IJNRD.ORG ISSN : 2456-4184

INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG

An International Open Access, Peer-reviewed, Refereed Journal

BIRD CLASSIFICATION BASED ON IMAGE OR AUDIO USING DEEP LEARNING

[1]Mr. Shreejith K B, [2]Maanikya, [3]Poojary Dheeraj Kumar, [4]Rachana Nayak, [5]Rakshit Lingappa Poojari

[1]Senior Assitant Professor, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India
[2] Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India
[3] Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India
[4] Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India
[5] Student, Mangalore Institute of Technology and Engineering, Moodabidre, Karnataka, India

Abstract: ***Birds are a vital group of animals that ecologists monitor using autonomous recording units as a crucial indicator of the health of an environment. Bird-watching is a popular hobby which offers relaxation in everyday life. Innumerable people visit bird sanctuaries to observe different species. Nowadays some bird species are found rarely and if found, classification of bird species prediction of the same is difficult. Numerous bird species have become extinct because of anthropogenic activities and climate change. Habitat destruction is a significant threat to biodiversity worldwide. Thus, monitoring the distribution of species and identifying the elements that make up the biodiversity of a region are essential for designing conservation stratagems. Bird classification has been an important task in the field of ornithology and wildlife conservation. With deep learning advancements, image and audio-based bird classification methods have gained significant attention. We employ convolutional neural networks (CNNs) to learn discriminative features from bird images and audio. We use a large dataset of bird images to train the CNN model. This model is capable of automatically extracting high-level features from images, audio and classifying birds into different species with high accuracy based on deep learning techniques using either images or audio data.***

Keywords - ***Bird Classification, Image Based, Audio Based, Biodiversity, CNN***

INTRODUCTION

In today's scenario, bird behavior and population has become an important issue. Birds help to detect other organisms in the environment. Basically, bird species identification from their sound is an important and challenging problem. There are also some different methods through which we can monitor bird species. As many birds migrate according to the environmental changes, so the use of automated methods for bird species identification is an effective way to evaluate the quantity and diversity of the birds which appear in the region.

Artificial intelligence and machine learning sounded like a science fiction prophecy of a technological feature. Image recognition is one of the most accessible applications of it. Deep Learning is a Machine Learning subfield which is in turn a subfield of Artificial Intelligence. Deep learning can be visualized as a platform where artificial, human brain inspired neural networks and algorithms learn from large amounts of data. Deep Learning allows computers to solve complex problems even though they use a very diverse, unstructured, and interconnected data set. The more Deep Learning algorithms learn, the better they perform. Nowadays, bird species identification is seen as a perplexing problem which often leads to confusion. Birds allow us to search for certain species within the ecosystem as they react rapidly to changes in the atmosphere; but collecting and gathering information on birds needs tremendous human effort. Many people visit bird sanctuaries to look at the birds, while they barely recognize the differences between different species of birds and their characteristics. Understanding such differences between species can increase our knowledge of birds, their ecosystems and their biodiversity.

The identification of birds with bare eyes is based solely on the basic characteristics due to observer constraints such as location, distance and equipment, and appropriate classification based on specific characteristics is often found to be tedious. Ornithologists have also faced difficulties in distinguishing bird species. To properly identify a particular bird, they need to have all the specificities of birds, such as their distribution, genetics, breeding climate and environmental impact. A robust system is needed for all these circumstances that can provide processing of large-scale bird information and serve as a valuable tool for scholars, researchers, and other agencies. The identification of the bird species from the input of sample data therefore plays an important role here. Bird identification can generally be done with the images, audio, or video.

Bird classification based on image or audio using deep learning has a wide range of applications in ornithology and conservation biology. For instance, these automated systems can be used to monitor bird populations in the wild, track bird migration patterns, and study bird behaviors, such as feeding, nesting, and mating. They can also aid in the identification of rare or endangered bird species, which can help in designing effective conservation strategies. Furthermore, these systems can be used by birdwatchers and citizen scientists to identify bird species in the field, thereby promoting citizen science initiatives and engaging the public in bird conservation efforts.

Despite the promising potential of deep learning for bird classification, there are several challenges that researchers are currently addressing. One major challenge is the need for large and diverse datasets to train accurate deep learning models, as obtaining such datasets can be time-consuming and resource-intensive. Another challenge is the need for robust models that can perform well under varying environmental conditions, such as changes in lighting or background noise. Additionally, the interpretability of deep learning models for bird classification is still an active area of research, as understanding the decision-making process of these models can be complex and challenging.

