Bird Species Classification from Acoustic Data Using a Convolutional Neural Network

Anuradha Ramachandran (anuram08@uw.edu), Monica Ramsey (mbr03@uw.edu)

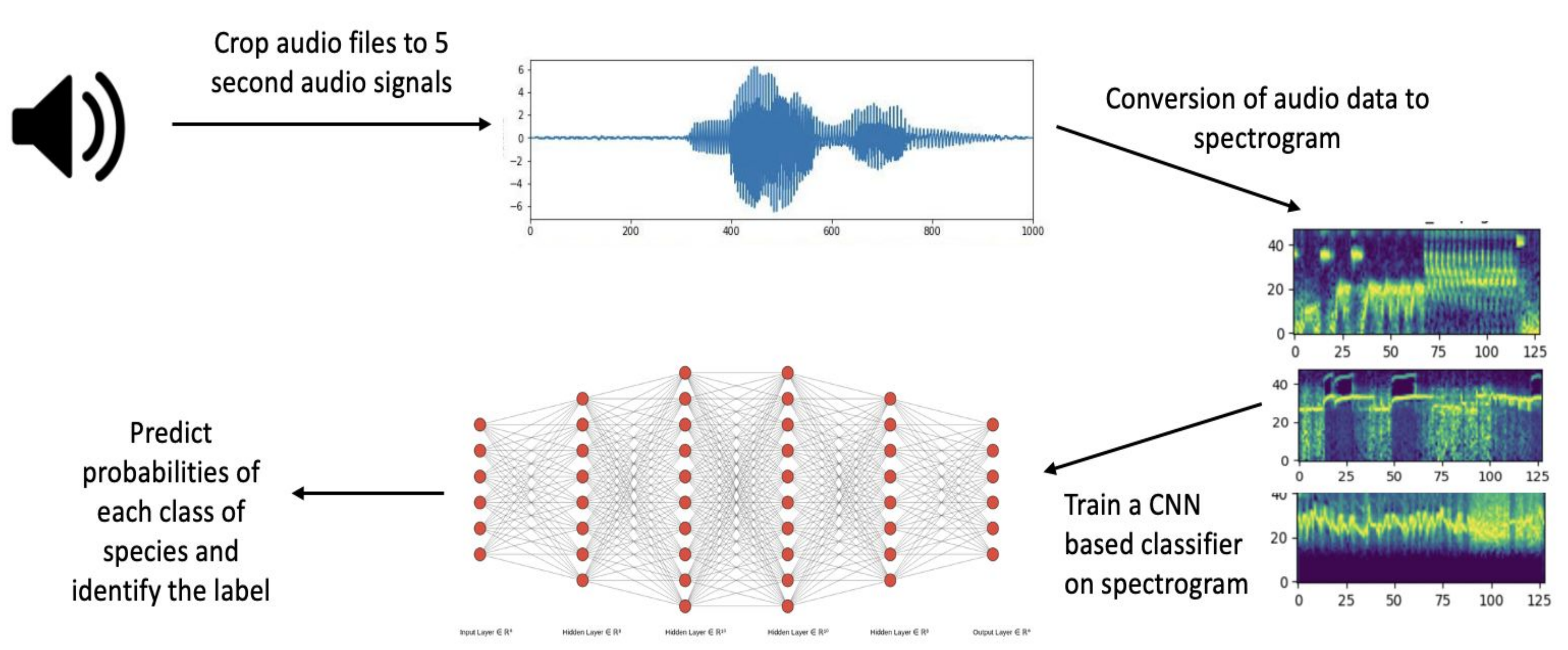
Background

- Acoustic monitoring is a widely-used method for conducting ecological and environmental research
- The most popular method for audio-based classification are based on a convolutional neural network (CNN) and have been successfully used classifying bird species in different geographic regions including North America, Europe, and Taiwan

Research Question

- Presently, there is no classification model to identify birds species in Eastern Africa
- To address the need for models that can classify avian species on a global scale, we develop and train a CNN-based model for the classification of bird species in Eastern Africa

