Modeling Habitat Shifts: Integrating Convolutional Neural Networks and Tabular Data for Species Migration Prediction

Min-Hong Shih*, **Emir Durakovic***
Northeastern University
360 Huntington Ave
Boston, MA 02115
durakovic.e [at] northeastern.edu
shih.minho [at] northeastern.edu

Abstract

Due to climate-induced changes, many habitats are experiencing range shifts away from their traditional geographic locations [1]. We propose a solution to accurately model whether bird species are present in a specific habitat through the combination of Convolutional Neural Networks (CNNs) [2] and tabular data. Our approach makes use of satellite imagery and environmental features (e.g., temperature, precipitation, elevation) to predict bird presence across various climates. The CNN model captures spatial characteristics of landscapes such as forestation, water bodies, and urbanization, whereas the tabular method uses ecological and geographic data. Both systems predict the distribution of birds with an average accuracy of 85%, offering a scalable but reliable method to understand bird migration.

Introduction

Climate change has resulted in the relocation of various species [3]; for example:

- Bird species are shifting their migratory routes to higher longitudes to find suitable nesting areas [4].
- Invasive aquatic animals favor warmer waters where the success of cold water invasive species is minimized [5].
- Reptiles and amphibians are finding locations that match their thermal and moisture needs [6].

We will tackle the first mentioned example above. Due to mobility, visibility, and well-documented historical ranges, birds have always been a good bio-indicator of environmental change [7]. Hence, understanding how bird species respond to climate change is critical to conserving our forests and introducing future environmental policies on undoing global warming effects.

Traditional methods of analyzing bird distributions usually rely on manual observations, a process that has major geographic coverage, temporal consistency, and resource limitations [8]. A field biologist can only be in one location at a time. Hence, there is a growing need for more scalable approaches, such as modeling migration patterns through the use of predictive AI agents. A method that is often used to model species distributions is climate envelope models.

*These authors contributed equally.

Climate envelope models are a subset of species distribution models that use climate variables such as climate, land cover, and topography to predict the environmental suitability of a species [9]. These models often use mathematical functions to describe the associations between variables and the presence of the species [10]. However, the limitations of climate envelope models lie in their inability to account for factors such as species interactions, dispersal limitations, and landscape features visible in satellite imagery [11].

In this paper, we examine ways to classify whether birds will be present in various climate types. More specifically, we focus on implementing a CNN model that takes satellite imagery and predicts species migration patterns through local landscape features that would affect the abundance of a bird species, such as bodies of water, forestation, and urbanization.

We also suggest a tabular model that generates a dataset that contains features such as latitude, longitude, elevation, precipitation, and temperature. This data set is then fed into a random forest classifier that determines whether a bird is present or absent in a given location.

Our main contributions include:

- A CNN-based method for using satellite imagery to predict species presence across landscapes.
- A feature-driven random forest classifier that uses climate and topography data.
- A comparative analysis of both models using real-world bird occurrence data from the eBird database and prebuilt models.

Our research study offers a more scalable technique to predict bird distributions and hope to inspire better climate conservation practices.

Background

In this section, we will be discussing how the models used in this paper were constructed.

- **Random Forests**
  A decision tree is represented by a node of a test attribute, and each corresponding branch is the result of that test. The random forest classifier constructs multiple decision trees and takes the average to decide the probability of a bird species being present at a given location [12].

- **Gradient Boosting Decision Trees**
  Similar to random forests, Gradient Boosting Decision Trees (GBDT) build trees sequentially with each new tree correcting the errors of the previous trees. We employ this technique as it generally performs better on smaller datasets due to its ability to reduce bias [13].
- **Convolutional Neural Networks**
  CNNs are designed to extract spatial features from image data. In our case, satellite imagery is processed through a CNN to capture important landscape features such as water bodies, vegetation, and urban structures that may influence bird presence [2].
- **Residual Network (ResNet)** Residual Network, otherwise known as ResNet are a type of neural network architecture that bypasses layers. The "skip connections" allow for a more direct information flow across the network eliminating the issue of "vanishing gradients" that occur when training on a model with hundreds of layers. ResNets learn the "residual" hence their name, between the desired output and the layer input. This improves its performance for tasks such as object detection and image classification [14].
- **Multi-label Classification** Since multiple bird species can be present in the same region, we frame the problem as a multi-label classification task. The model outputs a vector of presence probabilities for each species, rather than a single-class label, enabling simultaneous prediction of multiple species.

Related Work

Similar research projects have explored both bird migration distributions and habitat changes over several years. We begin by exploring the SatBird research study that generates a dataset based off of Sentinel-2 [19] satellite imagery, biological- and climate-oriented rasters, and the eBird dataset [17] for presence labeling. This dataset is divided into USA-summer, USA-winter, and Kenya-specific regions. The research project then inputs their dataset into baseline model classification and presents the dataset for further scientific development [38].

We also looked into a research project dealing with habitat prediction for spatial distribution of Japanese rice fish. The researchers of this project used Fuzzy Neural Networks, a system which mimics human cognitive abilities with Neural Networks to handle uncertainty and learn from the data, among other models. Their dataset involves features like depth and vegetation coverage, data which is then fed into the prediction model to evaluate habitat preference [16].

As discussed, there have been existing studies that have made efforts to species distribution modeling; however, limitations remain. The SatBird dataset, although comprehensive in its integration of satellite imagery with citizen science data, focuses primarily on static seasonal distributions and lacks the temporal resolution to capture progressive movement patterns. Studies using Fuzzy Neural Networks for habitat prediction show strong uncertainty that offer little transferability between areas. Our study addresses this gap [38] [16].

Methods

We define our problem as such: given a location (x_{lat}, x_{lon}), where x_{lat} and x_{long} represent the latitude and longitude coordinates respectively, with its respective features, H, and a binary label y_i generate a dataset, D(H, y_i), such that we can learn a function f(D) that minimizes the classification error for presence prediction across a range of birds with output, z.

