Predicting Sounds of Seattle Birds

Intro

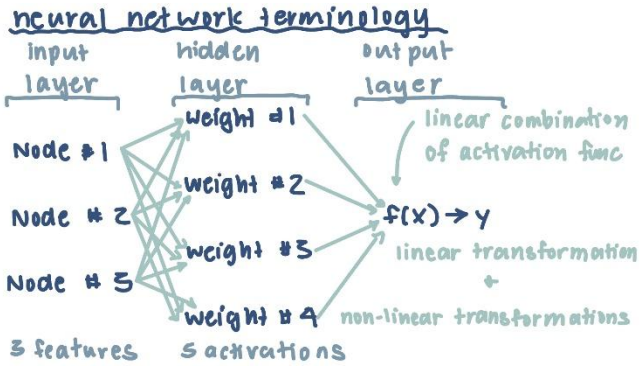
This report will utilize convolutional neural networks to classify common birds in Seattle by sound clips of their calls. The data comes from Xeno-Canto, which is a crowd sourced bird sounds archive (2). Particularly, this study will focus on classifying twelve bird species that are common in Seattle, including the American Crow, American Robin, Bewick's Wren, Black-capped Chickadee, Dark-eyed Junco, House Finch, House Sparrow, Northern Flicker, Red-winged Blackbird, Song Sparrow, Spotted Towhee, and White-crowned Sparrow (1). The following models explore both binary and multi-class classification tasks to evaluate the model performance of convolutional neural networks and the potential of using deep learning in bird call identification.

Theoretical Background

Neural networks are machine learning models that follow a decision-making process that weighs various input options and arrives at a conclusion. Neural network models are particularly effective at modeling complex, non-linear relationships in data. High-dimensional data like images or audio files are common use cases for neural network model implementations.

A neural network model contains an input layer, one or more hidden layers, and an output layer. Each node receives inputs and multiplies them by learned weights. Weights are determined by minimizing the value of a loss function, which quantifies the difference between the model's predicted and actual values. Examples of loss functions include binary cross entropy and mean squared error. These values are then passed through an activation function.

Figure 1. Neural Network Diagram



Activation functions apply linear transformations to nonlinear functions, enabling the model to capture complex relationships. Some examples of activation functions include sigmoid, which maps inputs between zero and one for binary classifications, and softmax which converts inputs to probabilities that sum up to one, for multi-class classifications.

Convolutional Neural Networks are a specific type of deep learning neural network that learns features by sliding a filter across data to detect patterns and create a summary. Max pooling is a method to reduce dimensionality and computational power by taking the maximum value within a given block and condense the data. Convolutional Neural Networks are commonly used for image classification problems, like analyzing a spectrogram.

Neural networks have many limitations and are not always the optimal model to implement. Neural networks are slow and computationally expansive. These models need lots of training data, and will often overfit to the patterns in training data. Techniques like drop out regularization will randomly reduce connections from certain nodes and data augmentation can add more training data, but these methods are not perfect solutions. Neural Networks are also not easily interpretable like a linear regression, so there is a loss of model application in many real-world problems and applications. It is always a good path to try simple and more easily explained models first, however there are some situations where neural networks might be the only model choice. Neural Networks are flexible and often the only choice to analyze sound and image data; in specific cases these models are strong and robust options.

Methodology

The original Kaggle dataset on birdcall data contains 264 species. The first step of the data preprocessing is narrowing down the dataset to the following twelve bird species that can be found in Seattle: the American Crow, American Robin, Bewick's Wren, Black-capped Chickadee, Dark-eyed Junco, House Finch, House Sparrow, Northern Flicker, Red-winged Blackbird, Song Sparrow, Spotted Towhee, and White-crowned Sparrow. Sound clips were selected from each species and subsampled to 22050HZ. The first three seconds of the sound clip are selected and a spectrogram, or an image of the bird call, is produced for each 2-second window. This results in a 128 frequency x 517 time image of the bird call, and 38-630 samples are produced for each of the selected bird species.