Modeling Approach

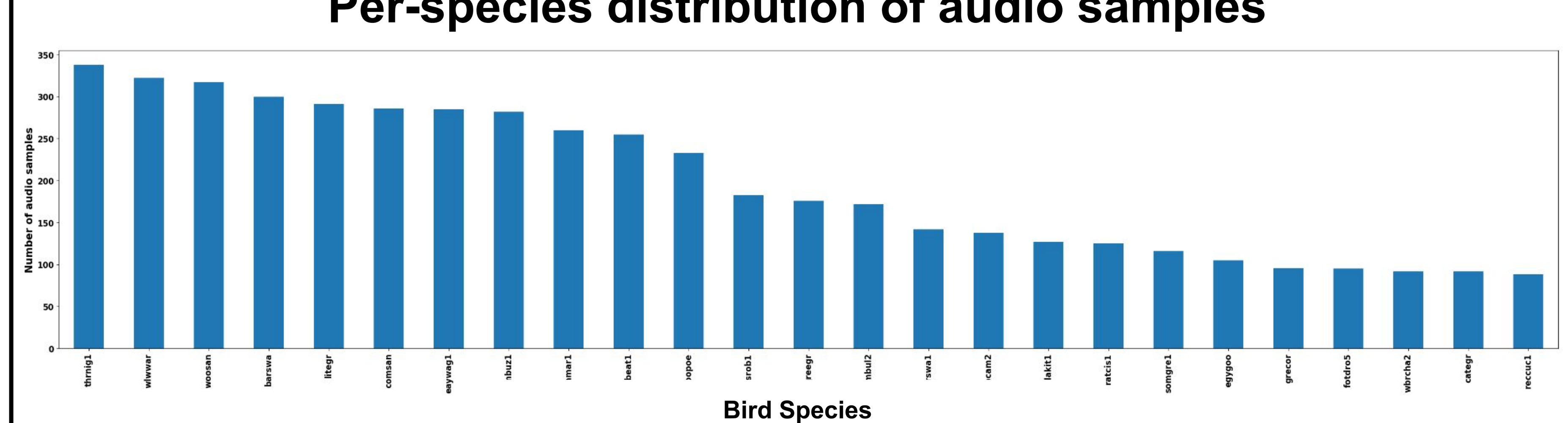


Performance Metrics

$$\text{RECALL} = \frac{\textit{true positive}}{\textit{true positive} + \textit{false negative}}; \quad \text{PRECISION} = \frac{\textit{true positive}}{\textit{true positive} + \textit{false positive}}$$

$$\text{ACCURACY} = \frac{\textit{true negative} + \textit{true positive}}{\textit{true positive} + \textit{false positive} + \textit{true negative} + \textit{false negative}}$$

Per-species distribution of audio samples



Maximum number of audio files per species = **338**
Minimum number of audio files per species = **88**

Dataset Information

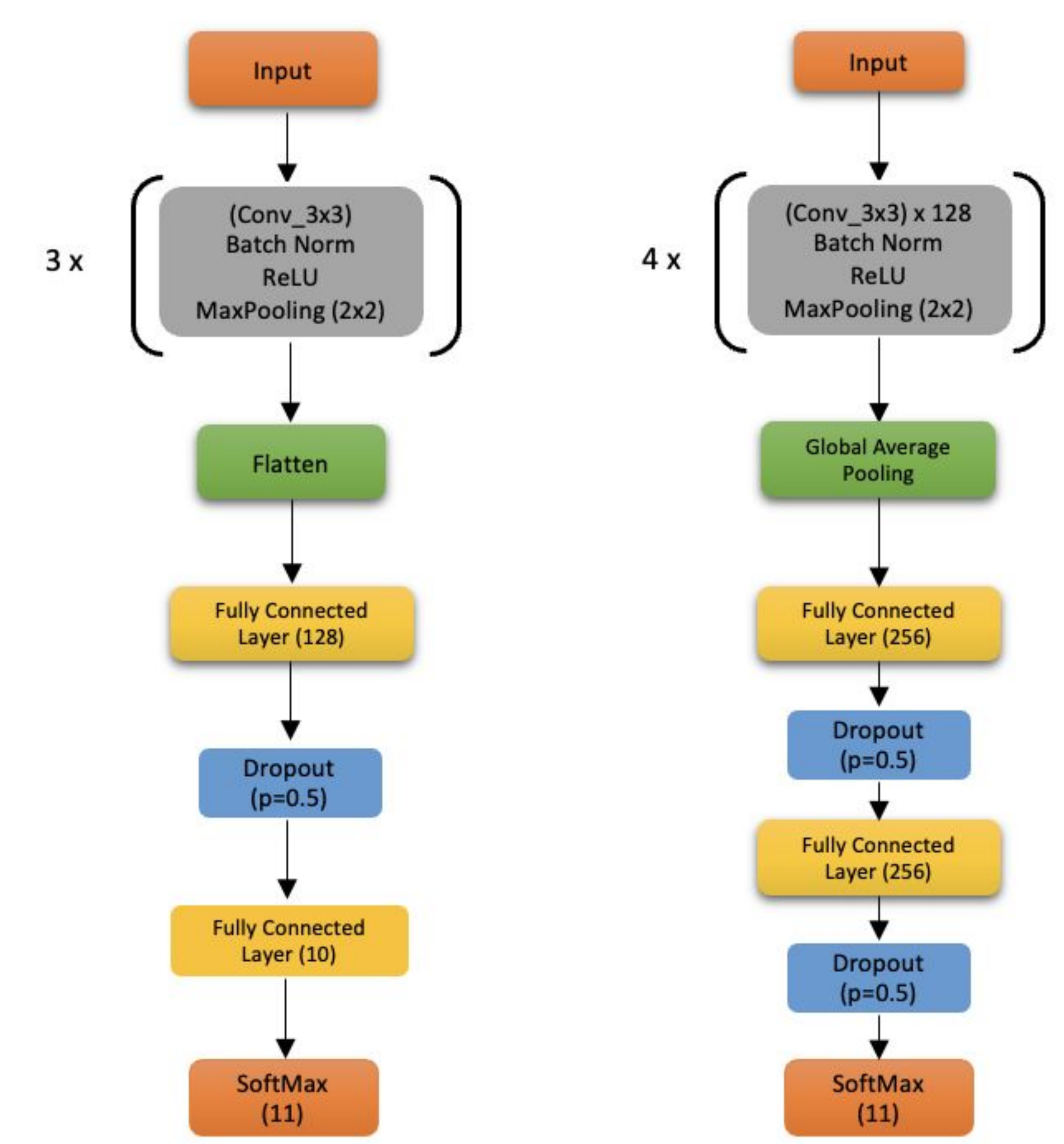
Obtained data from Kaggle's **BIRDCLEF 2023** challenge for training and testing:

- **16941 audio recordings** at 32kHz in corresponding to **264 bird species**
- Non-uniform audio quality and audio lengths

Training and Validation Data:

- Total of **4916** audio samples corresponding to **25 most represented birds** in data set
- Included only samples with **audio rating** of ≥ 4 and > 88 audio samples per species

Model Architecture



Model 1 Architecture *Model 2 Architecture*

References

1. Gabriel Morales et al. "Method for passive acoustic monitoring of bird communities using UMAP and a deep neural network ", *Ecological Informatics*, vol.72, 2022
2. Dan Stowell et al. "Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge", *Methods in Ecology and Evolution*, vol.10, 2018
3. Mario Lasseck, "Acoustic Bird Detection with Deep Convolutional Neural Networks",*Detection and Classification of Acoustic Scenes and Events*, 2018

Experiments Conducted

- **Spectrogram features**
  - Varied **height** and **width** of spectrograms
  - Varied **length** of fast Fourier transform window (N_FFT)
- **Model Architecture**
  - Varied **number of layers** in models
  - Varied **dropout** layers in models
- **Regularization**
  - **L1** in range (0.01 - 0.0001)
  - **L2** in range (0.01 - 0.0001)
- **Rebalance class distribution**
  - **Oversampling** techniques using SMOTE and ADASYN
  - **Weighted** loss
- **Hyperparameter tuning**
  - **Batch size** in range (16, 32, 64)
  - Number of **epochs** in range (30,50,70)

Baseline Model Performance

Parameters:

- Dim = 48x128x1
- Epoch = 50
- Lr = 0.001
- Batchsize = 32

Model	Validation set		
	Accuracy (%)	Recall(%)	Precision (%)
Model 1	73.96	64.52	89.34
Model 2	94.21	93.05	95.74

Conclusions

- **Oversampling** had little effect on Model 1 performance
- Weighted loss **decreased** performance for Model 2 and **minimally increased** performance for Model 1
- Increasing N_FFT led to **4% increase** in accuracy and **6% increase** in recall
- **L2 regularization** at 0.001 yielded **76% accuracy** and **67% recall**
- Increasing **number of layers** in Model 1 increased accuracy to **84%**
- Increasing **spectrogram width** increased recall by <1%
- Increasing the **spectrogram height** increased recall by 2%
- Decreasing **batch size** led to decrease in Model 1 performance and increase in Model 2
- Increasing **number of epochs** improved Model 1 performance by 2%

Future Studies

- Experiment with different **spectrogram features,** including window length, hop length, sampling rate, and spectrogram height
  - Temporal resolution and frequency resolution have inverse relationship
- Experiment with **data augmentation** techniques, including pitch shifting, time stretching, noise addition, adjusting signal-to-noise ratio
- Experiment with different **class rebalancing** techniques, including downsampling, or combining downsampling with upsampling