Currently, observations are simply recorded when a bird is spotted at a certain location. Manually transcribing observations takes effort and can grow tiring. We can further simplify manual classification by combining well-known datasets to create a more comprehensive dataset. We propose a solution to utilize this exhaustive dataset to generate predictions further enhancing bird distribution models. This exhaustive dataset generated through the longitude and latitude would not only contain tabular data similar to data generated through manual transcription, but would also contain satellite images of the longitude and latitude area. By fine-tuning our dataset to consider only necessary features that impact birds we can offer reliable and more effective classifications than traditional models.

Datasets

- **eBird Dataset**
  We use the Cornell Lab's eBird dataset [17] to pull bird observation data. This dataset contains the latitude, longitude, observation date, and presence data of multiple bird species in Northern America.
- **WorldClim Dataset**
  The WorldClim dataset [18] features raster files with locations spaced at a resolution of one square kilometer. Each coordinate maps to a real-world location and contains information like its latitude, longitude, elevation, precipitation, and temperature depending on the downloaded dataset .
- **Sentinel-2 Dataset**
  This dataset provides satellite imagery using Microsoft's Planetary Computer API. Using longitude and latitude, a satellite image of the area can be queried in a variety of bandwidth, from human visible wavelengths to non-visible ones. Through the use of queries to the API, it can be specified to a specific condition of weather like "no cloud condition" for a clear image of the landscape [19].

With regards to the tabular model, we combine the eBird and WorldClim datasets to generate a representative dataset of features like: latitude, longitude, elevation, precipitation, temperature, but also the bird species type and its presence at that location. Since eBird only provides presence observations, we generate pseudo-absent observations by checking whether a location is not too close to a known observed location for that bird (within a 1.1km radius).

Statistical Models

We divide our dataset into a 70:10:20 split for training, validation, and testing, respectively[32].

Tabular Model To model species occurrence, we use a random forest classifier trained on a tabular dataset combining environmental features from WorldClim and species observations from eBird. Since eBird only provides presence data, we generate pseudo-absence points from ecologically plausible but unobserved regions[20]. This allows the model to distinguish between suitable and unsuitable habitats.

The random forest builds multiple decision trees using bootstrap samples and feature bagging[21]. Each tree outputs a probability estimate for species presence, and the final prediction is the average of all trees. We apply a threshold θ (that is usually 0.5) on the validation set to convert probabilities into binary predictions, to evaluate our model for accuracy. We can then simply return the probabilistic estimate for the prediction of bird presence. The corresponding pseudo-code is presented in Algorithm 1.

In addition to random forests, we experiment with gradient boosting (e.g., XGBoost), a model which trains trees sequentially on residual errors from previous models. Boosting should improve accuracy for smaller datasets, but for our case it may struggle to classify due to the noise in an ecological setting. Still, we use this model to experiment and contrast between the two. See Algorithm 3.

We evaluate model performance using our 20% testing data split. Moreover, we analyze feature importance to identify key environmental drivers of species distribution[22]. The steps for evaluating this model are detailed in Algorithm 2 and Algorithm 4.

We chose these two models to compare how a more generalized model contrasts with a sensitive to noise model. Especially since we are using smaller datasets, we want to evaluate how the two use the features to differentiate whether a bird is present or absent at a given location.

Convolutional Neural Network For the ResNet CNN approach, we implemented a transfer learning approach using a modified ResNet-34 architecture[25]. The model extracts features from Sentinel-2 RGB imagery (bands B02, B03, B04) at a resolution of 10m by using pre-trained weights from ImageNet. The backbone for this remained the ResNet-34 architecture, while the final fully connected layer was substituted with a custom classifier for multi-label species prediction.

We implemented a dropout of (p = 0.5) in the classification head and froze early convolutional layers to reduce overfitting in our small ecological dataset. Binary cross-entropy loss with Adam optimization was used to train the model (weight decay = 1e-5, initial lr = 1e-4)[28]. Given the intrinsic variability in satellite imagery and the spatial correlation in ecological observations, these methods successfully strike a balance between fitting the training data and generalizing to unseen habitats when paired with weight decay. The pseudo-code is shown in Algorithm 5.

Next a custom CNN was implemented in comparison with the ResNet approach. Without using pre-trained weights, our unique CNN model was created especially for predicting bird habitat. Each of the four convolutional blocks that make up the architecture has two convolutional layers with batch normalization, ReLU activation, and max pooling. The filter depths of the blocks range from 64 to 128 to 256 to 512. Sentinel-2 satellite imagery features, which are very different from those of natural photography datasets like ImageNet, were the focus of this configuration's optimization. To guarantee appropriate gradient flow through the deep network, we used Kaiming initialization[30]. With a learning rate of 1e-4 and a weight decay of 1e-5 for regularization, the model was trained end-to-end using binary cross-entropy loss and Adam optimization. The pseudo-code is shown in Algorithm 6.

Algorithm 1: Train Random Forest

Input: Feature matrix X, labels y
Parameter: Number of trees T, max depth d, min samples split s, max features f
Output: List of trained trees

Initialize list $trees \leftarrow [\,]$
for $i = 1$ to T **do**
  $(X_{sample}, y_{sample}) \leftarrow$ SampleWithReplacement(X, y)
  $tree \leftarrow$ TrainDecisionTree$(X_{sample}, y_{sample}, d, s, f)$
  Append $tree$ to $trees$
end for
return $trees$

Algorithm 2: Predict with Random Forest

Input: Test data X_{test}, trained trees $trees$
Parameter: Threshold θ
Output: Predictions and average probabilities

Initialize list $all\_probs \leftarrow [\,]$
for each $tree$ in $trees$ **do**
  $probs \leftarrow$ PredictProba$(tree, X_{test})$
  Append $probs$ to all_probs
end for
$avg\_probs \leftarrow$ Average(all_probs, axis=0)
$predictions \leftarrow [1$ if $p \geq \theta$ else 0 for p in $avg\_probs]$
return $(predictions, avg\_probs)$

Algorithm 3: Gradient Boosting Training

Input: Feature matrix X, target labels y
Parameter: Number of trees T, learning rate η
Output: Final model $F_T(x)$

