Bird Identification Through Sounds using Convolution Neural Network

Ralph Bryllemond Y. Diaz,
Dr. Prospero C. Naval Jr.
Computer Vision & Machine Intelligence Group
Department of Computer Science
College of Engineering
University of the Philippines-Diliman

ABSTRACT

Birds play a vital role in their ecosystems and biodiversity. Though they are very agile and hard to see because of their natural behavior and habitat, their songs and calls are very evident. Since there is a uniqueness to their voice, songs and calls, they are prime candidates to monitor biodiversity and their function in their environments. In the Philippine context, there is a gap for studies in avian fauna. Thus, there is a need for accurate bird identification where birds present in the Philippines were the target. In this paper, a system for identifying bird species through their voice, calls, or songs is proposed. The system is divided into two phases: training phase and testing phase. The first stage is data acquisition where audio recordings were acquired and converted to data that can be interpreted by the model. The next stage entails converting the data into mel-spectrograms through data augmentation and preprocessing. Lastly, the third stage is the classifier where supervised learning is done through a Convolutional Neural Network (CNN). The inputs to the CNN were the mel-spectrograms obtained from the second stage. For training phase, the training dataset were fed to the CNN for training and pattern recognition. While, in the testing phase, the model is simulated and the extracted features were compared to the input to make a prediction about the identity of the bird species present in the recording.

1. INTRODUCTION

Accurate species identification is the basis for many aspects of research and a valuable component of various biological studies including medicine, ecology, and physiology. Additionally, the usefulness of accurate identification of species is necessary for many activities such as monitoring endangered species, observing biodiversity health, and determining the impacts of climate change on the species' distribution (Austen et al., 2016)[2]. Automating this task would greatly benefit researchers, farmers, foresters, conservation biologists, and ecologists considering the persisting problem of loss of diversity (Ceballos et al., 2015)[5].

The group of animals with the most interest and pressing need for observation and accurate identification are birds. Birds are considered good indicators of biodiversity response in the ecosystem they belong to. Due to their high activity rate, accessibility to observation practices, and sensitivity to the changes around them, researchers were able to determine other species' health in diversity through correlation with the birds' population and activities (Bowler et al., 2019)[3]. However, continuous human exploitation of the environment, such as urban expansion, illegal logging, and wildlife hunting, disturbs the whole environment. This leads to birds' populations disappearing, finding new habitats, or changing their migratory patterns (Xu et al., 2018)[26].

Thus, there is a need for the preservation and conservation of bird species in their natural habitats. With the advent of fast development of technologies and the rise of machine learning, researchers are now taking advantage of high-quality images and recordings to train models to determine the species captured (Waldchen & Mader, 2018)[22]. Various methods were developed for researchers to be able to gather effective data on birds. Some of these techniques include camera traps and acoustic captures. A study held by Chalmers et al. (2023)[6], determines that camera traps are expensive in terms of storage demand and equipment costs. This makes acoustic trails a promising prospect for identification techniques. This can be further effective as birds are often heard rather than seen. Additionally, they can be identified due to their unique bird calls or songs in varying frequencies (Fishbein, 2022)[7].

The goal of an accurate bird species identification system is to effectively and efficiently aid researchers, ecologists, and birdwatchers in monitoring various bird species in an ecosystem. Studies have shown relevant results when it comes to the proposal and comparison of various methods and techniques for accurate species identification (Waldchen & Mader, 2018)[22]. However, this is different when looking at the Philippine context. A limitation of currently available studies is that an application or a baseline model for a bird identification system is not widely available in the Philippines. Furthermore, there is a lack of data gathered for wider regions, with each study only offering a local survey of bird biodiversity in a certain area (Serrano et al., 2019)[18]. Nonetheless, national research about avian wildlife would prove valuable in the collation of data.

Overall, the paper aims to develop an effective bird identification model that recognizes bird species in present in the Philippines to provide a baseline model for avian biodiversity on-field monitoring,

2. METHODOLOGY

2.1 Dataset Description & Preprocessing

The audio dataset for the study consists of the Xeno-Canto website which consists of 879 relevant audio recordings of 175 bird species. For the current baseline model, 10 prominent species with a substantial number of available audio clips were selected for the dataset. The clips selected ensure that birds' songs or calls are included. Additionally, the dataset could possibly include environmental recordings with various difficulties such as multiple bird calls, different audio levels, wind, or other animal sounds. Therefore, there is a need for careful and proper data selection, clean-up, and validation.