Binary CNN Classification: American Crow and White-crowned Sparrow

This binary classification will use a convolutional neural network to identify the call of an American Crow or White-crowned Sparrow. First, the data for both species were extracted and the shape of their data is reordered to the number of samples, frequency, and time. The X data frame is created by combining the data for both bird species. The y data frame is created similarly, but only selecting the number of samples for each species and labeling the American Crow as one and the Sparrow as zero. These data frames are split into a train/test split of 67% training and 33% testing and a fixed random state to ensure reproducibility. The y data frames are passed through a label binarizer to ensure the target variable is converted to a binary class.

Various convolutional neural networks were tested to determine the optimal model by tuning for a combination of convolutional layers and their complexity, kernel sizes, and dropout rates (see Test_Models.ipynb for full code). The optimal model includes two convolutional layers with 32 and 64 filters using ReLu activation and a 3x3 filter, max pooling to reduce dimensionality, flatten to pass a one-dimensional vector through 64 dense layers, and a dropout layer that randomly drops 50% of neurons for regularization. The model is compiled using binary cross entropy as the loss function to compare predicted and actual values, while using accuracy to evaluate the binary classification. Finally, the model is trained on 50 iterations using an early stopping method tracking the validation accuracy, with a batch size of 16.

Multi-class CNN Classification: All 12 Common Bird Species in Seattle

This multi-class classification will utilize a convolutional neural network to identify any of the bird calls from the subset of 12 bird species that can be found in Seattle. The preprocessing of the data is like the binary classification CNN, where all bird species data are extracted, shapes are reordered, and data is concatenated into X and y data frames. Similarly, these data frames are split into a train/test split of 67% training and 33% testing with a fixed random state, and target variables are passed through a label binarizer.

The multi-class CNN model was also tested on various tuning metrics to determine the optimal model, altering amount of convolutional layers and degree of dropout rates. The optimal model includes two convolutional layers with 32 and 64 filters using ReLu activation and a 3x3 filter, max pooling to reduce dimensionality, flatten to pass a one-dimensional vector through 64 dense layers, and a dropout layer that randomly drops 50% of neurons for regularization. The model is compiled using categorical cross entropy as the loss function to compare predicted and actual values, while using accuracy to evaluate the multi-class classification. Finally, the model is trained on 50 iterations using an early stopping method tracking the validation accuracy, with a batch size of 16.

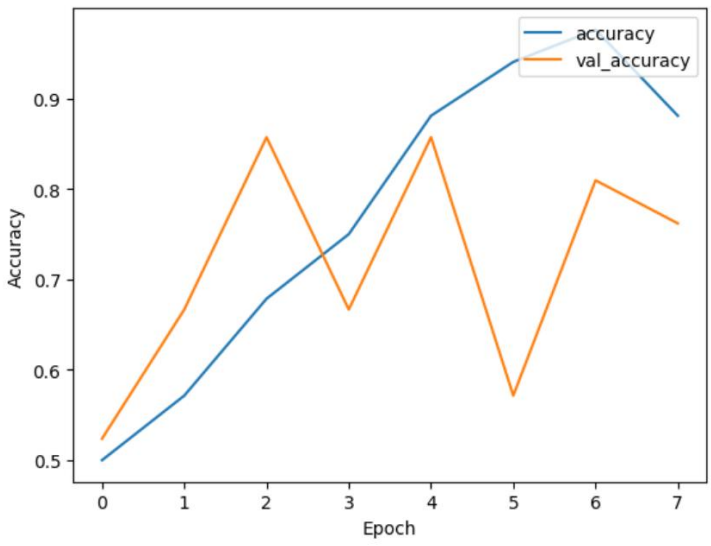
Results

Binary CNN Classification: American Crow and White-crowned Sparrow

The goal of this convolutional neural network model is to classify spectrogram images of bird calls into either an American Crow or White-crowned Sparrow. This model returns a relatively high test accuracy, which means that this model correctly classifies 76.92% of the spectrograms it has never seen before. A strong accuracy implies that this model has learned general patterns in the training data and can reliably apply these patterns to data the model has never seen before.

The following plot shows the training and validation accuracy over the number of iterations of the training data or epochs. The blue line represents the change in training accuracy and the orange line represents the change in validation accuracy. Initially, the model does a good job learning patterns in the data and applying them to the test set, as both accuracies rise steadily together. Around the middle of the plot, the training accuracy continues to rise while the validation accuracy dips and

Figure 2. Binary CNN Accuracy Plot

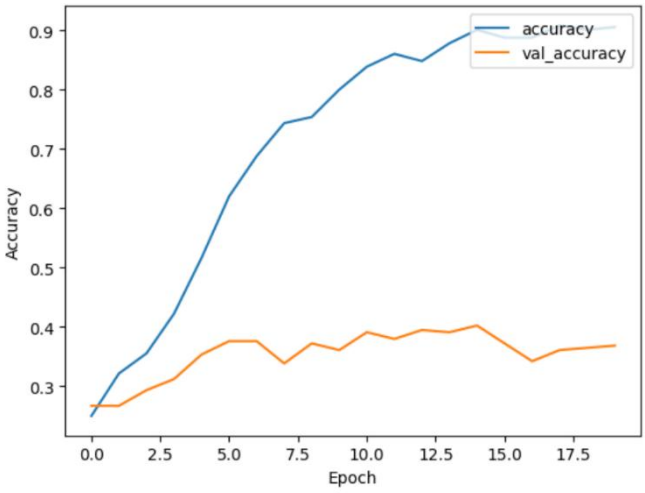


diverges slightly from the training accuracy. This divergence shows that the model begins to overfit to the training data for a few epochs, so the early stopping allows the model the stop training before the overfitting gets worse. Overall, this binary classification convolutional neural network does a decent job at classifying between an American Crow and White-crowned Sparrow.

Multi-class CNN Classification: All 12 Common Bird Species in Seattle

The goal of the multi-class classification convolutional neural network is to classify spectrogram images of twelve bird calls that can commonly be found in Seattle. This model returns a decent test accuracy, where model is able to correctly classify 40.98% of the spectrograms it has never seen before. This model is significantly better than guessing, where guessing one out twelve species correctly would be about a 9% accuracy, so the model has clearly learned some patterns among the training data.