Bird classification based on image or audio using deep learning is a promising field that has the potential to revolutionize the way bird species are identified and monitored. The advancements in deep learning algorithms, the availability of large datasets, and the diverse applications in ornithology and conservation biology make this field an exciting area of research. However, challenges still exist, and further research is needed to improve the accuracy, robustness, and interpretability of deep learning models for bird classification. Nevertheless, with continued advancements in technology and research, deep learning-based bird classification systems have the potential to significantly contribute to our understanding and conservation of avian biodiversity.

NEED OF THE STUDY.

Birds are an important group of Birds that ecologist monitor using autonomous recordings units as a crucial indicator of health of an environment. There is not yet an adequate method for automated bird call recognition in acoustic recordings due to high variations in bird calls and the challenges associated with bird call recognition. We do not have an effective way to classify birds for a common man, especially for those who are into for Birds-Observation or Analysis oriented Hobbies and or Professions. Our application helps common bird enthusiasts, researchers, photographers, and others to identify Bird Species based on the image captured or the audio. One important group of Bird that ecologist monitor in acoustic recordings are birds. Birds are regarded as an important indicator of biodiversity as the number and diversity of bird species in an ecosystem can directly reflect biodiversity, ecosystem health and suitability of the habitat

The goal of bird classification using a CNN model is to identify bird species accurately and reliably from images or audio recordings, leveraging the power of convolutional neural networks (CNNs) to automatically learn and extract relevant features from the data. The ultimate objective is to achieve high accuracy in bird species classification, enabling applications such as bird species identification, biodiversity monitoring, and ecological research. Additionally, the goal may also include optimizing the model for efficiency, scalability, and deploy ability on various platforms, such as mobile devices, to enable real-time bird classification in the field.

LITERATURE REVIEW

1. Recognition of Endemic Bird Species Using Deep Learning Models by Yo-Ping Huang and Haobijam Basanta:

In this study, they used Inception-ResNet-v2, which is a hybrid convolutional neural network (CNN) architecture of Inception and a residual network connection. These modules were incorporated with different configuration parameters that make use of the Inception approach by internally attached residual connections with the entire Inception part of the module by replacing the filter concatenation stage of the Inception architecture. This model achieved an accuracy of 98.39% in the classification of 29 endemic bird species and an accuracy of 100% in the detection of birds among different object categories. Moreover, the model achieved a precision, recall, and F1-score of 98.49%, 97.50%, and 97.90%, respectively, in the classification of bird species.

2. Automatic acoustic detection of birds through deep learning by Dan Stowell, Michael D. Wood, Hanna Pamuła, Yannis Stylianou, Hervé Glotin:

To conduct the evaluation campaign, we designed a detection task to be solved—specific but illustrative of general-purpose detection issues—gathered multiple datasets and annotated them, and then led a public campaign evaluating the results submitted by various teams. After the campaign, we performed detailed analysis of the system outputs, inspecting questions of accuracy, generality, and calibration. In revalidating the testing set, we examined those items with the strongest mismatch between manual and automatic detection, to determine which was in error: 500 presumed negative and 1,243 presumed positive items. This showed inter-rater disagreement in 16.6% of such cases predominantly, the most ambiguous cases with barely audible bird sounds with amplitude close to the noise threshold.

3. Bird Call Recognition using Deep Convolutional Neural Network, ResNet-50 by Mangalam Sankupellay and Dmitry Konovalov:

The Inception-v4 architecture is an architecture that utilizes residual learning (Szegedy et al, 2016). In CNN, as the layers get larger, training of deep-CNN becomes difficult and the accuracy starts to saturate and then degrade. Residual learning help solve this degrading accuracy problem (He et al. 2016). Residual learning uses shortcut connections as a training method to directly connect input to some other subsequent layers (not just to the next adjacent layer), to train deepCNN. The ResNet-50 (a 50 layer deep-CNN architecture), is the first deep-CNN architecture that utilized residual learning.

4. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Mingxing Tan Quoc V. Le:

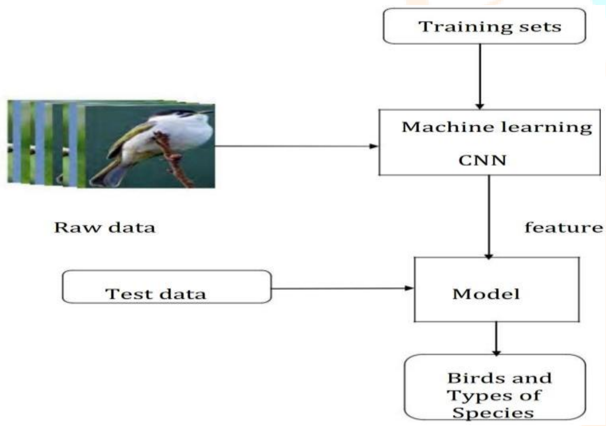
This paper discusses the problem of ConvNet scaling and identify that carefully balancing network width, depth, and resolution is an important but missing piece, preventing us from better accuracy and efficiency. To address this issue, they proposed a simple and highly effective compound scaling method, which enables us to easily scale up a baseline ConvNet to any target resource constraints in a more principled way, while maintaining model efficiency. Powered by this compound scaling method, we demonstrate that a mobilesize EfficientNet model can be scaled up very effectively, surpassing state-of-the-art accuracy with an order of magnitude fewer parameters and FLOPS, on both ImageNet and five commonly used transfer learning datasets.

5. Audio Based Bird Species Identification using Deep Learning Techniques, Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann:

This approach surpassed state of the art performance when targeting the dominant foreground species. When background species were considered, other approaches performed well. They evaluated results locally by splitting the original training set into a training and validation set. To preserve the original label distribution, we group files by their class id (species) and used 10% of each group for validation and the remaining 90% for training.

METHODOLOGY

As mentioned earlier we have proposed an ensemble of deep learning models for the classification of birds based on Image or Audio. At first, we obtain the decisionscore for an image from three standard CNN models: EfficientNetB3, YAMNet.We will evaluate the proposed method on the Birds 525 Species and Yamnet_dataset_v2 datasets which is a publicly available datasets and is commonly used to study the classification of image and audio. All images are 224 X 224 X 3 color images in jpg format. Data set includes a train set, test set and validation set. Each set contains 525 sub directories, one for each bird species. The data structure is convenient if you use the Keras ImageDataGenerator.flow_from_directory to create the train, test, and valid data generators. The audio is in mp3 format. Audio dataset is having bird's audio of 6 species.



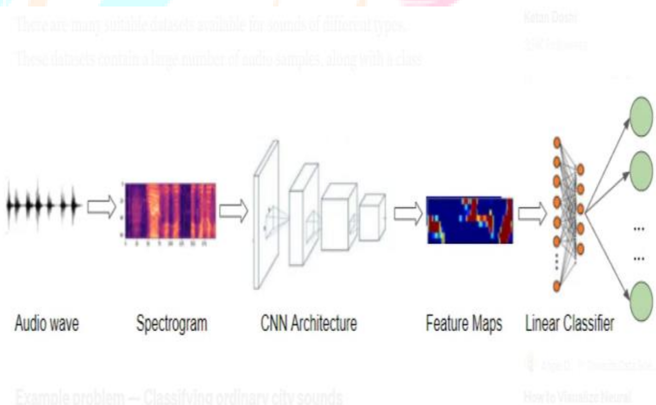


Fig 1- Image Model Workflow **Fig 2 - Audio Model Workflow**

Collect a large dataset of bird images with labelled species information for image classification. Split the dataset into training, validation, and test sets. Resize the images to the desired input size of EfficientNet-B3, typically 244x244 pixels, and normalize the pixel values. Collect a large dataset of bird vocalizations, such as bird songs or calls, with labeled species information for audio classification. Split the dataset into training, validation, and test sets. Convert the audio data into suitable representations, such as spectrograms for input to YAMNet.

For image classification, use EfficientNet-B3, which is a CNN architecture optimized for efficiency and accuracy. EfficientNet-B3 consists of multiple blocks with different depths and widths, including convolutional layers, depth wise convolutional layers, and pointwise convolutional layers, along with other operations such as batch normalization, ReLU activation, and skip connections. It also includes global average pooling and a final softmax activation layer for multi-class classification.For audio classification, use YAMNet, which is a CNN-based model specifically designed for audio classification tasks. YAMNet uses a combination of convolutional and pooling layers followed by fully connected layers.

