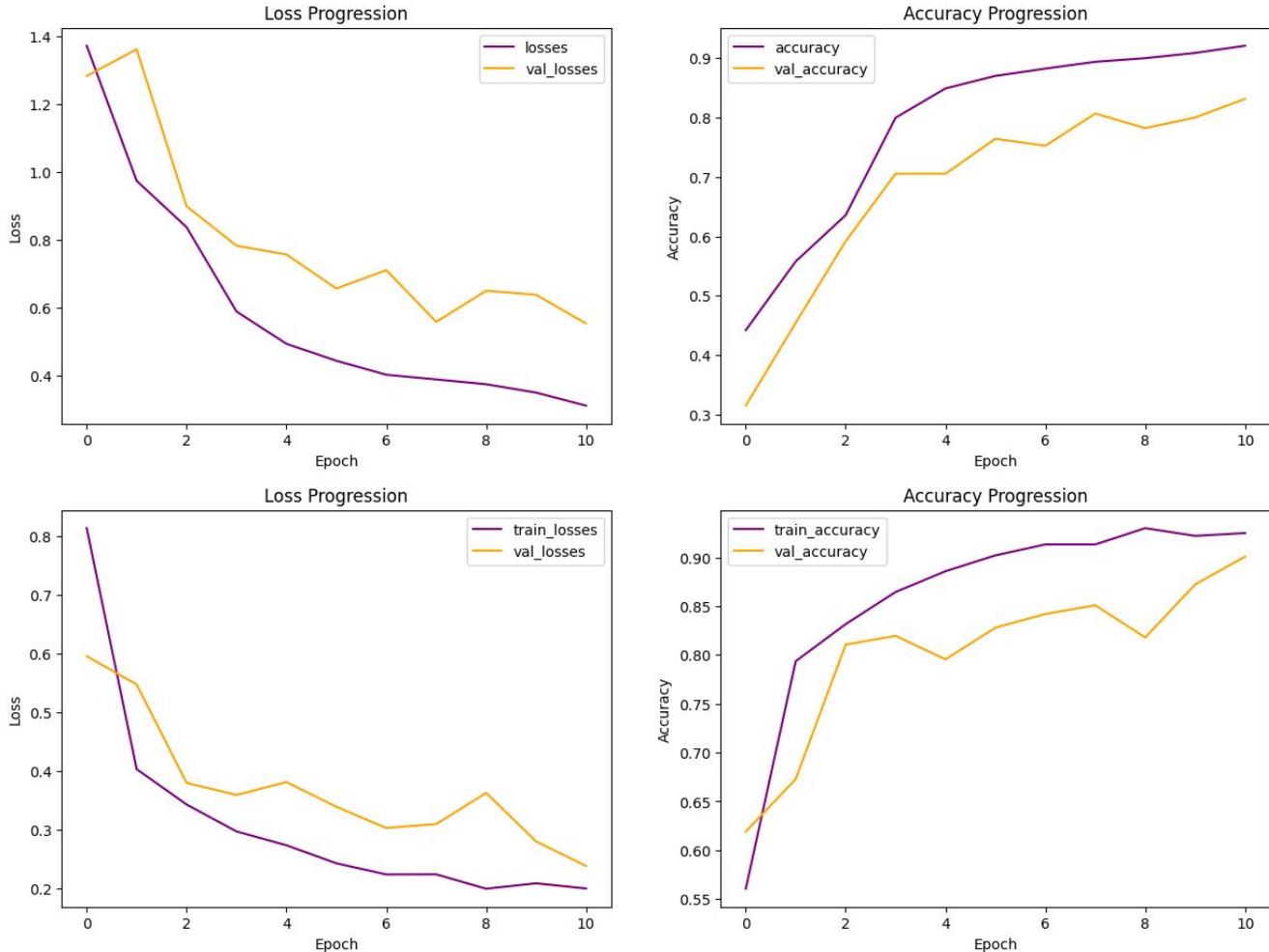
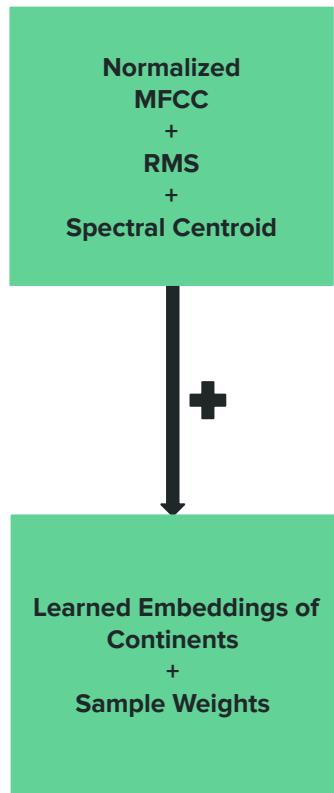
Predicted Labels