Figure 3. Multi-Class CNN Accuracy Plot

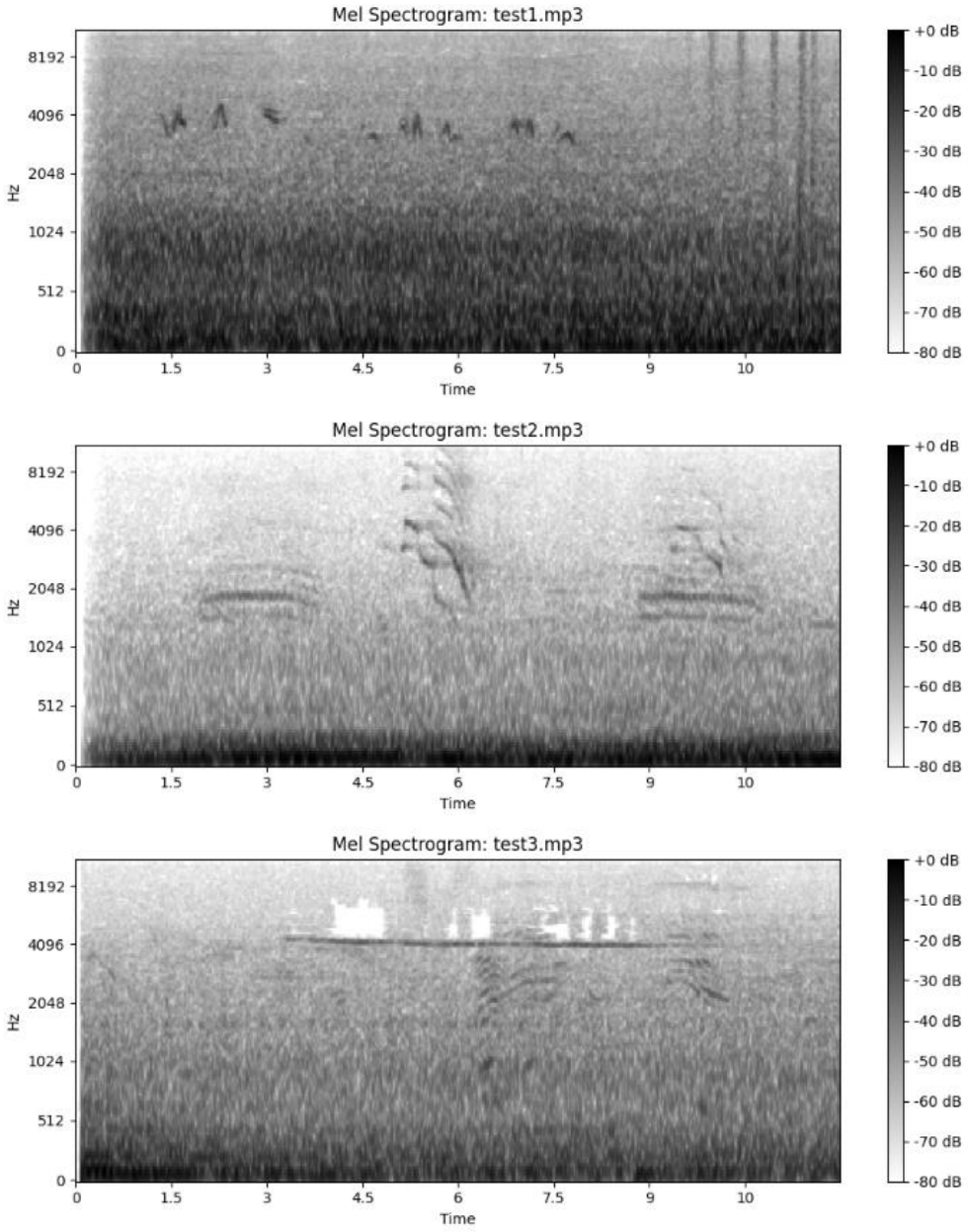


The following plot shows the training and validation accuracy over the number of iterations of the training data or epochs. The blue line represents the change in training accuracy and the orange line represents the change in validation accuracy. Initially, the model does an okay job learning patterns in the data and applying them to the test set, as both accuracies rise steadily together. However, at around 5 iterations the model starts to overfit to the training data learning nuances of different species but failing to generalize these patterns to new data the model has never seen before. Overfitting is a common issue when training convolutional neural networks, and future implementations might include data augmentation to balance the classes of training data so the model can better learn patterns from all twelve species.

Predicting on External Test Data

The multi-class classification convolutional neural network model will be used to predict three bird species using their sound from an mp3 file. Firstly, to preprocess the data, the same steps will be used as discussed previously in the methodology section. Each mp3 file will be subsampled to 22050HZ. The first three seconds of the sound clip are selected and the following spectrogram images are produced for each sound clip.

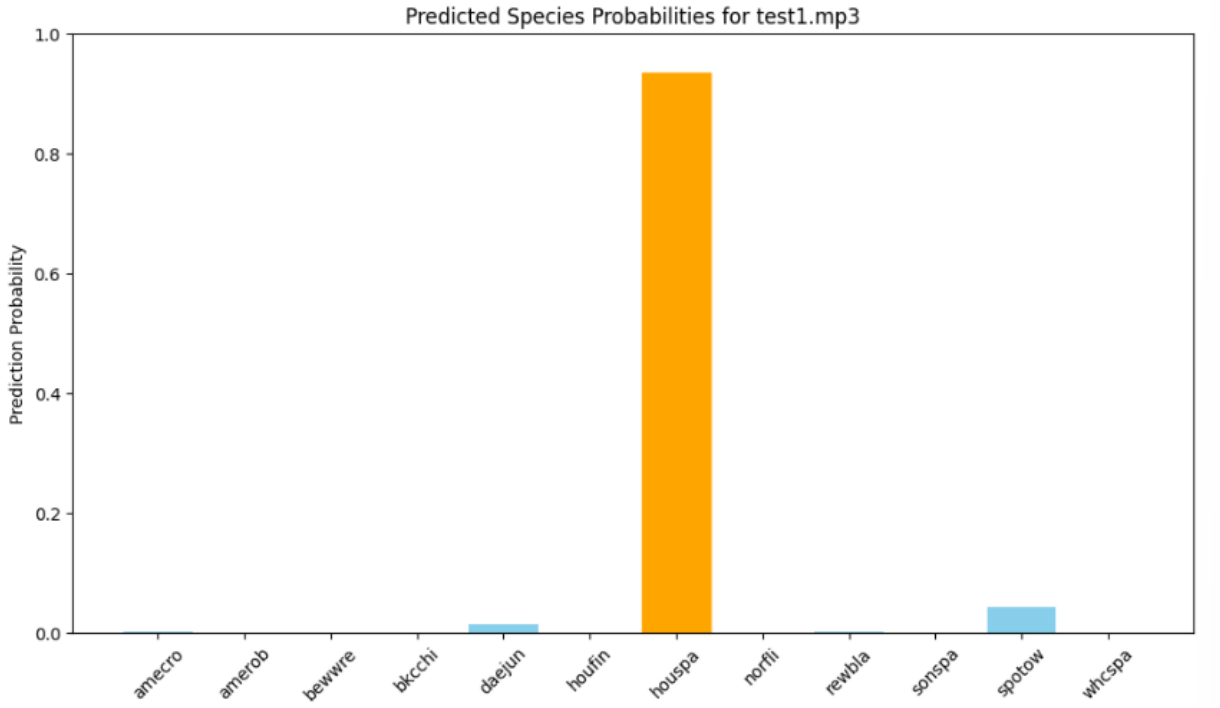
Figure 4. Test Data Spectrograms



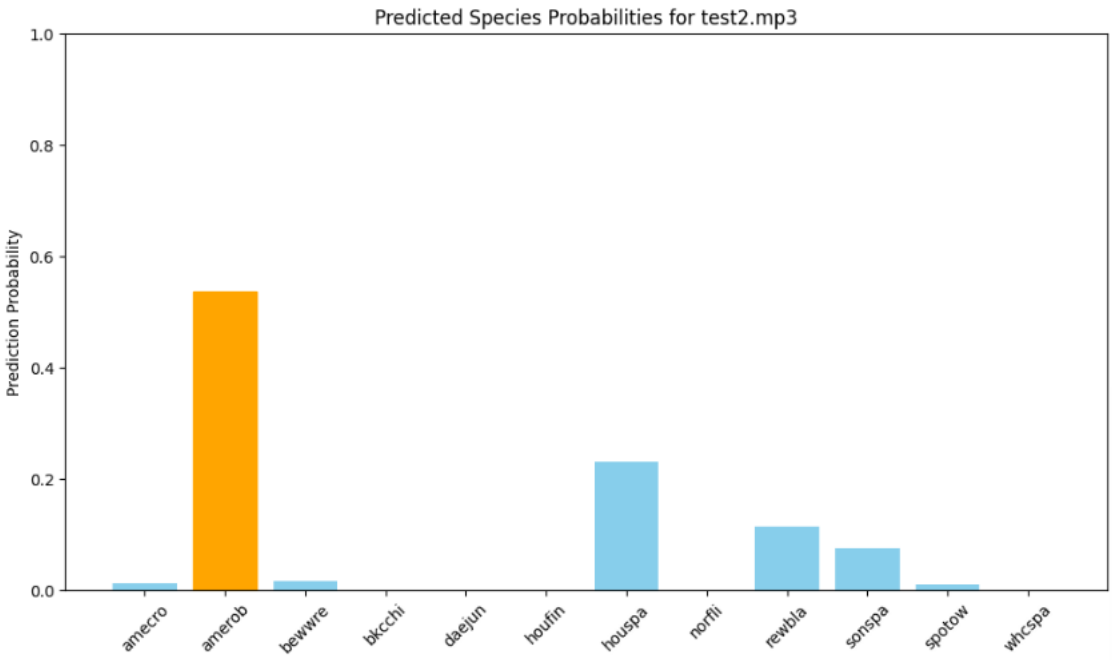
Next, an image of the bird class is produced for every 2-second window which results in a 128 frequency by 517 time image of the bird call. All image data for the three bird species are saved into one file. The multi-class convolutional neural network model is used to predict on this test data. The bird species predictions are House Sparrow, American Robin, Red-winged Blackbird for the respective files 1,2 and 3.