For preprocessing, the following transformation was used to normalized the audio recordings to prepare them for the dataset: (a) Resampling and Conversion to Mono, (b) Resizing to fixed length, (c) Audio Augmentation, (d) Conversion to Mel-spectrogram, and (e) Spectrogram Augmentation.

The dataset would be normalized to an uncompressed WAV format (44.1kHz sampling frequency, 32 bits per sample, mono, 4 seconds duration of a random frame from the clip). Furthermore, recordings would be subjected to pre-processing steps that include filtering and data segmentation. Filtering through spectral gating reduces the background noise to improve the signal-to-average ratio (Kumar et al., 2023)[11].

Meanwhile, segmentation divides the recording into recognizable syllables for each bird call that would be useful in identification. Lastly, time shifting and pitch stretching have been used for data augmentation using the method described by Rajan et al. (2021)[17]. For spectrogram augmentation, the method used by Wang et al. (2021)[24] utilizes applying horizontal and vertical bars to the mel-spectrogram through time and frequency masking. Frequency masking is defined as randomly masking out consecutive frequencies by adding horizontal bars. Consequently, time masking is defined as randomly masking blocks of time by adding vertical bars to the spectrogram[14].

The preprocessing used by the paper is described as below.

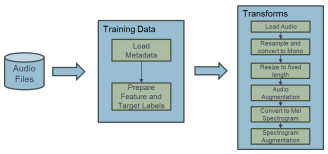

Figure 1: Enter Caption

2.2 Mel-Spectrogram Extraction

Extraction and selection of proper features are important for high classification accuracy in classification tasks. The Mel-frequency Cepstral Coefficients (MFCC) and Mel-spectrograms are the most popular and valuable feature selection methods for voice recognition. Plotting the vocalization of bird species in a spectrogram is an effective method for the identification of birds especially those who can be identified with their certain call patterns and frequencies. The *librosa* and the *torchaudio* python package can be used for computations of MFCCs and Mel-spectrogram in the front-end. Meanwhile, a study by Rajan et al. (2021)[17] suggests that the time-domain waveform is converted to a time-frequency representation using the Short-Time Fourier Transform (STFT) with a frame size of 30 ms and a hop size of 10 ms. Lastly, the linear frequency spectrogram will be converted to a mel-scale which can preserve important harmonic characteristics.

After the audio clips were preprocessed, there is a need to get valuable insight of the datasets produced. One technique proposed is to extract meaningful data through the conversion of the sound into useful data visualization. This is done by applying Mel-spectrogram to the given data.

In the Mel-spectrogram Extraction, the following theories were used.

2.2.1 Fourier Transform

The Fourier Transform is a mathematical formula that enables a signal to be decomposed into individual frequencies and frequency's amplitude. This essentially converts the signal from time-domain into frequency-domain. Meanwhile, the Fast Fourier Transform (FFT) is an algorithm that is used to compute for the Fourier Transform of an audio[16].

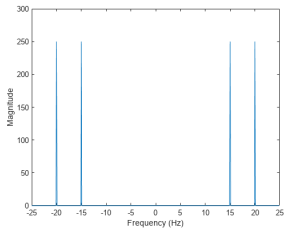

Figure 2: Fourier Transform

2.2.2 Spectrogram

Though the Fourier Transform is a powerful tool for analyzing the frequency content of a signal, most bird songs and other audio signals are non-periodic. This means that there is a need to represent the spectrum of signals as they change over time. The Short-Time Fourier Transform (STFT) computes the signal on overlapping windowed segments and this results into spectrograms. This gives a visual representation on a signal's loudness strength as it varies over time. Additionally, the y-axis is converted to a log scale and colors are represented by decibels. This is useful for manual analyzation as humans can only perceive limited range of frequencies and amplitudes.

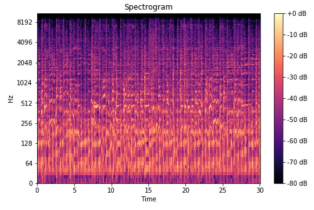

Figure 3: Spectrogram

2.2.3 *Mel-Spectrogram*

Since humans do not perceive frequencies on a linear scale, mel-scale is introduced such that the pitch sounded equally distant to the listener. A Mel-spectrogram converts these frequencies through the application of the mel-scale. This is done by mapping the y-axis or the frequency onto the mel-scale [16].

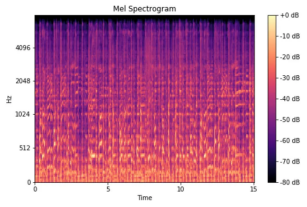

Figure 4: Mel-spectrogram

The following figures show mel-spectrograms produced by sample audio clips from White Vented Whistler and Balicassiao.

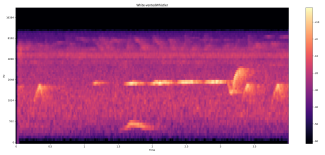

Figure 5: White Vented Whistler Mel-Spectrogram

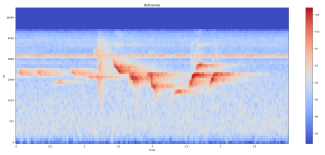

Figure 6: Balicassiao Mel-Spectrogram

2.3 Classifier Module

The classifier module is the most important part of the model. This is where training and testing of the dataset is done. The objective of the training phase is to create a model that is consistent and fits with the pattern representation of an individual class. Thus for each bird call or song pattern of a species, the goal is for the specific species to be identified. Meanwhile, the representation can be a pattern itself, or a signal model that characterizes the statistical variations of the bird species class. The function of the testing phase is to get the vector of features and compare them with the patterns produced in the training phase. The input voice and each voice pattern is computed. Moreover, the best match is considered to be the bird species that called.

In this study, the artificial neural network approach is used.

2.3.1 *Artificial Neural Network*

Artificial neural networks attempts to mimic how a person applies their intelligence to learn or recognize patterns and tasks. It visualizes, analyzes, and makes decisions on the given features similar to how a person thinks[8]. Since neural networks have been developed by generalizing human's cognitive behavior using mathematical models, they can be characterized by:

- The pattern of connections similar to neurons. The architecture usually represents the connection between neurons.
- Method of determining weights on the connections between neurons. Represented by the learning algorithm.
- Activation functions.

In audio classification, the most common neural network is the Convolution Neural Network (CNN). Therefore, the paper utilized Convolution Neural Network (CNN) to produce a bird identification model. It is important to note that selecting classical image classifications such as kNN is not desirable due to the complexity of images (Wang, Q. et al., 2019)[23].

2.3.2 *Convolution Neural Network*

Generally, CNNs are good image and acoustic classifiers. The CNN architecture utilizes deep feed-forward approach that enables it to generalize compared to networks with fully connected layers. Indolia et al. (2018)[9] describes the strenght of CNN lies on its capability of weight sharing which reduces the number of parameters that needs training and lessening the probability of overfitting. In addition, the classification section also includes feature extraction which benefits from the learning process. Lastly, CNN implementation, compared to other models of artificial neural networks, is much easier (Tivive, et al., 2005)[21].

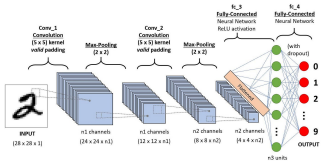

Figure 7: A Convolution Neural Network

For the current status of the project, a generalized CNN is used. It has a single input and output layer along with multiple hidden layers. A neuron takes input of vector X and produces output Y by performing some function F on it with a W denoting the weight vector denoting the strength of connection between two neurons or layers. It is represented by

$$F(X, W) = Y$$

The model consists of four components: (a) convolution layer, (b) pooling layer, (c) activation function, and (d) fully connected layer.

2.3.3 *Convolution Layer*

The convolution layer is considered as the main building block of the CNN. The image to be classified is fed to the

input layer and output is the predicted class label computed using extracted features from the image. The local features from the input are extracted using receptive field. The receptive field of a neuron that is associated in previous layer forms a weight vector. As neurons share same weights, similar features at different locations in the input data are detected[12][15].

For the paper, four blocks of convolution layer are implemented. The first layer accepts [3x1x244x244] which is the shape of the mel-spectrogram. The four convolution layer is described as

$$self.conv1 = nn.Conv2d(1, 32, 3, 1)$$

$$self.conv2 = nn.Conv2d(32, 64, 3, 1)$$

$$self.conv3 = nn.Conv2d(64, 128, 3, 1)$$

$$self.conv4 = nn.Conv2d(128, 256, 3, 1)$$

2.3.4 Pooling Layer

The convolution layer is followed by pooling layer which reduces number of trainable parameters. It basically reduce the spatial dimensions of the feature maps while preserving depth. Few pooling approaches exist such as average pooling and max pooling[27].

An adaptive average pooling is implemented for the project. It is described as

$$self.pooling = nn.AdaptiveAvgPool2d((8, 8))$$

2.3.5 Activation Function

Activation function's main purpose is whether to decide if a neuron should be activated or not. This means it considers if the neuron's input in the network will be helpful. There are many activation functions used. However, common functions used are the linear, sigmoid, and ReLU (Rectified Linear Unit) functions. For the paper, the ReLU function is used for the activation function due to its back propagation capability and efficiency in terms of computation. [9].

It is described in the model as

$$x = F.relu(x)$$

Meanwhile, its mathematical representation is described as

$$f(x) = max(0, x)$$

2.3.6 Fully Connected Layer

Fully connected layer refer to a neuron which applies linear transformation to the input vector. The current layer's neurons are connected to the neurons from the previous layer which completes its linkage. Lastly, the layer is responsible for producing the final output or prediction[1].

The paper has implemented two Fully Connected Layer. The last fully connected layer should have the same number of output classes described by the paper which are 10 classes. They are described as

$$self.fc1 = nn.Linear(16384, 128)$$

$$self.fc2 = nn.Linear(128, 10)$$

2.3.7 Learning Algorithm

Optimization is another key in maximizing the network's capability in learning. Learning algorithms minimized the objective or loss function depending on some parameters like weights and biases[19].

The learning approach that the paper utilized is the ADAM (Adaptive Moment Estimation) Optimization. The ADAM proposed by [10] uses first and second moment of gradient for computation of the learning rates. This means that it is computationally efficient and performs well on large datasets. The algorithm of the ADAM is described as below

Figure 8: ADAM Optimization Algorithm

ADAM implementation is described as

$$optimizer = torch.optim.Adam(model.parameters(), lr = 0.001)$$

Scheduling learning rate is also an essential step for CNNs. This enables the gradual decrease of learning rate over time. This improves the stability and accuracy of the model through smoothing and convergence to the optimal solutions [25].

2.3.8 Testing and Simulation

For the testing phase, the network is simulated. The test audio features are input to the network and the network returns a vector of size 10 which has an associated node depicting the class or simply the bird species determined to match the pattern if the value is greater than or equal than the threshold. The system outputs the bird species identified, otherwise, it returns a message saying that the bird species is unrecognized by the machine.

For the analyzation of the performance of the model, the paper applied the strategy described in [13] . It is based on the strategy which applies Average Accuracy, Average Precision, Average Loss and Confusion Matrix as metrics for analysis of the given CNN.

3. EXPERIMENTS

3.1 Environment

The system runs under the Windows Operating System and uses Jupyter Notebook and Visual for its user interface. The environment in which the model and network is executed is listed as

- Hardware: ASUS TUF A15
- Software: Windows 11
- User Interface: Visual Studio Code & Jupyter Notebook
- Language: Python
- Compiler: Python 3.9.2

3.2 Specifications

The system is trained to recognize 10 bird species with varying bird sounds classified as song, mating call, flight call, and alarm call. One pitfall for the dataset gathered is that it has an unbalanced number of data for each class. The audio recordings have to be recorded and normalized following these specifications:

- Sampling Rate: 44.1 kHz
- Format: Uncompressed WAV File (32 bits, Mono)

Additionally, each audio recording were converted to mel-spectrograms. Both the dataset for training and validation follows the same format mentioned before. These were the inputs for the training of the model. The network was trained for 50 epochs. The system then computes the prediction after taking in the inputs. It will then display the bird species present in the recording (or an "unknown species" if there is a low confidence).

3.3 Experimental Results

Three bird classification tests were performed each with a different number of recordings for the dataset. Each run was done in 50 epochs and performance metrics were determined using Validation & Training Loss, Accuracy, and Confusion Matrix.

3.3.1 1st Experimental Run

The first experiment run was done in 50 epochs. Overall recognition rate is classified to be about 60%. The following is the result of the 1st experimental run:

Bird Species	Prediction (Correct/Total)	Accuracy
Philippine Bulbul	23/43	0.535
White-browed Short Wing	33/42	0.786
Philippine Coucal	9/29	0.310
Long-tailed Bush Warbler	29/30	0.967
Luzon Hawk-Owl	30/34	0.882
White-eared Brown Dove	29/46	0.630
White-vented Whistler	15/30	0.500
Balicassiao	17/57	0.298
Elegant Tit	11/26	0.423
Grey-backed Tailorbird	31/41	0.756

Table 1: Results of 10 Simulations of the 1st Network

The performance metrics for the 1st Network shows an average validation loss significantly greater than the training loss. Meanwhile, the validation and training accuracy diverges around 13 epochs, with the training accuracy being significantly higher than the validation accuracy. This are seen on figures 9 and 10.

3.3.2 2nd Experimental Run

The second bird identification test also used the same classes set with a decrease on the number of data available for the dataset. The test is done in 50 epochs. Furthermore, the test were run under the same conditions as the 1st network. It achieved an accuracy of about 0.57%. This is a decrease on the performance of the 1st experimental run.

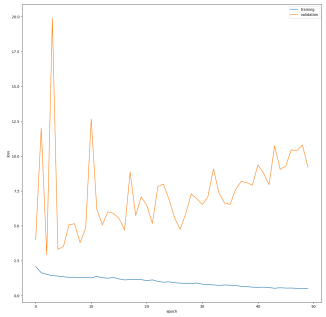

Figure 9: Validation and Training Loss for the 1st Network

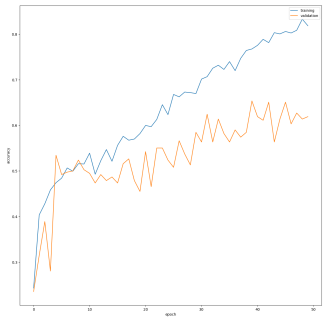

Figure 10: Validation and Training Accuracy for the 1st Network

Meanwhile, its performance exhibits the same characterizations as the 1st experimental run. However, the losses are much more nearer to each other and the accuracy of both training and validation were near each other until it converges at around epoch 35. The graphs still show erratic numbers as shown on figures 11 & 12. The result of the 2nd experimental run is described in table 2.

Bird Species	Prediction (Correct/Total)	Accuracy
Philippine Bulbul	14/26	0.538
White-browed Short Wing	7/17	0.412
Philippine Coucal	8/20	0.400
Long-tailed Bush Warbler	21/22	0.955
Luzon Hawk-Owl	14/17	0.824
White-eared Brown Dove	13/26	0.500
White-vented Whistler	9/15	0.600
Balicassiao	5/20	0.250
Elegant Tit	9/16	0.563
Grey-backed Tailorbird	8/10	0.800

Table 2: Results of 10 Simulations of the 2nd Network

3.3.3 3rd Experimental Run

Meanwhile, the 3rd test also follow the exact classes described previously with a different number of data available in the dataset. The 3rd test show the same characteristics despite having a different learning rate and scheduler (0.001 instead of 0.0001). It achieved an accuracy of about 0.45%. The performance is the worst among the three experimental runs. In terms of performance, it shows similar characteristics as the previous two experimental runs as shown on figure 13. The result of the 3rd experimental run is described in table 3.