True Labels

Transfer Learning with 1D CNN Architecture

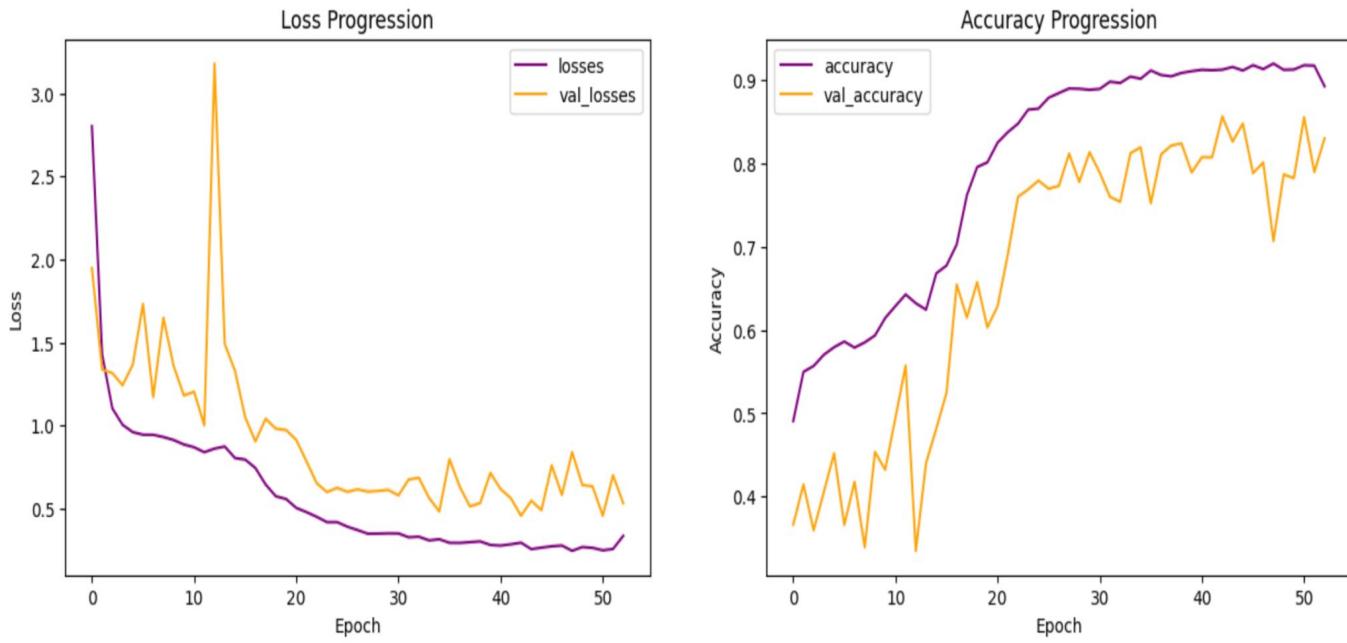


2D CNN

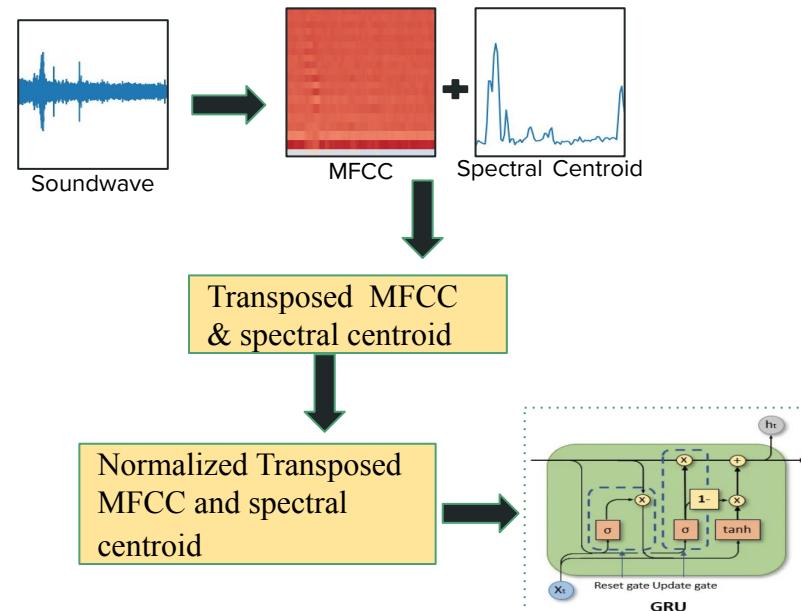


LSTM RNN

Normalized
MFCC
+
Chroma
+
Audio time stretch



Architecture - GRU RNN



Barn Swallow



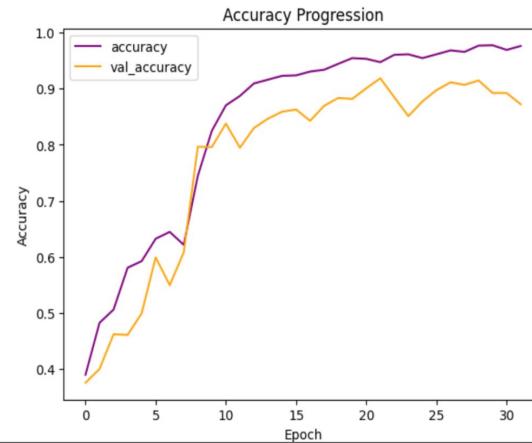
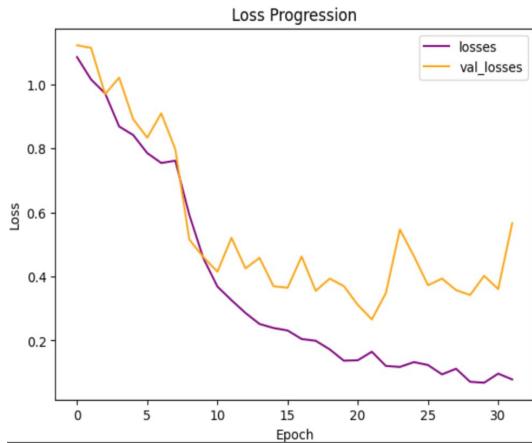
Common Sandpiper



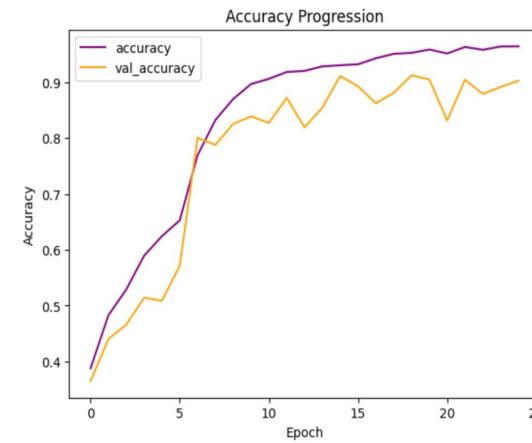
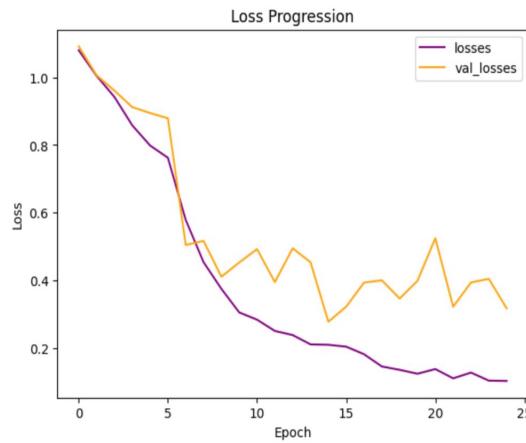
Western Yellow Wagtail