Graphing the prediction probabilities for each of the twelve bird species will help analyze how confidently the model predicts the specific bird call from each file as well as if there may have been multiple bird calls within the sound clip. Each graph represents the prediction probabilities

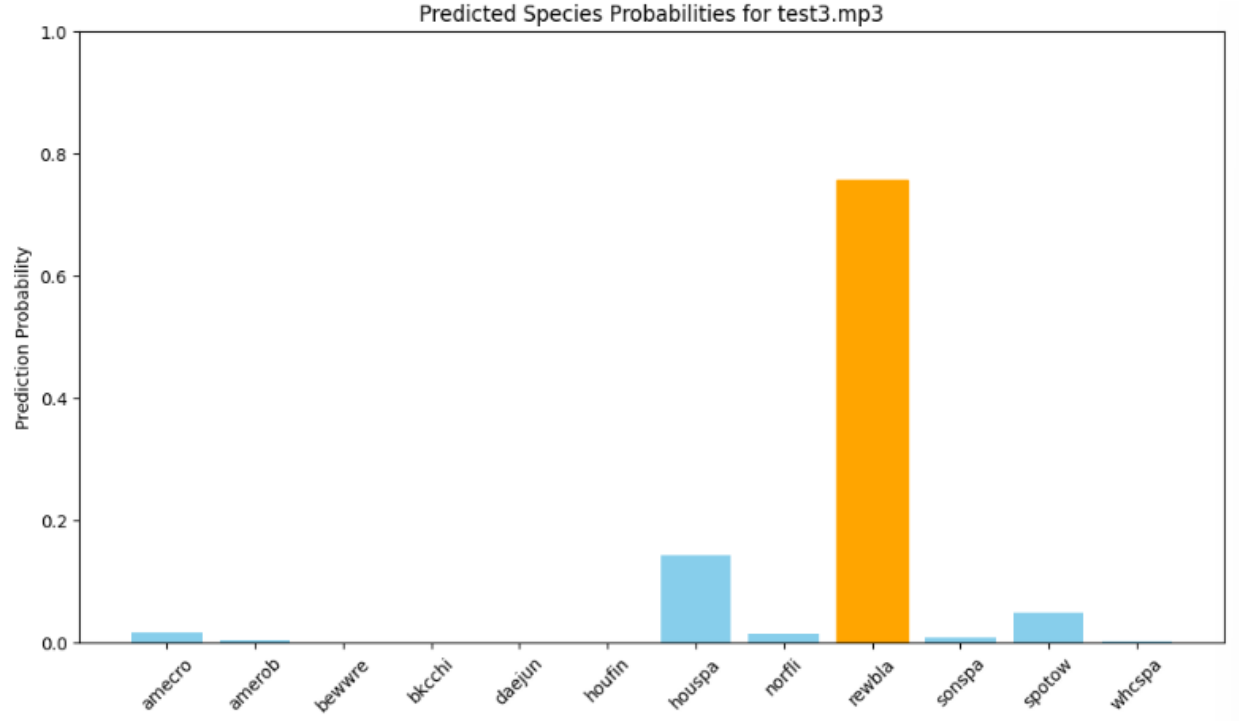
among all twelve species for each sound file. The blue bars represent the prediction probabilities, and the orange bar represents the highest prediction probability or the species that the model predicts to classify the bird call. The first graph represents the prediction probabilities for the first file, which classifies the bird call as the House Sparrow. Since the prediction probability for the sparrow is extremely high (93.53%), the model pretty confidently classifies this species. This sound clip is very likely the House Sparrow and there are probably not any other bird calls present in this sound clip.



The following graph represents prediction probabilities for the second file, which classifies the bird call as the American Robin. The multi-class CNN model classifies the robin with 53.63% confidence, which is not as high as the previous sound file but still a pretty solid prediction. The prediction probabilities for House Sparrow (23.16%), Red-winged Blackbird (11.55%), and Song Sparrow (7.52%) are also notably high. This might imply that there is a good change the American Robin is the main bird call that is present in the sound clip, but other bird calls like the house sparrow or red-winged blackbird may also be present.



Finally, the last graph represents prediction probabilities for the third file, which classifies the bird call as the Red-winged Blackbird. The multi-class CNN model classifies the blackbird with 75.82% confidence, which is a relatively high prediction probability. The prediction probabilities for House Sparrow (14.36%) is also slightly higher probability than the rest of the bird species. This prediction is likely a Red-winged Blackbird and either other bird species may be present in the sound clip or the model may be misclassifying the House Sparrow.



Discussion

In this report, two types of convolutional neural network models were implemented to classify between two birds and twelve bird species that are common to the Seattle area. The binary model, one that predicts only between two bird species, performs significantly better than the multi-class model, which can classify between all twelve bird species. The high accuracy of the binary model suggests that simpler classification tasks, one that identifies fewer classes, can learn clearer differences between bird calls and allows the convolutional neural network to learn and generalize patterns more effectively. The multi-class model still performs substantially better than a random guess, which demonstrates that the model was still successful at identifying prominent patterns and characteristics of different bird calls.

Both models present with signs of overfitting, which is indicated by the divergence of training and validation accuracy after an increasing number of epochs. Both models utilized an early stopping method when training the neural networks, which allows the model to reduce the chances of severe overfitting, however the issue is still present. Overfitting is likely caused by the wide range of class sizes, where the samples for bird species range anywhere from 38 to 630 samples. Underrepresented species in the training data likely skew the training of the model and reduce overall performance for unseen data. Future implementations might address this class imbalance by oversampling for the underrepresented classes or utilizing data augmentation to increase the number of training samples to address the overfitting issue.

Implementing the model on real-world audio clips of bird calls has promising results, where the convolutional neural network was able to confidently classify the first audio as a House Sparrow with 93% accuracy and moderately classify the other two clips as Rid-winged Blackbird and the American Robin. The second clip was classified as the American Robin with 53% accuracy, but might show ambiguity as other bird species may also be present in the background of the audio clip. Even though the model was not specifically trained to identify multiple bird species in one sound clip, it does a decent job at analyzing real world audio that might include difference species as well as background noises.

Conclusion

This report demonstrates the feasibility of using convolutional neural networks to classify bird species based on audio spectrograms, showing promising results for both binary and multi-class classifications. Even with limited training data, these models are able to extract meaningful patterns and make reasonably accurate predictions. These bird classification models can be implemented in the real world as a means to monitor biodiversity and bird species populations. With climate change and urbanization on the rise, it is important to protect the diversity and populations of species that could be at risk (3). Having a means to automate and track bird species without having to manually listen to live audio or being present at the scene could be the first step towards protecting biodiversity.

References

[1] Pyle, P. (2024). Alpha codes for 2022 bird species. https://www.birdpop.org/docs/misc/Alpha_codes_eng.pdf

[2] Xeno-cano Foundation. (2025.). *Canto*. xeno. https://xeno-canto.org/

[3] 2025 National Audubon Society. (2024, January 19). *Survival by degrees: 389 bird species on the Brink*. Audubon. https://www.audubon.org/climate/survivalbydegrees