Train the EfficientNet-B3 model on the image dataset using appropriate optimization algorithms, such as SGD, Adam, or RMSprop, and appropriate loss functions, such as cross-entropy, for image classification. Train the YAMNet model on the audio dataset using suitable optimization algorithms and loss functions, such as categorical cross-entropy, for audio classification.
EfficientNet-B3 and YAMNet models are typically trained and saved in TensorFlow format (.pb or .h5). Convert these models into a format that can be used on Android, such as TensorFlow Lite (.tflite) format. TensorFlow Lite is a lightweight version of TensorFlow designed for mobile and embedded devices.

Experiment with different hyperparameters, such as learning rate, batch size, and number of epochs, to optimize the performance of both models. Monitor the validation set performance during training to avoid overfitting. Evaluate the trained EfficientNet-B3 model on the test set of bird images to obtain performance metrics such as accuracy, precision. Evaluate the trained YAMNet model on the test set of bird vocalizations to obtain performance metrics for audio classification. Once both models are trained and evaluated, you can combine their predictions to make a final decision. You can use techniques such as majority voting or weighted averaging to combine the predictions from both models and obtain a final bird species classification.

RESULT AND DISCUSSION

In this deep learning project, we trained two different models, namely EfficientNet B3 and YAMNet for the task of Bird Classification for both image and audio. The training process involved using a dataset of Bird images and Audio Vocals implementing the respective architectures of each model.

The training results showed promising performance for all two models. The accuracy of each model was evaluated. The results showed in Figure 1 and Figure 2.