Initialize model prediction: $F_0(x) =$ mean of y
for $t = 1$ to T **do**
  Compute residuals: $r_i = y_i - F_{t-1}(x_i)$
  Train a regression tree $h_t(x)$ on inputs X and residuals r
  Update model: $F_t(x) = F_{t-1}(x) + \eta \cdot h_t(x)$
end for
return final model $F_T(x)$

Algorithm 4: Evaluate Gradient Boosting

Input: Test data X_{test}, true labels y_{true}, trained model $F_T(x)$
Parameter: Threshold θ
Output: Binary predictions and evaluation metrics

Compute predicted probabilities: $probs \leftarrow F_T(X_{test})$
Initialize empty list $predictions \leftarrow [\,]$
for each p in $probs$ **do**
  if $p \geq \theta$ **then**
    Append 1 to $predictions$
  else
    Append 0 to $predictions$
  end if
end for
Compare $predictions$ with y_{true} to compute:
  Accuracy, Precision, Recall, F1-score, AUC
return Evaluation metrics

Algorithm 5: ResNet-based Bird Habitat CNN

Input: Sentinel-2 RGB imagery $X \in \mathbb{R}^{3\times224\times224}$, species presence labels $y \in \{0,1\}^n$
Output: Species presence probabilities $p \in [0,1]^n$

Load pre-trained ResNet-34 with ImageNet weights
for each parameter in ResNet layers[1:len(layers)-8] **do**
  parameter.requires_grad $\leftarrow$ false {Freeze early layers}
end for
Replace ResNet.fc with Identity layer
Define classifier as Sequential(
  Dropout(0.5),
  Linear(512, 256),
  ReLU(),
  Dropout(0.5),
  Linear(256, n),
  Sigmoid()
)
function forward(X)
features $\leftarrow$ backbone(X) {Extract features using modified ResNet}
$p \leftarrow$ classifier(features) {Predict species probabilities}
return p
end function

Results

Tabular Data

We evaluate our models based off accuracy which computes the proportion of all predictions that were correct and the "Area Under the Curve of the Receiver Operating Characteristic curve" (AUC-ROC), a measurement which shows how well a model separates data. We choose these two metrics because birds may be rarer in some regions than others, hence AUC helps us understand whether the model is learning relevant features from the dataset and accuracy shows how well our models are performing in general.

Our Random Forest and Gradient Boosting models will be compared alongside the baseline Scikit-learn implementations of the two mentioned models. We compare both the validation and test accuracies to understand both the generalized performance and hyperparameter choices of our models and our actual accuracy of the model. For demonstration purposes, we will be training on four different bird species—American Robin, Pileated Woodpecker, Blue Jay, and Carolina Wren. For each species, we sample 250 presence and 250 pseudo-absence statistics respectively. The summarized report for each bird species and their respective models can be found in Table 1. As shown, our Random Forest (RF) model performs with approximately 86% validation and lingers around 80% testing accuracy respectively. We also conclude that our model is separating the two classes with a high accuracy of 92% validation and 84% testing AUC accuracy. Our Gradient Boosting Tree (GBT) model performs with a slightly worse accuracy, an observation which presents some interesting facts about our dataset.

Since GBTs perform slightly worse than RFs, this discrepancy shows that our environmental features may include noise or unwanted information that may be skewing our results. RFs average this noise, hence provides better results than a GBT, a model which is more sensitive to noisy data. We will get back to this observation when we discuss future work in the conclusion section of our paper. Additionally, the features may separate presence/absence well, since RFs are effective when decision boundaries are clear. Hence, GBT may be overfitting based on some features, whereas RFs generate a more even spread of feature influence. Indeed, when we compare the importance of the features (see Figure 1), we notice that the RF implementation has a more even spread out of feature importance, a remark that shows that RFs may be building more generalized trees. However, when we compare the feature importance of RFs to Figure 2, the GBTs may be overfitting based on longitude.

We note, however, that longitude seems to be the most important feature in determining where bird species are located. This could indicate that birds are looking to migrate towards warmer areas or coastal bodies. Take both Figure 4 and Figure 3 as an example: notice how birds are strongly tied to coastal bodies as they provide food sources. With global warming, many food sources for birds may alter their presence and would thus affect bird migration.

When compared to the baseline Scikit-learn models, we notice that our custom implementations are similar in accuracy. This insight shows that our models are classifying the data accurately. Comparing the confusion matrices between the two RF models as shown in Figure 5 and Figure 6 we notice similarities in how well the model is predicting true positives and decently well in true negatives. This is most likely due to presence locations actually stemming from observations, whereas our pseudo-absent dataset is being generated and might not accurately reflect where birds may be absent.

Algorithm 6: Custom CNN for Bird Habitat Prediction

Input: Satellite imagery $X \in \mathbb{R}^{c \times 224 \times 224}$, labels $y \in \{0, 1\}^n$
Output: Species probabilities $p \in [0, 1]^n$

features ← Sequential(
  Conv2d(c, 64, kernel_size=7, stride=2, padding=3),
  BatchNorm2d(64), ReLU(), MaxPool2d(3, stride=2),
  Conv-BN-ReLU(64, 128) × 2, MaxPool2d(2),
  Conv-BN-ReLU(128, 256) × 2, MaxPool2d(2),
  Conv-BN-ReLU(256, 512) × 2, MaxPool2d(2)
)

classifier ← Sequential(
  AdaptiveAvgPool2d((1, 1)), Flatten(),
  Dropout(0.5), Linear(512, 256), ReLU(),
  Dropout(0.5), Linear(256, n), Sigmoid()
)

{Apply Kaiming initialization to all layers}
for each module m **do**
  Apply appropriate initialization based on layer type
end for

function forward(X)
$X \leftarrow$ features(X)
$p \leftarrow$ classifier(X)
return p
end function

Convolutional Neural Network

Significant differences were found between the experimental results of the ResNet transfer learning approach to building the model and the custom CNN built from scratch. We will compare the two models in the following categories: performance comparison, classification patterns in the form of confusion matrices, feature importance, and discriminative power. Finally, we will be discussing the results of the two wholistically and their relevance to solving the problem of habitat shift modeling.