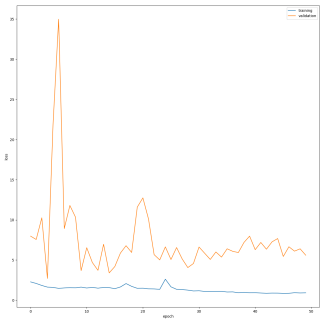

Figure 11: Validation and Training Loss for the 2nd Network

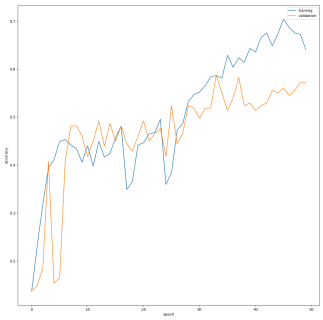

Figure 12: Validation and Training Accuracy for the 2nd Network

Bird Species	Prediction (Correct/Total)	Accuracy
Philippine Bulbul	17/30	0.567
White-browed Short Wing	10/26	0.385
Philippine Coucal	17/29	0.586
Long-tailed Bush Warbler	23/28	0.821
Luzon Hawk-Owl	12/32	0.375
White-eared Brown Dove	14/37	0.378
White-vented Whistler	8/25	0.320
Balicassiao	5/34	0.147
Elegant Tit	10/25	0.400
Grey-backed Tailorbird	10/18	0.556

Table 3: Results of 10 Simulations of the 3rd Network

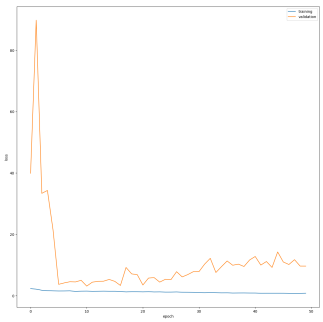

Figure 13: Validation and Training Loss for the 2nd Network

4. SUMMARY AND CONCLUSION

Bird identification task requires the machine to be able to identify the bird species based on calls or songs of the given bird. The theories applying to bird identification have been covered and computational techniques and algorithms used were presented. The scope of this work is limited by the available dataset on Xeno-Canto about the birds present in the Philippines. The species covered in the paper are the most prominent in terms of number of data recordings available. Though, a pitfall of the dataset from Xeno-Canto is that the dataset is heavily unbalanced. Consequently, training for each species whose number of data recordings are low might be difficult when it comes to generalization and requires careful data augmentation.

In this study, a new baseline for bird identification system for the birds present in the Philippines that utilizes mel-spectrogram to convert audio recordings to visual data that is useful for both human and machine analysis. In addition, the application of the Convolutional Neural Network (CNN) and enables bird species identification through their songs or calls. Three bird identification tests were performed. The first test yielded 60% on its accuracy, the second test yielded 57%, and lastly, the third test yielded 45% accuracy. All tests were run for 50 epochs.

All three performance tests resulted in significant validation loss which means that the model tends to overfit. This essentially hampers the learning process of the model. A tweak to a hyper-parameter, upgraded architecture, and careful selection of dataset could improve the model's learning capacity and its accuracy. From this, the researcher concluded that there is a need for improvement to clean up data and experiment more with CNN architectures. Having a good CNN model could greatly improve the chances of generalization and ease the implementation of other bird species for identification. Thus, when a human struggles to learn from his current approach, it is effective for a human to change his technique and find a way to bring out his best. This is the same for machines.

5. REFERENCES

[1] Alzubaidi, L., Zhang, J., & Humaidi, A. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 8, 53. doi:10.1186/s40537-021-00444-8

[2] Austen, G. E., Bindemann, M., Griffiths, R. A., & Roberts, D. L. (2016). Species identification by experts and non-experts: Comparing images from field guides. Scientific Reports, 6.

[3] Bowler, D., Heldbjerg, H., Fox, A., de Jong, M., & Böhning-Gaese, K. (2019). Long-term declines of European insectivorous bird populations and potential causes. Conservation Biology, 1120–1130.

[4] Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X., Raich, R., Hadley, S.& Betts, M. (2012). Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. J. Acoust. Soc. Am., 131 (6), 4640–4650.

[5] Ceballos, G., Ehrlich, P. R., Barnosky, A. D., García, A., Pringle, R. M., & Palmer, T. M. (2015).

Accelerated modern human–induced species losses: Entering the sixth mass extinction. Science Advances, 1(5).

[6] Chalmers, C., Fergus, P., Wich, S., Longmore, S. N., Walsh, N. D., Stephens, P. A., ... Nuseibeh, A. (2023). Removing Human Bottlenecks in Bird Classification Using Camera Trap Images and Deep Learning. Remote Sensing, 15(10).

[7] Fishbein, A. (2022). How Birds Hear Birdsong. Scientific American, 326, 36–43.

[8] Han, S., Kim, K., Kim, S., & Youn, Y. (2018). Artificial Neural Network: Understanding the Basic Concepts without Mathematics. Dementia and Neurocognitive Disorders, 17(3), 83–89.

[9] S. Indolia, Goswami, A., Mishra, S. P., & P.Asopa. (2018). Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach. Procedia Computer Science, 132, 679–688. doi:10.1016/j.procs.2018.05.069

[10] Kingma, D., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv Preprint arXiv:1412. 6980.

[11] Kumar, E., Surya, K., Varma, K., Akash, A., & Reddy, K. (2018). Noise Reduction in Audio File Using Spectral Gatting and FFT by Python Modules. Volume 32: Recent Developments in Electronics and Communication Systems, 624, 510–515.

[12] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature 521 (7553), 436–444.

[13] Nasierding, G., & Kouzani, A. (2012). Comparative evaluation of multi-label classification methods. 2012 9th International Conference on Fuzzy Systems and Knowledge Discovery, 679–683.

[14] Necciari, T., Balazs, P., Kronland-Martinet, R., Ystad, S., Laback, B., Savel, S., & Meunier, S. (2012). Auditory Time-Frequency Masking: Psychoacoustical Data and Application to Audio Representations. 146–171.

[15] Palsson, F., Sveinsson, J.R., & Ulfarsson M.O. (2017). Multispectral and Hyperspectral Image Fusion Using a 3-D-Convolutional Neural Network. IEEE Geoscience and Remote Sensing Letters 14 (5), 639–643.

[16] Rabiner, L., & Schafer, R. (2010). Theory and Applications of Digital Speech Processing.

[17] Rajan, R., & A, N. (2021). Multi-label Bird Species Classification Using Transfer Learning. 2021 International Conference on Communication, Control and Information Sciences (ICCISc), 1–5.

[18] Serrano, J. E., Guerrero, J. J., Quimpo, J. D., Andes, G. C., Bañares, E. N., & General, M. A. (2019). Avifauna Survey within a University Campus and Adjacent Forest Fragment in Bicol, Eastern Philippines. Applied Environmental Research, 41(2), 84–95.

[19] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv Preprint arXiv:1409. 1556.

[20] Tanalgo, K. C., Achondo, M. J., & Hughes, A. C. (2019). Small Things Matter: The Value of Rapid Biodiversity Surveys to Understanding Local Bird Diversity Patterns in Southcentral Mindanao. Tropical Conservation Science, 12.

[21] Tivive, F., & Bouzerdoum, A. (2005). Efficient training algorithms for a class of shunting inhibitory convolutional neural networks. IEEE Transactions on Neural Networks 16, 3, 541–556.

[22] Wäldchen, J., & Mäder, P. (2018). Machine learning for image-based species identification. Methods Ecol Evol., 9, 2216–2222.

[23] Wang, Q., Jia, N., & Breckon, T. (2019). A Baseline for Multi-Label Image Classification Using An Ensemble of Deep Convolutional Neural Networks. arXiv [Cs.CV]. Retrieved from http://arxiv.org/abs/1811.08412

[24] Wang, H., Zou, Y., & Wang, W. (2021). SpecAugment++: A Hidden Space Data Augmentation Method for Acoustic Scene Classification. arXiv [Eess.AS]. Retrieved from http://arxiv.org/abs/2103.16858

[25] Wen, L., Li, X., & Gao, L. (12 2020). A New Reinforcement Learning Based Learning Rate Scheduler for Convolutional Neural Network in Fault Classification. IEEE Transactions on Industrial Electronics, PP, 1–1. doi:10.1109/TIE.2020.3044808

[26] Xu, X., Xie, Y., Qi, K., Luo, Z., & Wang, X. (2018). Detecting the response of bird communities and biodiversity to habitat loss and fragmentation due to urbanization. Science of The Total Environment, 624, 1561–1576.

[27] Zhou, Y., Wang, H., Xu, F., & Jin, Y. (2016). Polarimetric SAR image classification using deep convolutional neural networks. IEEE Geoscience and Remote Sensing Letters 13 (12).