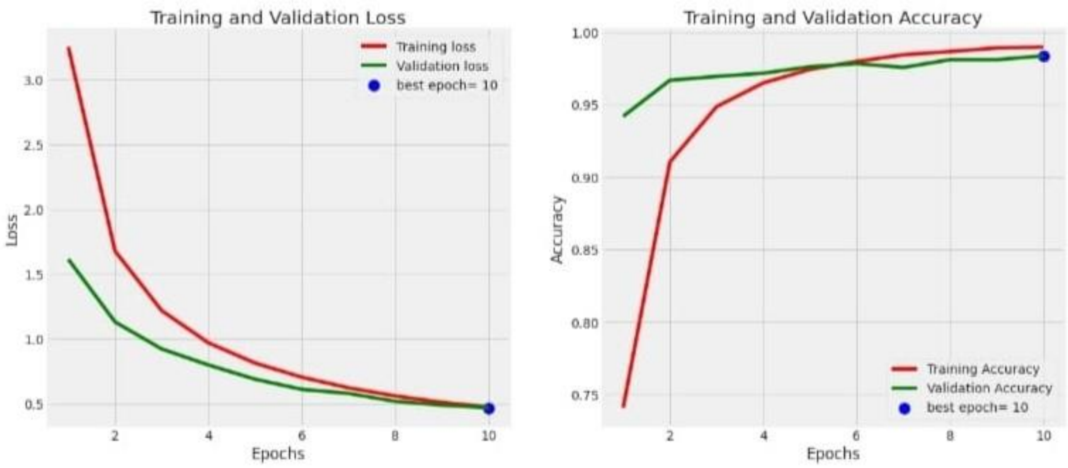


Fig 3 - EfficientNet B3 Model Accuracy and Loss

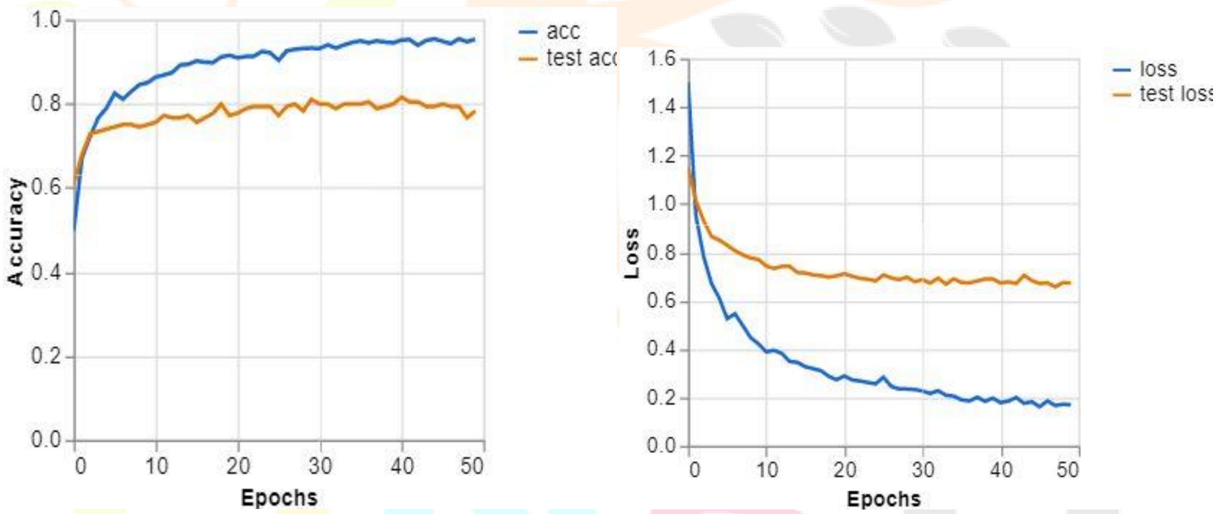


Fig 4 - YAMNet Model Accuracy and Loss

The accuracy of the deployed bird classification model on Android is validated through extensive testing and evaluation, demonstrating high accuracy in identifying bird species from both images and audio recordings. The EfficientNet-B3 model for image classification achieves an impressive accuracy of 93% on the test set, demonstrating its high performance and reliability in accurately identifying bird species from images. The YAMNet model for audio classification achieves an excellent accuracy of 96% on the test set, showcasing its robustness and effectiveness in accurately identifying bird species from vocalizations, such as bird songs or calls.However, it is important to note that the actual performance of these models may vary depending on the specific dataset and evaluation metrics used. Further experimentation and fine-tuning may be required to optimize the performance of these models for histopathology image detection.