First, from Figure 7 we can see that the ResNet model outperforms the CNN consistently across all performance metrics. For precision, ResNet shown an average precision of 84.3% versus 61.3% for the custom CNN. For Recall, ResNet shows an average of 92.2% compared to 69.3% of the CNN. For the F1 score, ResNet averages a score of 88% compared to the custom CNN of 65.8%. Lastly the accuracy of ResNet shows an average of 91% compared to the 61% of the CNN. This substantial difference in all categories showcases the limitations of the CNN architecture to fully capture the complex landscape features that would influence bird distribution patterns during climate-induced range shifts. It also shows the fundamental advantages of using pre-trained ResNet on images as a backbone for a transfer learning on satellite images rather than a custom CNN without the image training of ResNet34.

Next, when comparing the confusion matrices generated from both models as shown from Figure 8 and Figure 9 about the pileated woodpecker. We can see that the ovver-all accuracy differs greatly, 59.4% for the custom CNN and 86.9% for the ResNet. This might be the result of the insufficient model depth of the custom CNN's four block architecture compared to the ResNet architecture. It lacks the depth needed to capture complex ecological patterns. It also was not trained on anything other than the satellite image, which might be insufficient compared with ResNet which has already learned general image features. This allows the ResNet to adapt to the satellite images with relatively less additional data.

Table 1: Validation and Test Accuracy/AUC for Each Species and Model

Species	Model	Val Acc / AUC	Test Acc / AUC
American Robin	RF	0.857 / 0.924	0.803 / 0.842
	GBT	0.771 / 0.848	0.732 / 0.722
	SK-RF	0.857 / 0.938	0.789 / 0.809
	SK-GBT	0.886 / 0.942	0.831 / 0.785
Pileated Woodpecker	RF	0.829 / 0.872	0.873 / 0.895
	GBT	0.857 / 0.768	0.873 / 0.926
	SK-RF	0.829 / 0.910	0.916 / 0.914
	SK-GBT	0.800 / 0.884	0.887 / 0.864
Blue Jay	RF	0.914 / 0.958	0.845 / 0.837
	GBT	0.771 / 0.720	0.747 / 0.801
	SK-RF	0.914 / 0.962	0.817 / 0.869
	SK-GBT	0.886 / 0.936	0.761 / 0.843
Carolina Wren	RF	0.971 / 0.942	0.884 / 0.885
	GBT	0.941 / 0.981	0.899 / 0.872
	SK-RF	0.941 / 0.904	0.870 / 0.859
	SK-GBT	0.882 / 0.942	0.884 / 0.883

Thirdly, we must compare the feature importance between both the models itself. Looking at Figure 10 and Figure 11 we see that the custom CNN show more diffuse and less defined activation patterns compared to the ResNet one with a more structured and focused activations. ResNet was able to also demonstrate better spatial coherence where it was able to identify forest edges, water bodies, and vegetation patterns. These were features that the custom CNN was struggling to identify. The custom CNN had trouble with edge detection, where the transition zones should have shown as brighter activation patterns. This capability is relevant for predicting range shifts, as bird species follow habitat edges during migration.

Finally, we must discuss the difference in discriminative power between the custom CNN and ResNet. As shown in Figure 12 and Figure 13, the custom CNN's ROC curves show greater irregularity and proximity to the random classifier line, which is shown as a straight diagonal dotted line. On the other hand, the ResNet's curve rises sharply to the top left corner. This is an indicator of excellent separation between positive and negative cases. Within the custom CNN, it shows highly variable performance across species, while ResNet maintains consistently high AUC scores for all bird species predicted. This would suggest that the ResNet has

better generalization over diverse habitats preferences of the different bird species.

Comparing Tabular Data and ResNet/CNN

The first apparent difference between the two approaches other than the method in which a prediction is made, is the feature interpretation between these approaches. Although tabular models identified longitude as the most significant characteristic indicating latitudinal migration patterns with a strong preference for the coast, CNN models used activation maps that emphasized landscape elements such as water bodies, vegetation patterns, and edge boundaries to capture visual patterns. However, it was true that birds were often found also residing near large bodies as supported by the results of both approaches.

Interesting differences were also revealed by the classification patterns between the approaches. Tabular models demonstrated strong true positive identification across species with reasonable true negative performance, though potentially challenged by the artificial nature of the pseudo-absence data.

The ecological significance of both methods demonstrates how well they work together to model habitat shifts. Important climate and geographic factors influencing large-scale migration patterns were identified by tabular models that were in line with conventional climate envelope methodologies. By capturing landscape features that are difficult to measure in tabular data—particularly ecological transition zones that are crucial for migration—the CNN/ResNet models introduced a new dimension. Given that it captures both macro-level climate factors and micro-level landscape characteristics that impact habitat selection during climate-induced range shifts, an integrated approach that combines both methodologies may offer the most complete solution for predicting bird species distribution under climate change scenarios.

Conclusion

Our current work seems to be performing well, however, we note several areas for improvement. For one, we could consider other models like Logistic Regression, Support Vector Machine (SVM), or K-Nearest Neighbors [32]. Furthermore, we could look into changing the activation functions from a sigmoid to something like ReLU or tanh. We also want to look into combining our models to see whether the hybrid model performs better than having two separate models that predict based on features in image landscape and tabular data.

We also noticed that the GBT may be overfitting based on noise features in the dataset. We could improve our dataset for filtering for noise by providing more spatial resolutions of data points. Additionally, we could improve upon generating bird absence data points by gathering actual observation data for absent bird species' locations.

The more obvious limitation to our CNN/ResNet model is its small ecological dataset that the model was trained on[35]. Since ResNet was already pre-trained in general images, it performed better when confronted with new satellite images in comparison with CNN which only had the satellite images that it was provided. To improve this, more images might need to be used to train the custom CNN than the ResNet.

Another limitation to the CNN/ResNet models is its reliance on RGB bands from Sentinel-2 imagery, constraining the model to patterns visible to the human eye rather than the full spectral information available. By incorporating non-visible spectral bands from Sentinel-2 imagery, such as near-infrared, short-wave infrared, and red-edge bands would enable the detection of vegetation health, moisture content, and biomass[34]. As a result, these variables that could effect the bird habitat would be included.

Ultimately, we proposed a solution to predict bird species distributions. We hope that our approach will be used by ecologists in forecasting how bird habitats may shift in response to changing climate conditions. We wish our methods would contribute to policies to combat climate change and are inspired to further develop our approach to generate stronger models to predict where species are likely to relocate in the future.

Contributions

Min-Hong Shih was responsible for generating the CNN model; Emir Durakovic integrated the Tabular Data models. Both contributors contributed equally in writing the paper as well as helping to debug issues in the other's code.

Code

Our code can be found using this GitHub link: https://github.com/Matt940624/Bird-Species-Distribution-Modeling

References

[1] Piguet, E., Pécoud, A., & de Guchteneire, P. (2011). Migration and Climate Change: An Overview. *Refugee Survey Quarterly*, 30(3), 1–23. https://doi.org/10.1093/rsq/hdr006

[2] O'Shea, K., & Nash, R. (2015). An Introduction to Convolutional Neural Networks. *arXiv preprint arXiv:1511.08458*.

[3] McDonald-Madden, E., Runge, M. C., Possingham, H. P., and Martin, T. G. 2011. Optimal timing for managed relocation of species faced with climate change. *Nature Climate Change* 1: 261–265.

[4] Lemoine, N., and Böhning-Gaese, K. 2003. Potential Impact of Global Climate Change on Species Richness of Long-Distance Migrants. *Conservation Biology* 17(2): 577–586.

[5] Rahel, F. J., and Olden, J. D. 2008. Assessing the Effects of Climate Change on Aquatic Invasive Species. *Conservation Biology* 22(3): 521–533.

[6] Bickford, D., Howard, S. D., Ng, D. J. J., and Sheridan, J. A. 2010. Impacts of climate change on the amphibians and reptiles of Southeast Asia. *Biodiversity and Conservation* 19: 1043–1062.

[7] Morrison, M. L. 1986. Bird Populations as Indicators of Environmental Change. In Johnston, R. F., ed., *Current Ornithology*, volume 3. Springer, Boston, MA.

[8] Sullivan, B. L., Phillips, T., Dayer, A. A., Wood, C. L., Farnsworth, A., Iliff, M. J., Davies, I. J., Wiggins, A., Fink, D., Hochachka, W. M., Rodewald, A. D., Rosenberg, K. V., Bonney, R., and Kelling, S. 2017. Using open access observational data for conservation action: A case study for birds. *Biological Conservation* 208: 5–14.

[9] Pearson, R. G., and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* 12(5): 361–371.

[10] Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C., Aspinall, R., and Hastie, T. 2006. Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology* 43(3): 386–392.

[11] Ferrier, S., Powell, G. V. N., Richardson, K. S., Manion, G., Overton, J. M., Allnutt, T. F., Cameron, S. E., Mantle, K., Burgess, N. D., Faith, D. P., Lamoreux, J. F., Kier, G., Hijmans, R. J., Funk, V. A., Cassis, G. A., Fisher, B. L., Flemons, P., Lees, D., Lovett, J. C., and Van Rompaey, R. S. A. R. 2004. Mapping More of Terrestrial Biodiversity for Global Conservation Assessment. *BioScience* 54(12): 1101–1109.

[12] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[13] Si, S., Zhang, H., Keerthi, S. S., Mahajan, D., Dhillon, I. S., & Hsieh, C.-J. (2017). Gradient Boosted Decision Trees for High Dimensional Sparse Output. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 3182–3190). PMLR. https://proceedings.mlr.press/v70/si17a.html

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *arXiv preprint arXiv:1605.07146*.

[15] Teng, M., Elmustafa, A., Akera, B., Bengio, B., Radi, H., Larochelle, H., & Rolnick, D. (2023). SatBird: A Dataset for Bird Species Distribution Modeling using Remote Sensing and Citizen Science Data. In *Conference on Neural Information Processing Systems (NeurIPS), Datasets and Benchmarks Track*.

[16] Fukuda, S., De Baets, B., Onikura, N., Nakajima, J., Mukai, T., & Mouton, A. M. (2013). Modelling the distribution of the pan-continental invasive fish *Pseudorasbora parva* based on landscape features in the northern Kyushu Island, Japan. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 23(6), 901–910. https://doi.org/10.1002/aqc.2336

[17] Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: a citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282–2292.

[18] Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315.

[19] European Space Agency. (2023). Sentinel-2 User Guide. Retrieved from https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi

[20] Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* 3(2): 327-338.

[21] Prasad, A. M., Iverson, L. R., and Liaw, A. 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. *Ecosystems* 9(2): 181-199.

[22] Mi, C., Huettmann, F., Guo, Y., Han, X., and Wen, L. 2017. Why choose Random Forest to predict rare species distribution with few samples in large undersampled areas? Three Asian crane species models provide supporting evidence. *PeerJ* 5: e2849.

[23] Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., and Lawler, J. J. 2007. Random forests for classification in ecology. *Ecology* 88(11): 2783-2792.

[24] Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G. 2019. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* 10(2): 225-232.

[25] He, K., Zhang, X., Ren, S., and Sun, J. 2015. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778.

[26] Christin, S., Hervet, É., and Lecomte, N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10(10): 1632-1644.

[27] Buda, M., Maki, A., and Mazurowski, M. A. 2018. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* 106: 249-259.

[28] Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations*.

[29] Kattenborn, T., Eichel, J., and Fassnacht, F. E. 2019. Convolutional Neural Networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. *Scientific Reports* 9(1): 17656.

[30] He, K., Zhang, X., Ren, S., and Sun, J. 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, 1026-1034.

[31] Pecl, G. T., Araújo, M. B., Bell, J. D., Blanchard, J., Bonebrake, T. C., Chen, I. C., et al. 2017. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science* 355(6332): eaai9214.

[32] Cutler DR, Edwards TC, Beard KH, et al (2007) Random forests for classification in ecology. Ecology 88:2783–2792. doi: 10.1890/07-0539.1

[33] Arik S, Pfister T (2021) TabNet: Attentive interpretable tabular learning. Proceedings of the AAAI Conference on Artificial Intelligence 35:6679–6687. doi: 10.1609/aaai.v35i8.16826

[34] Kattenborn T, Eichel J, Fassnacht FE (2019) Convolutional neural networks enable efficient, accurate and fine-grained segmentation of plant species and communities from high-resolution UAV imagery. Scientific Reports. doi: 10.1038/s41598-019-53797-9

[35] Aodha OM, Cole E, Perona P (2019) Presence-only geographical priors for fine-grained image classification. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). doi: 10.1109/iccv.2019.00969

[36] Kim, J.-Y.; Yoon, J.; Choi, Y.-S.; and Eo, S. H. 2022. The influencing factors for distribution patterns of resident and migrant bird species richness along elevational gradients. *PeerJ* 10: e13258.

[37] Schekler, I.; Smolinsky, J. A.; Troupin, D.; et al. 2022. Bird migration at the edge – geographic and anthropogenic factors but not habitat properties drive season-specific spatial stopover distributions near wide ecological barriers. *Frontiers in Ecology and Evolution* 10: 822220.

[38] Teng, M.; Elmustafa, A.; Akera, B.; et al. 2023. Satbird: Bird species distribution modeling with remote sensing and citizen science data. *arXiv preprint arXiv:2311.00936*.

[39] Watling, J. I.; Brandt, L. A.; Mazzotti, F. J.; and Romañach, S. S. 2013. Use and interpretation of climate envelope models: A practical guide. *USGS Open File Report* 2013-1057.

[40] Weerasooriya, P. 2019. Implementing CART algorithm from scratch in Python. *Medium*.

[41] Withgott, J. 2000. Taking a bird's-eye view…in the UV: Recent studies reveal a surprising new picture of how birds see the world. *BioScience* 50(10): 854–859.

Appendix

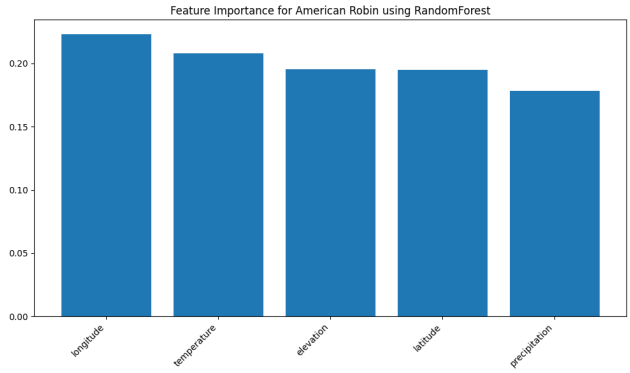


Figure 1: Feature importance graph for the Random Forest model.

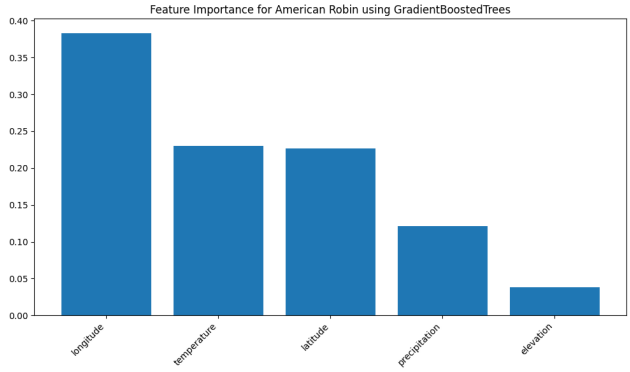


Figure 2: Feature importance graph for the Gradient Boosted Trees model.

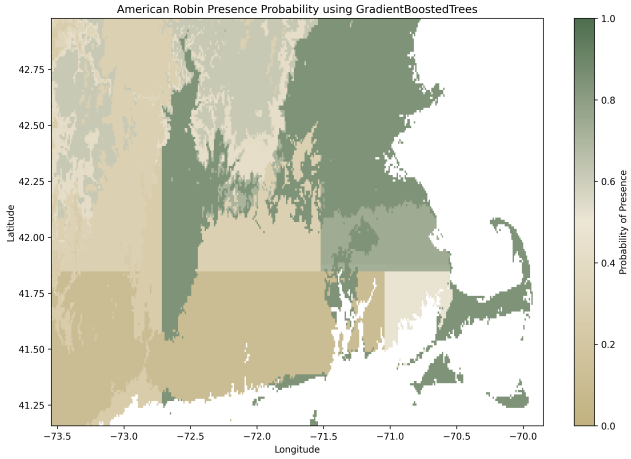


Figure 3: Predicted distribution map for the American Robin using the Gradient Boosted Trees model.

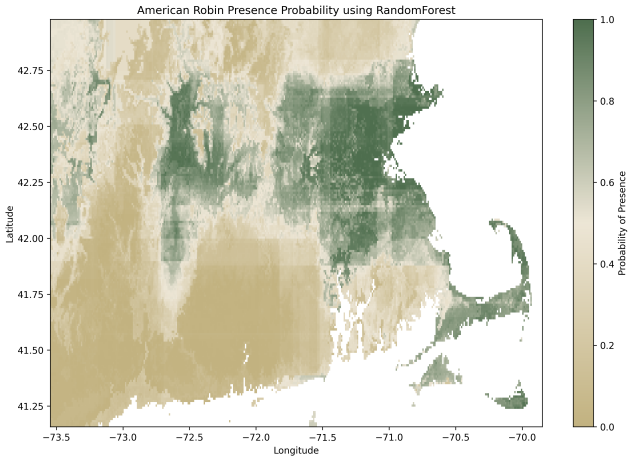


Figure 4: Predicted distribution map for the American Robin using the Random Forest model.

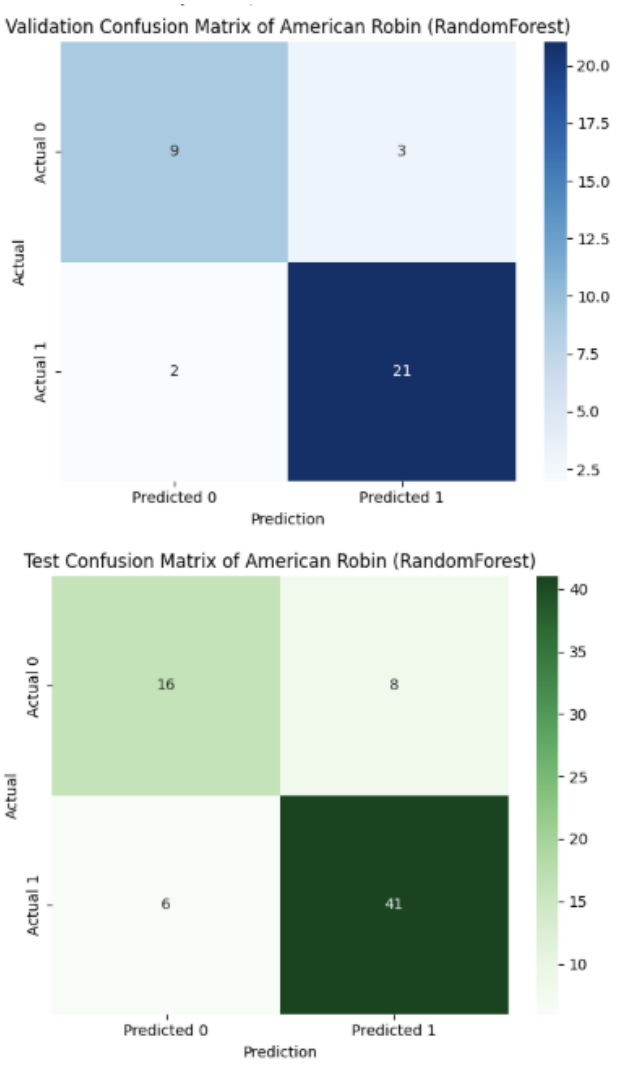
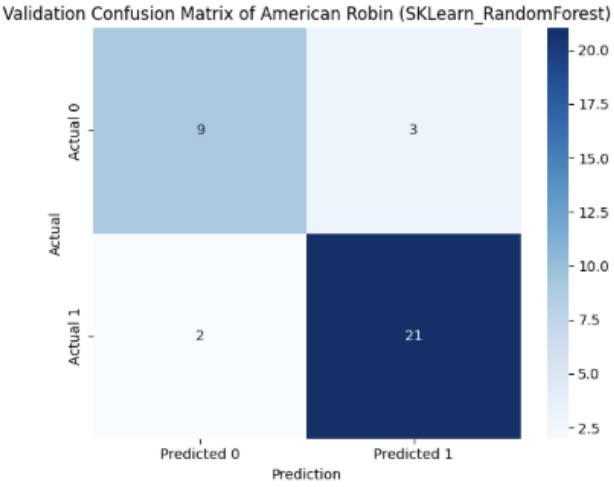


Figure 5: Random Forest Confusion Matrix.



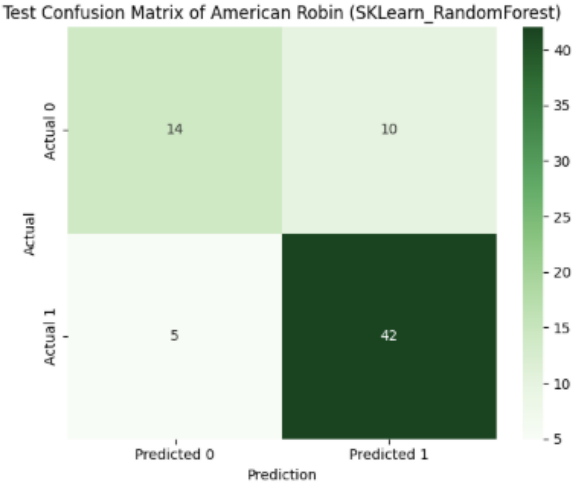


Figure 6: Scikit-learn Random Forest Confusion Matrix.

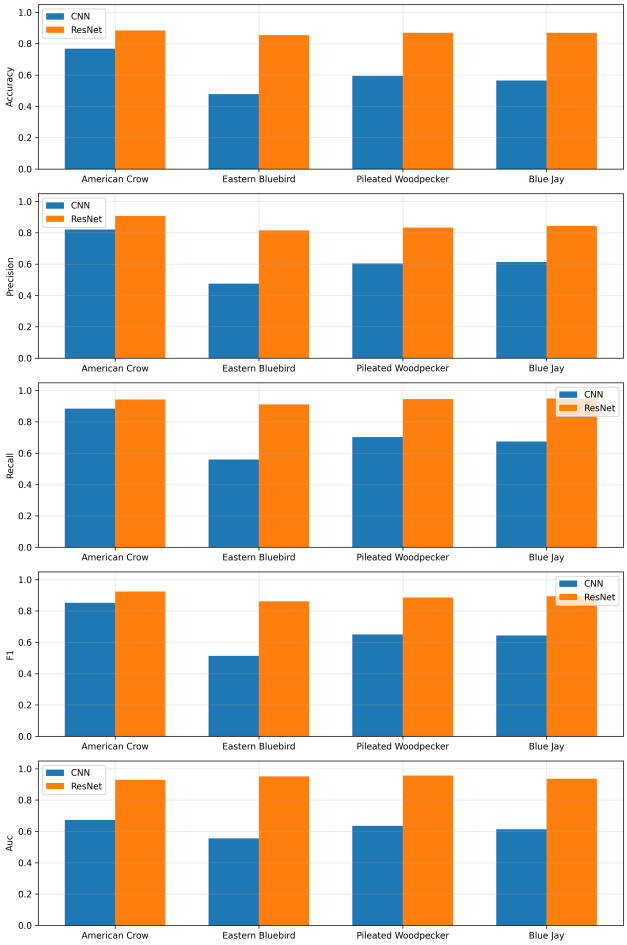


Figure 7: ResNet vs CNN model results comparison

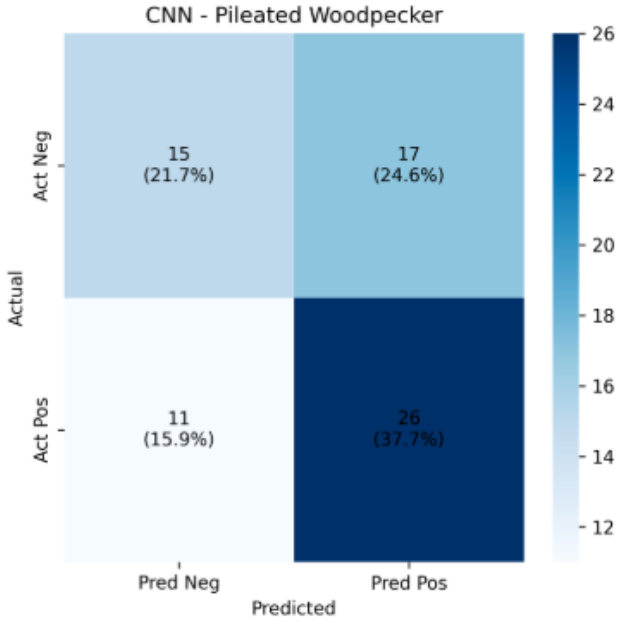


Figure 8: Cnn Confusion Matrix for Pileated WoodPecker

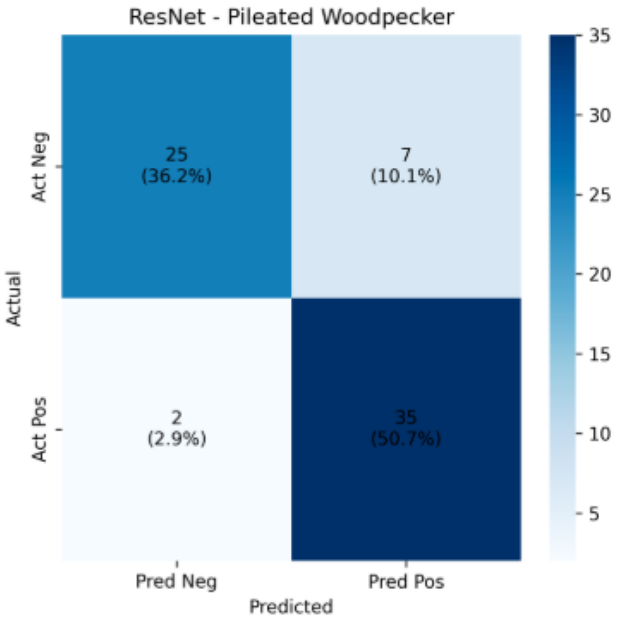


Figure 9: ResNet Confusion Matrix for Pileated Woodpecker

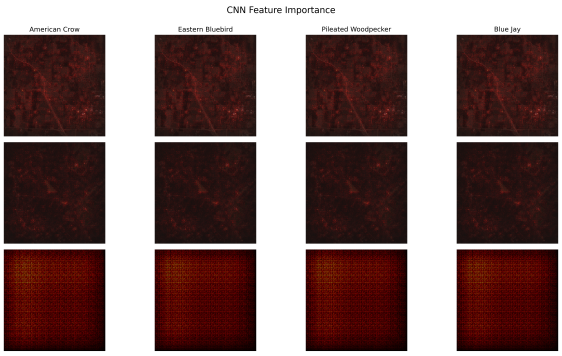


Figure 10: CNN Feature importance

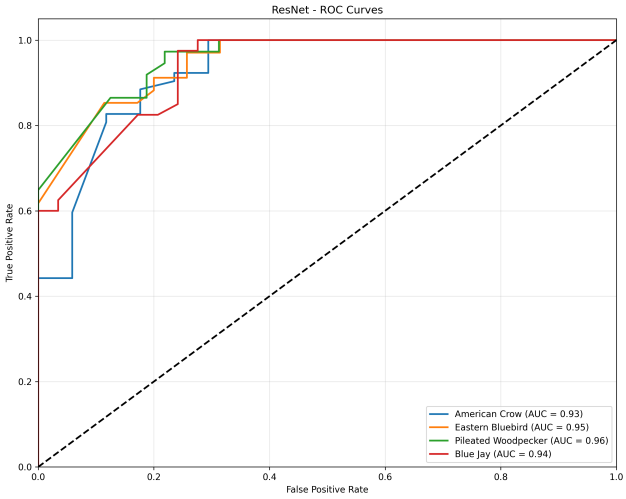


Figure 13: ResNet ROC curve

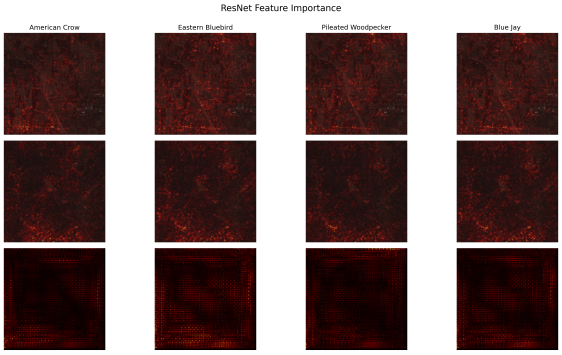


Figure 11: ResNet Feature Importance

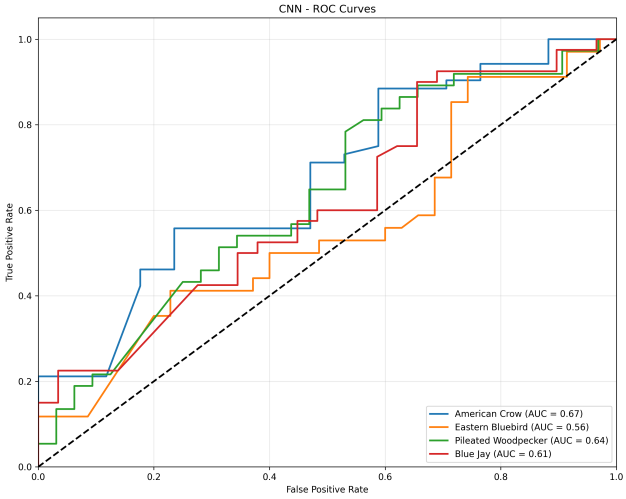


Figure 12: CNN ROC curve