GRU RNN

Normalized
MFCC
+
Spectral Centroid
+
Audio time stretching



Normalized
MFCC
+
Spectral Centroid

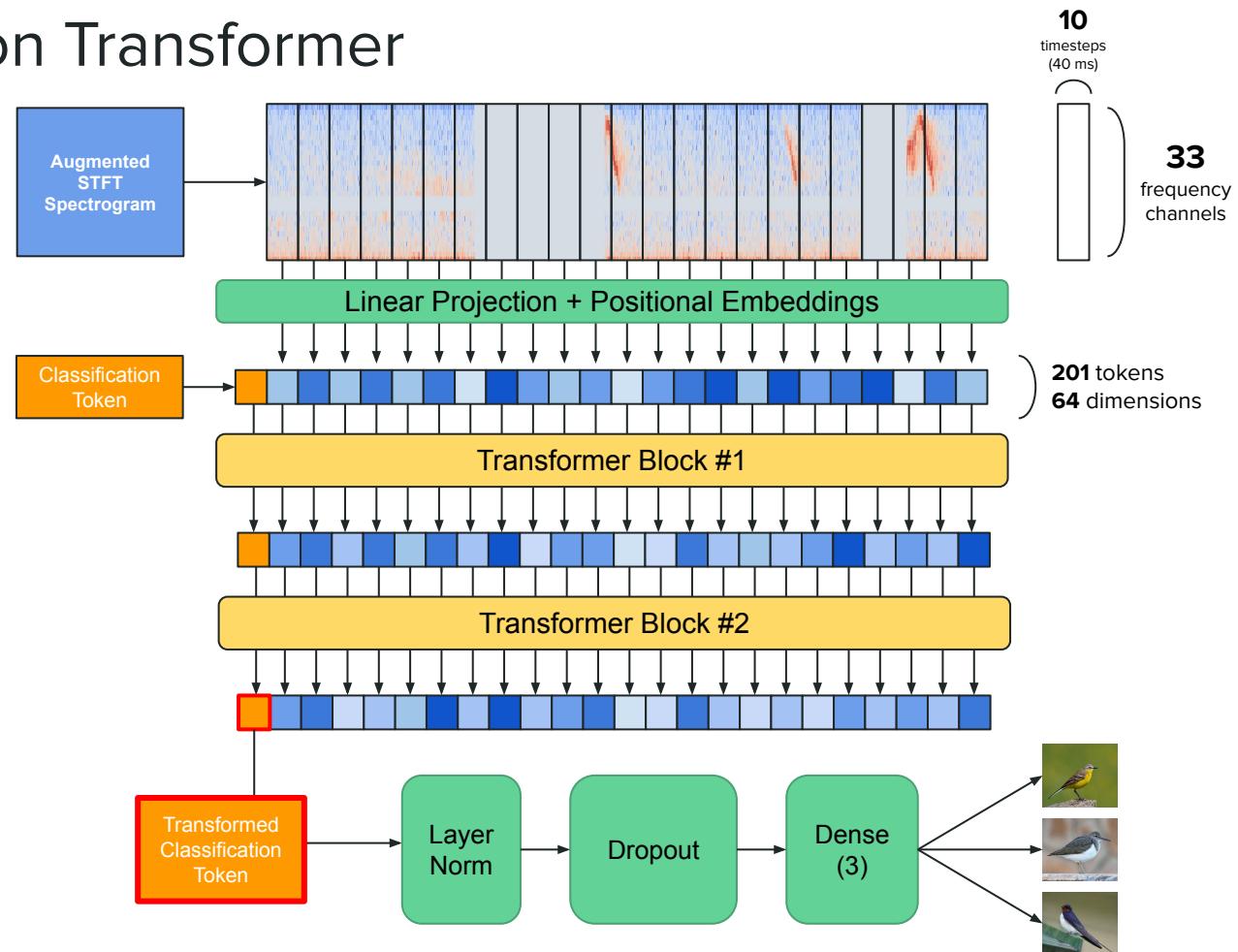
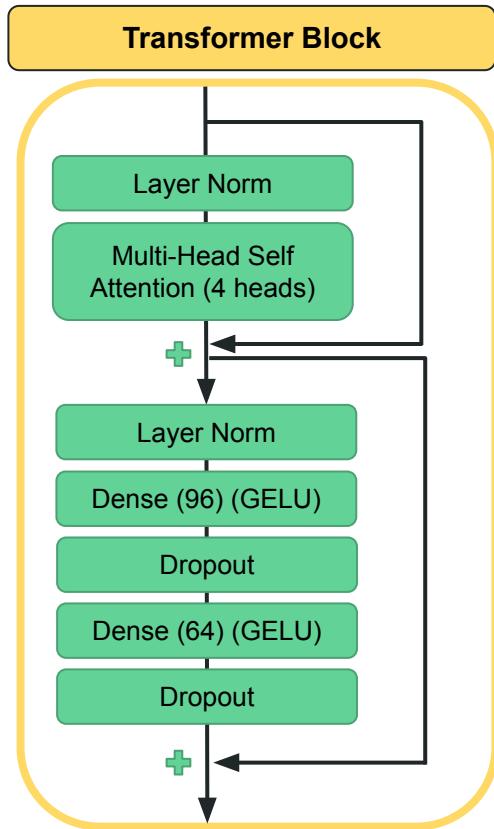


Inference results - GRU RNN

	Precision	Recall	F1-score	Support
Barn Swallow	0.89	0.83	0.86	1288
Common Sandpiper	0.80	0.93	0.86	1063
Western Yellow Wagtail	0.95	0.86	0.90	1054
Overall accuracy	0.87			

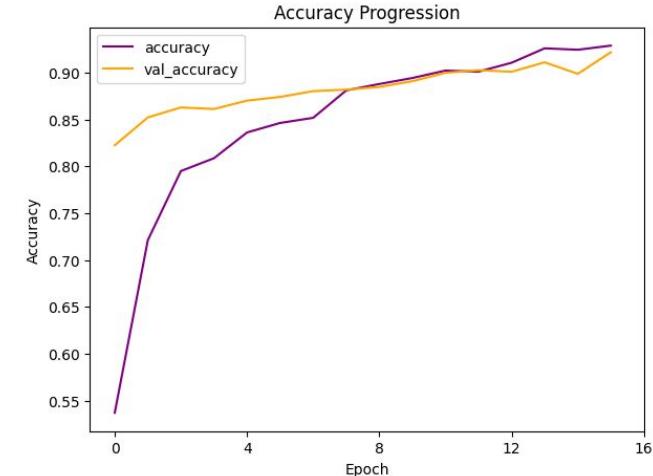
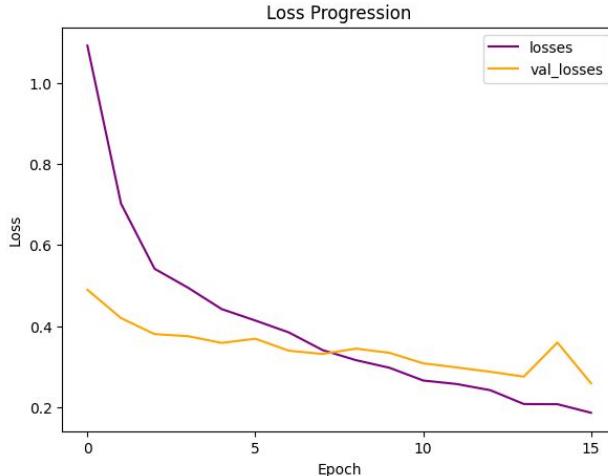


Architecture - Vision Transformer

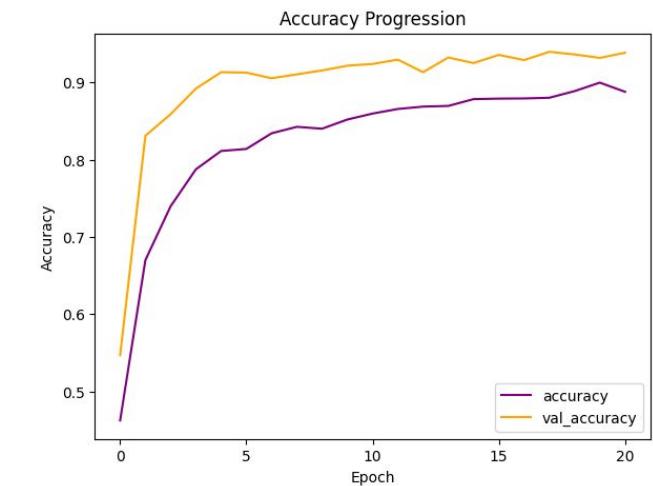
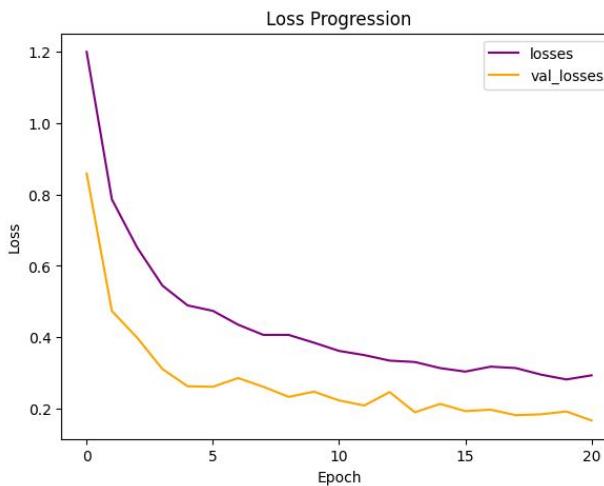


Vision Transformer

**Normalized
MFCC**
+
**Training-time
Spectrogram
Augmentation**

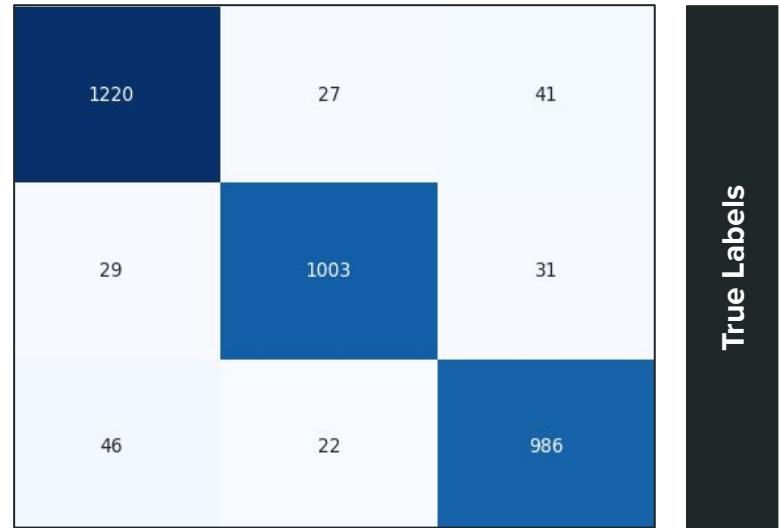


**Normalized
STFT**
+
**Training-time Audio &
Spectrogram
Augmentation**



Inference results - Vision Transformer

	Precision	Recall	F1-score	Support
Barn Swallow	0.94	0.95	0.94	1288
Common Sandpiper	0.95	0.94	0.95	1063
Western Yellow Wagtail	0.93	0.94	0.93	1054
Overall Accuracy	0.94			

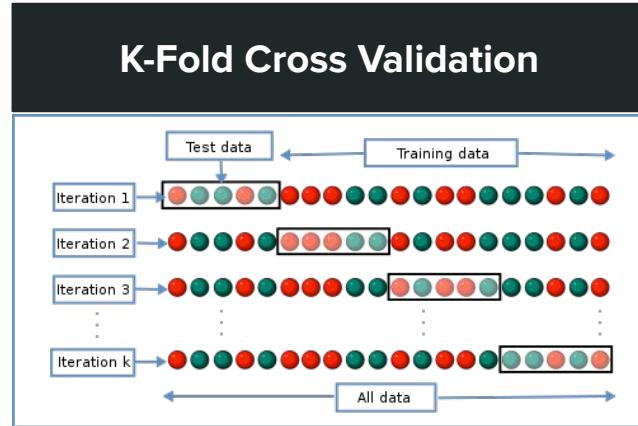


Predicted Labels

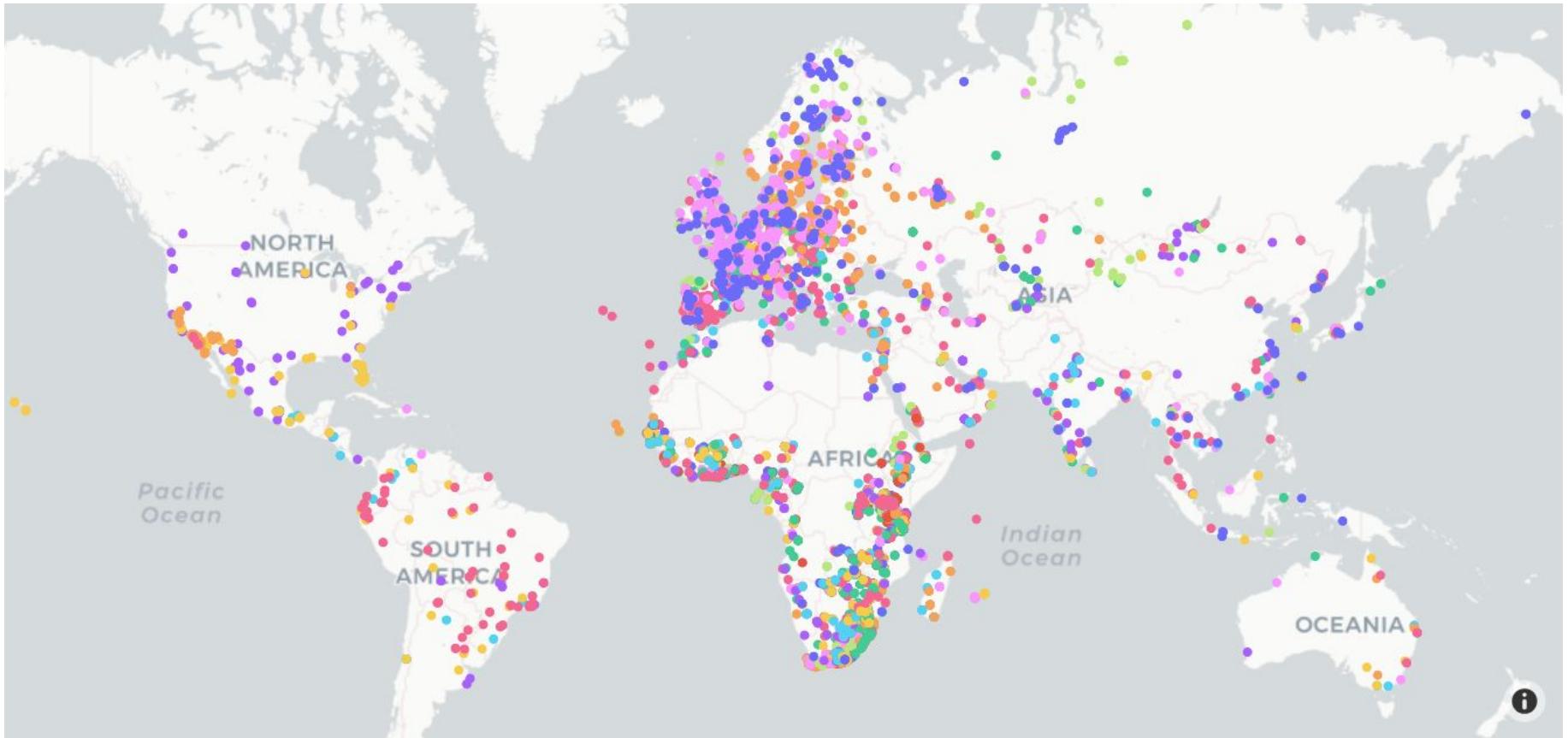
True Labels

Inspiration for future research

Looking Ahead



Fairness

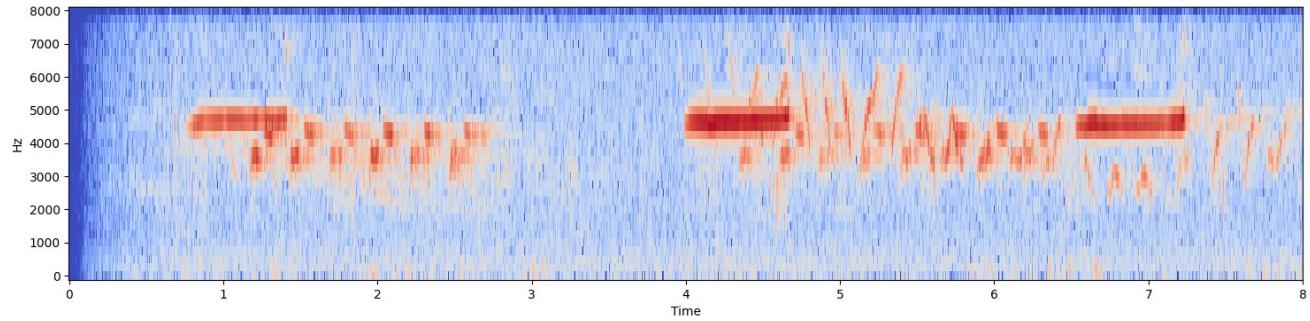


Thank you!

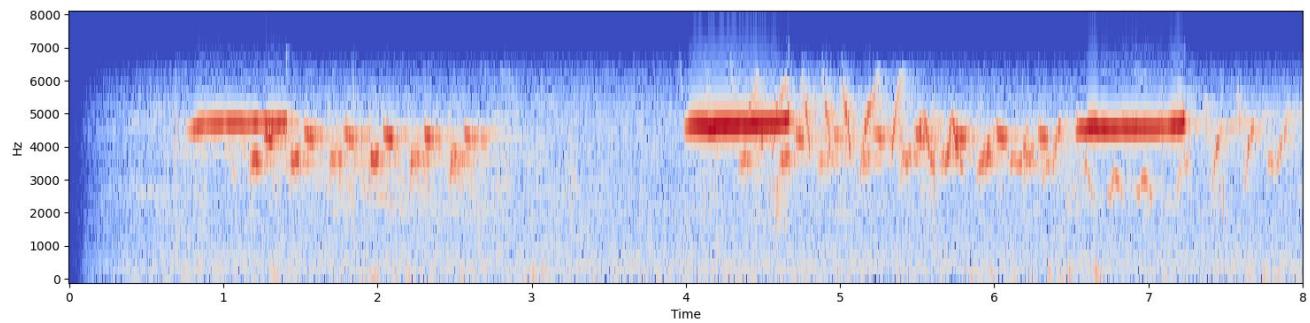
Link to Team GitHub Repo: https://github.com/rachelgaoMIDS/207_final_project

Augmentation Example

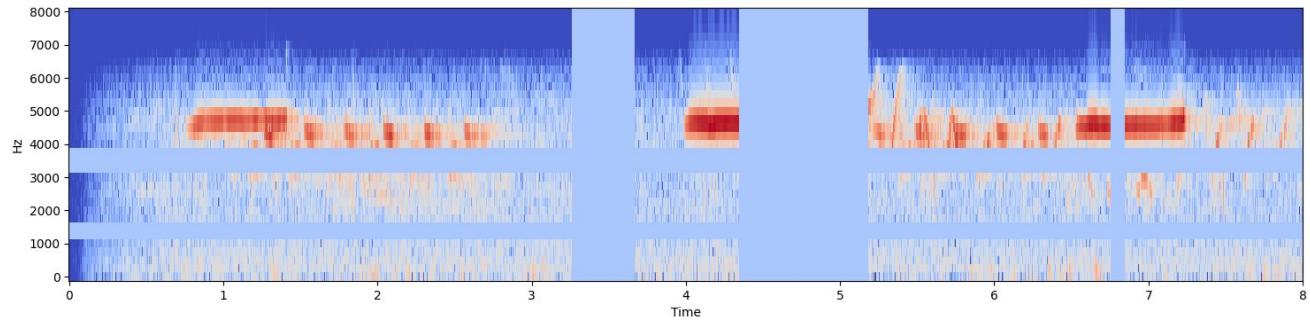
Original STFT



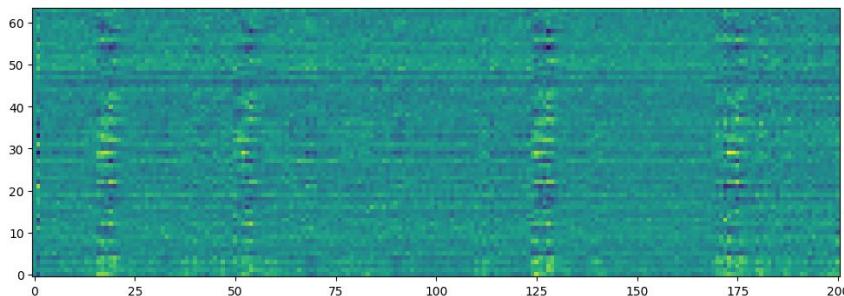
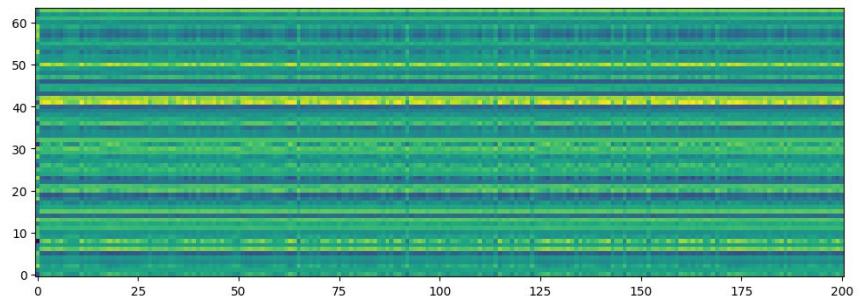
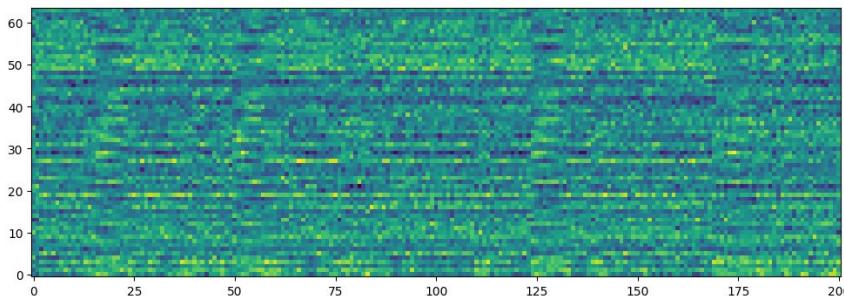
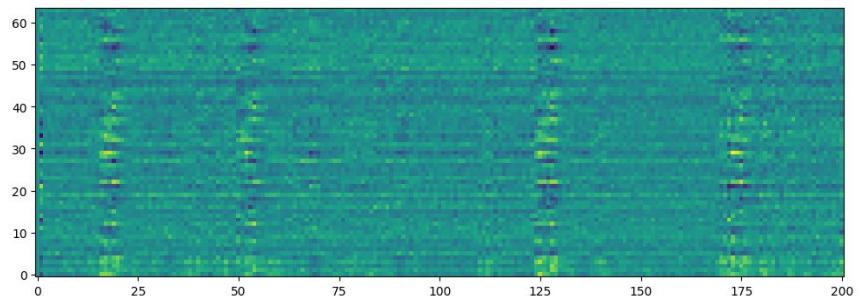
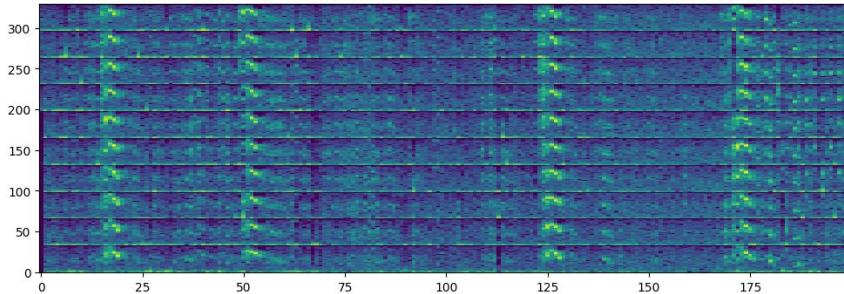
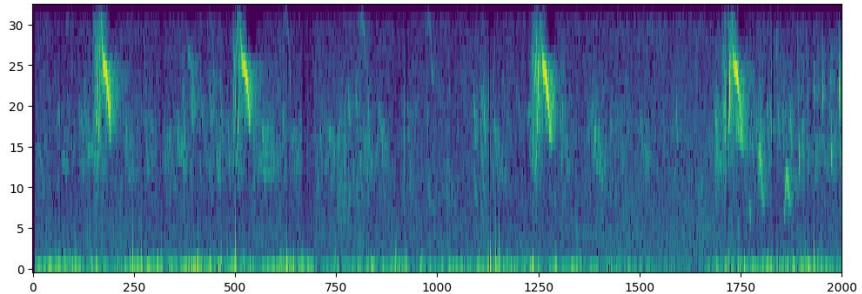
+ Audio
Augmentation

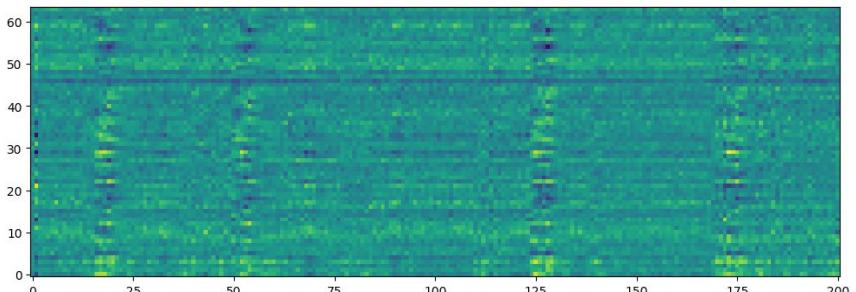
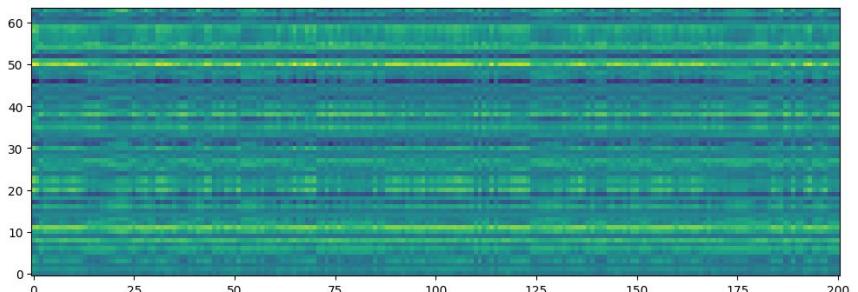
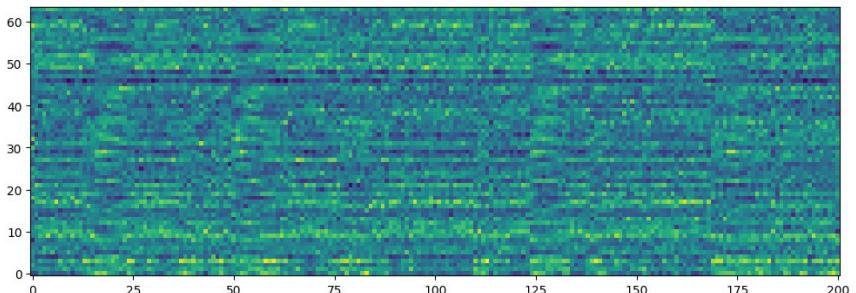
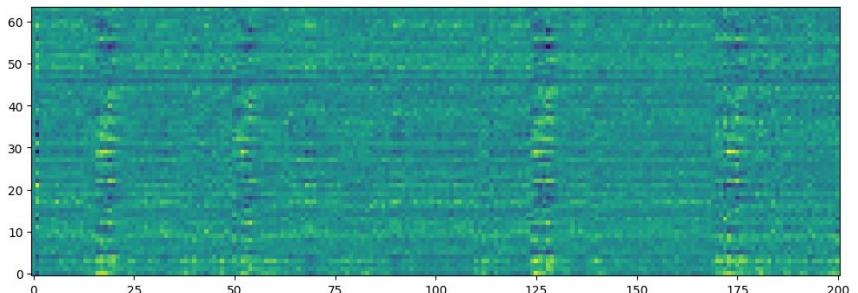
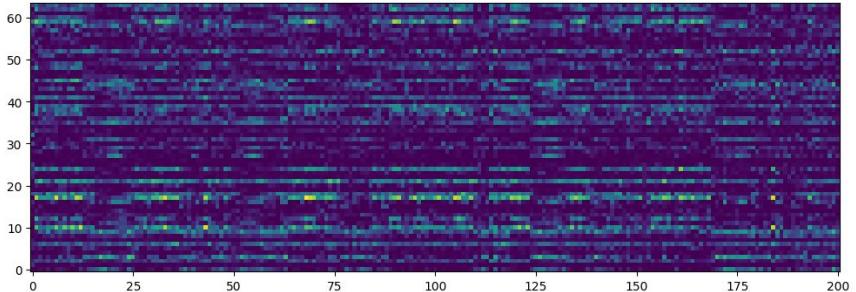
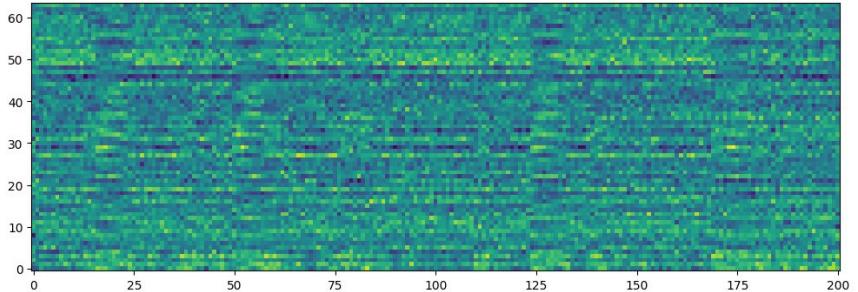


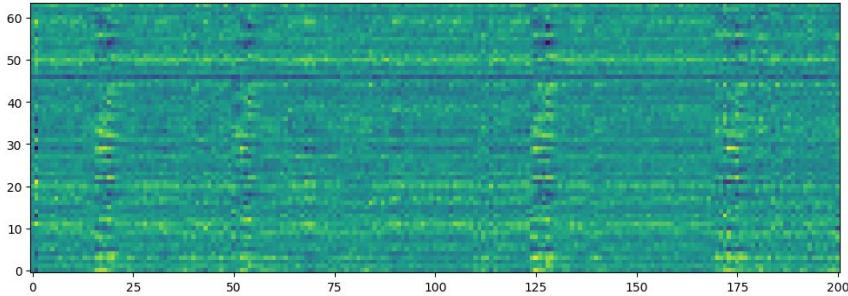
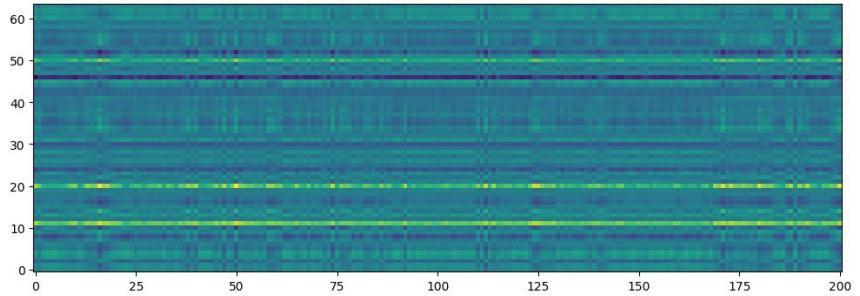
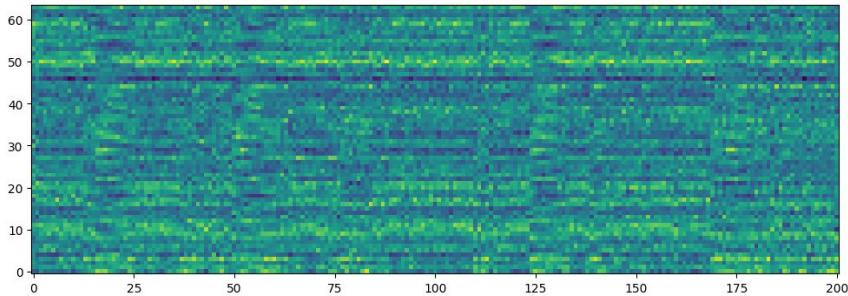
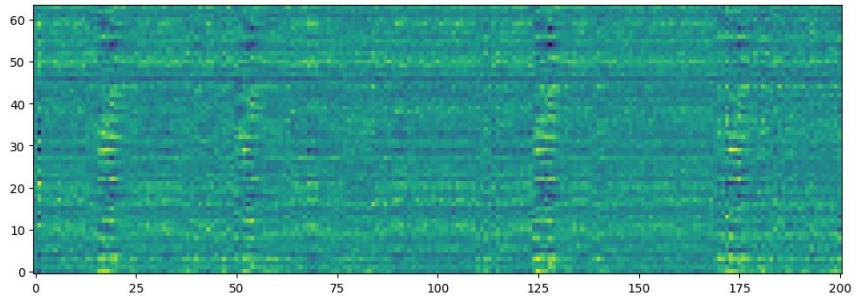
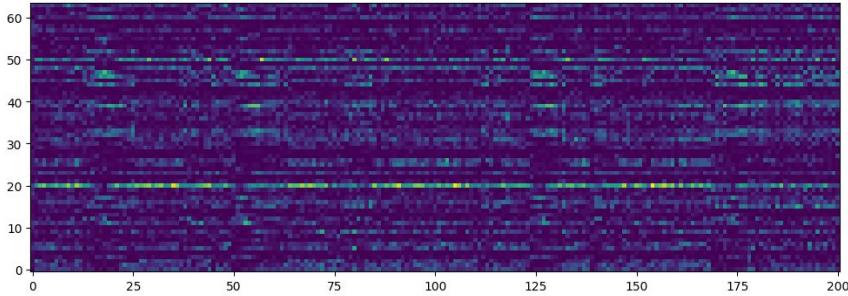
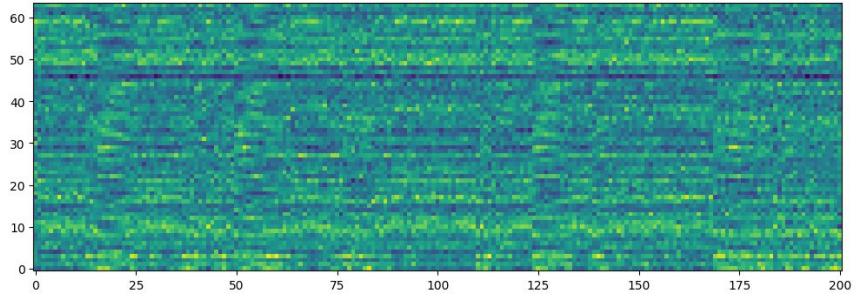
+ SpecAugment

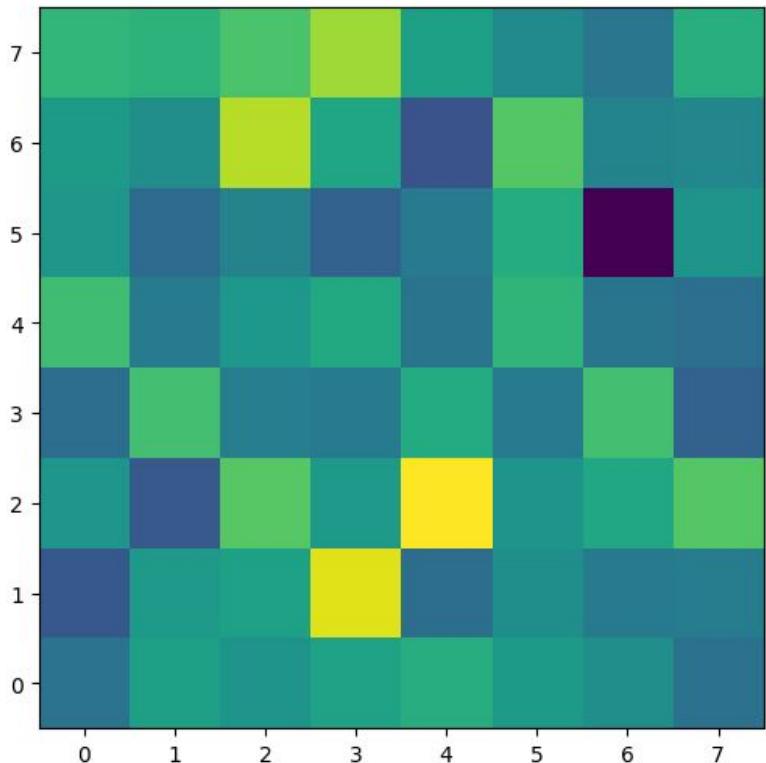


ViT Intermediate Layer Visualizations









	Logits	Probability
Barn Swallow	-1.2	0.5%
Common Sandpiper	-3.4	0.06%
Western Yellow Wagtail	4.0	99.4%