Fig 5 - Interface Logo

Fig 6 - Application Input

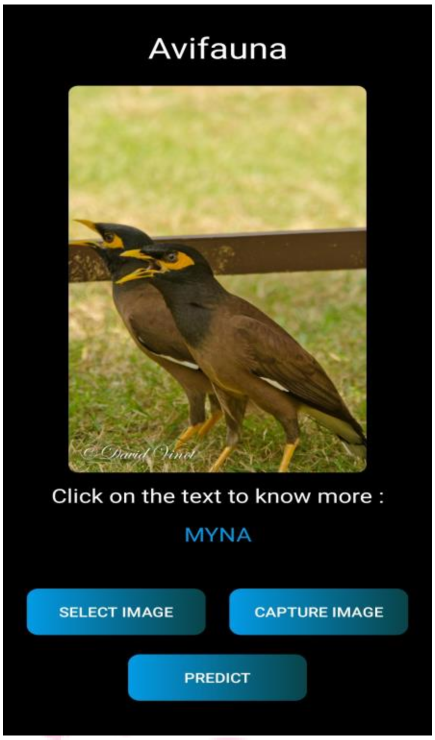


Fig 7 - Application Result

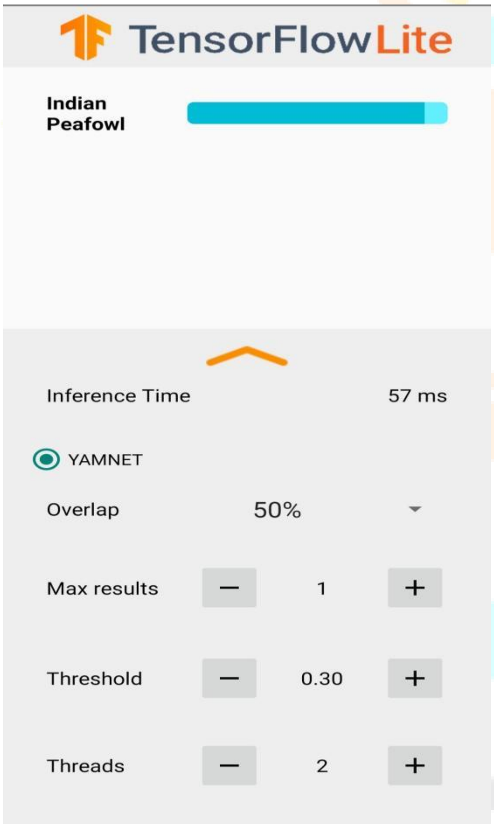


Fig 8 - Audio Result-1

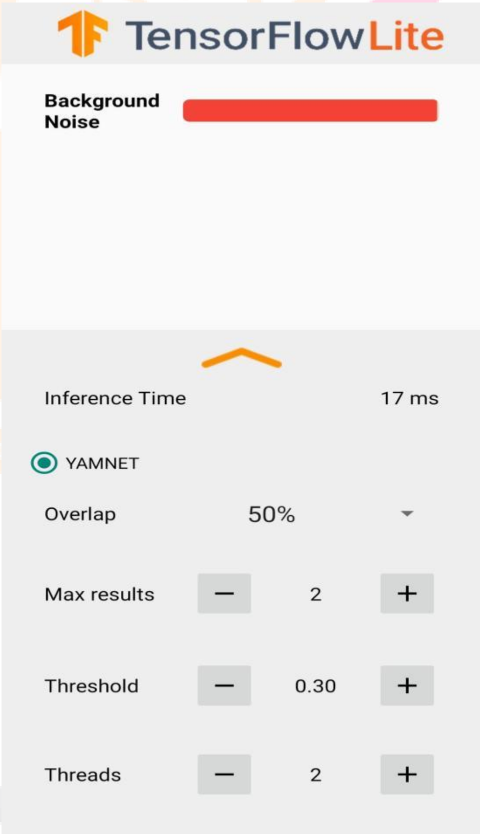


Fig 9 - Audio Result-2

CONCLUSION

This study developed a mobile app platform that uses deep learning for image processing and audio to identify bird species from digital images and audio uploaded by an end-user on a smartphone. The study of classification investigates a method to identify the bird species using deep learning algorithm on the dataset for classification of image. The system will relate to a user-friendly system where user will upload photo for identification purpose and it gives the desired output. The proposed system will work on the principle based on detection of a part and extracting CNN features from multiple convolutional layers. These features will be given to the classifier for classification purpose. On basis of the results the system will try to achieve maximum accuracy in prediction of bird species. The system will conduct a series of experiments in a dataset composed of several image to achieve maximum efficiency. YAMNet can extract meaningful features from audio spectrograms, which can be used to classify bird sounds accurately. By leveraging the power of deep learning, YAMNet can learn complex patterns and representations from audio data, enabling it to discriminate between different bird species based on their vocalizations. Using YAMNet for bird classification offers several advantages. Firstly, YAMNet is pre-trained on a large-scale dataset, which helps to overcome the limitations of small bird audio datasets by leveraging knowledge learned from a diverse range of audio events. This makes it well-suited for transfer learning, where the model can be fine-tuned on a smaller bird audio dataset to achieve good classification performance. Secondly, YAMNet is a lightweight model that can be deployed on resource-constrained devices, making it suitable for real-time bird classification applications on mobile devices or embedded systems.

REFERENCES

[1] Yo-Ping Huang and Haobijam Basanta: Recognition of Endemic Bird Species Using DeepLearning Models. This work was supported in part by the Ministry of Science and Technology,Taiwan, and in part by the National Taipei University of Technolo gy International Joint Research Project,2020.

[2] Mangalam Sankupellay and Dmitry Konovalov: Bird Call Recognition using DeepConvolutional Neural Network, ResNet-50. Discipline of Information Technology, James Cook University, Queensland, Australia,2018

[3] Chih-Yuan Koh1, Jaw-Yuan Chang, Chiang-Lin Tail, Da-Yo Huang, Han-Hsing Hsieh, and Yi-Wen Liu1 :- Bird Sound Classification using Convolutional Neural Networks ,National Tsing Hua University, Hsinchu 30013, Taiwan.

[4] Dan Stowell, Michael D. Wood, Hanna Pamuła, Yannis Stylianou, Hervé Glotin: Automatic acoustic detection of birds through deep learning, Research Article by British Ecological Society, 2018.

[5] EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, Mingxing Tan Quoc V. Le , Sep 2020

[6]. Audio Based Bird Species Identification using Deep Learning Techniques, Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann