

Kathmandu University

Department of Computer Science and Engineering   
 Dhulikhel, Kavre

A Project Report   
 on

“Birds Audio Classification Using HRNET Model”

[Code No: AISP-121]

(For partial fulfillment of I Year/II Semester in B.Tech in Artificial Intelligence)

Submitted by,

Shova Gelal (Roll No. 08)   
 Bhawana Ojha (Roll No. 21) Baarosh Manandhar (Roll No. 21)

Submitted to,

(Dr. Yagya Raj Pandeya)

Department of Computer Science and Engineering

Submission Date:

20 June, 2024

Bonafide Certificate

This project work on

“Birds Audio Classification Using HRNET Model”

is a Bonafide work of

“Shova Gelal, Bhawana Ojha, and Baarosh Manandhar”   
 who carried out the project under my supervision.

Project Supervisor

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Name: Umesh Hengaju

Academic Designation: Assistant Lecturer

1

Acknowledgement

We want to thank our supervisor, Sir, Umesh Hengaju for his constant help and

guidance during the project. He supported us all the way, giving us new ways to make our project successful. His role as our supervisor was super important in making sure we had what we needed to finish the project well. We are grateful for all his support. Working under his guidance was an amazing experience.

We extend our heartfelt gratitude to the Department of Artificial Intelligence (DoAI), School of Engineering (SOE), Kathmandu University (KU), for providing us with the opportunity to embark on this intellectually enriching journey, blending academic theory with practical application.

We owe special thanks to Assistant Professor Dr. Yagya Raj Pandey Sir, whose guidance, insights, and encouragement have been invaluable throughout the conceptualization and planning of this project.

2

Abstract

This project is focused on creating an audio classification system using the HRNet model, to classify bird species based on their audio signals. The audio files were converted into mel spectrograms and mel phasegrams, which were then fed into the HRNet model for classification. By utilizing advanced training techniques such as early stopping and learning rate scheduling, we were able to improve the accuracy of our model. Our dataset comprises thousands of audio files representing different bird species, and we evaluated our system using accuracy, loss, confusion matrices, and ROC curves. The HRNet model demonstrated high accuracy and robustness in audio classification tasks, making it a promising candidate for broader audio-based applications.

The main objective of this project is to develop a birds species audio classification model for birds species conservation that enables to balance the ecosystem.

Keywords:

Artificial Intelligence ,HRNet, mel spectrograms ,Birds Species Audio Classification,

3

Table of Contents

Acknowledgement 2

Abstract 3

List of Figures 5

Acronyms/Abbreviations 6

Chapter 1 Introduction 7

1.1 Background 7

1.2 Objectives 7

1.3 Motivation and Significance 8

Chapter 2 Related Works/ Existing Works 9

Chapter 3 Design and Implementation 10

3.1 Procedures and Methods 10

Chapter 4 System Requirements and Specifications 15

4.1. Software Specifications 15

4.2 Hardware Specifications 15

Chapter 5 Discussion on the achievements 16

Chapter 6 Conclusion and Recommendation 20

References 21

4

List of Figures

Figure 1 Workflow

Figure 2 Mel Spectrogram Example 10

Figure 2 HRNet Model Architecture 11

Figure 3 Confusion Matrix of Model performance 12

Figure 4 ROC Curve For Birds Audio Classification 13

5

Acronyms/Abbreviations

HRNet High-Resolution Network

LR Learning Rate

GPU Graphics Processing Uint

CNN Convolutional Neural Network

AI Artificial Intelligence

ROC Reciever Operating Curve

6

Chapter 1 Introduction

In this chapter we have introduced the techniques used for birds species audio classification using HRNET model. It further elaborates the abstract providing a more detailed explanation on project’s background, objectives,motivation,expected outcomes, and significance of the work.. 2

1.1 Background

Birds plays a crucial role in the nature ecosystem as it directly affects the human health , food production, ecology balance and so on. Audio classification, while having many beneficial uses, also poses challenges due to its complexity, particularly in noisy environments where sounds from various sources overlap. Our project addresses the specific need for an efficient and accurate system capable of recognizing bird species from their calls. Traditional methods of manual bird identification are often tedious and prone to error, highlighting the need for an automated solution. By leveraging the capabilities of advanced machine learning models, such as HRNet, this project aims to simplify and enhance the process of identifying bird species through their audio signatures.

The proposed system will use deep learning algorithms to process audio data and extract relevant features, converting the audio into spectrograms that reveal the distinctive characteristics of each bird call. These visual representations of sound are then fed into the HRNet model, which classifies the species. This project contributes to the growing field of AI-based audio analysis, offering a tool that not only streamlines the identification process but also supports efforts in wildlife conservation and ecological studies..

1.2 Objectives

● To classify bird species based on their audio recordings using deep learning techniques.

● To utilize mel spectrogram and mel phasegram features for robust audio classification.

● To evaluate the effectiveness of HRNet architecture for audio classification tasks.

7

1.3 Motivation and Significance

The ability to classify bird species through their audio signals holds great potential for advancing wildlife research and conservation efforts. With the increasing availability of technology, there is a growing demand for automated solutions that can assist in identifying bird species, particularly in remote or ecologically sensitive areas where manual observation is challenging. Audio classification systems provide an efficient way to monitor biodiversity and track changes in bird populations, supporting conservationists in preserving habitats.

However, the misuse of technology in the field of audio classification can also raise ethical concerns, such as misidentifying species or generating incorrect data that could mislead conservation efforts. The need for accurate and reliable systems is therefore paramount, particularly when such systems are used in critical ecological studies or by government agencies for decision-making. By developing a precise classification system using advanced deep learning models like HRNet, we aim to overcome these challenges and contribute positively to the field of environmental monitoring.

This project will empower researchers, conservationists, and enthusiasts by offering a tool that can accurately and efficiently classify bird species through audio recordings. It will also enhance our understanding of bird populations and their behaviors, providing critical data for biodiversity studies. This technology's impact can extend beyond research, contributing to the protection of ecosystems and raising awareness about the importance of preserving bird species.

.

8

Chapter 2 Related Works/ Existing Works

In recent times, significant advancements have been made in using computational techniques to classify and recognize bird sounds, enabling more efficient identification of bird species. These advancements enhance the study of birds, allowing researchers to automate identification processes and gain deeper insights into bird behavior, biodiversity, and ecological trends. Such innovations hold potential for furthering research in ornithology and environmental science, with improved accuracy and reduced manual efforts.

“"Investigation of Different CNN-Based Models for Improved Bird Sound Classification," in *IEEE Access*, vol. 7, pp. 175353-175361, 2019, doi:” by J. Xie, K. Hu, M. Zhu, J. Yu and Q. Zhu investigate different CNN based models for improved sound classification. A key function of automatic classification of bird sounds is the observation and preservation of biodiversity. A new method for continually monitoring birds is made possible by recent developments in deep learning algorithms and acoustic sensor networks. Several deep learning-based classification frameworks for the identification and categorization of birds have been developed in earlier research. In order to further enhance bird sound classification ability, we compare various classification models in this work and selectively fuse them. To be more precise, we use two distinct deep learning architectures to build the fused model in addition to using the same deep learning architecture with various inputs. In order to define the various acoustic components of birds, three methods of time-frequency representations (TFRs) of bird sounds are studied: mel-spectrograms, harmonic-component based spectrograms, and percussive-component based spectrograms.

“Bird Sounds Classification Using Linear Discriminant Analysis” by M. Ramashini, P. E. Abas, U. Grafe and L. C. De Silva. In this study, bird species from the Borneo region are classified and identified based on their sounds. The researchers focused on five local bird species and applied audio signal processing techniques to extract 35 features from bird sounds. These features were reduced using Linear Discriminant Analysis (LDA) and classified using the Nearest Centroid (NC) method. The proposed method achieved a prediction accuracy of 96%, outperforming more complex algorithms such as Support Vector Machines (SVM) and K-Nearest Neighbor (KNN). This research demonstrates an efficient approach to bird sound classification in dense jungle environments.

“Active Learning for Bird Sounds Classification” by S. Hirzel Verlag · EAA. This research highlights the use of bird sound recognition as a valuable tool for ornithologists and ecologists to study behaviors such as mating, evolution, biodiversity, and the effects of climate change. The study introduces active learning techniques to reduce the need for extensive human-labeled data, which is often costly and time-consuming. The proposed sparse-instance-based and least-confidence-score-based active learning methods reduce the required human annotations by 16.0% and 35.2% respectively, compared to passive learning methods. These techniques maintain an acceptable performance with an unweighted average recall greater than 85% when recognizing sounds from 60 bird species

9

Chapter 3 Design and Implementation

3.1 Procedures and Methods

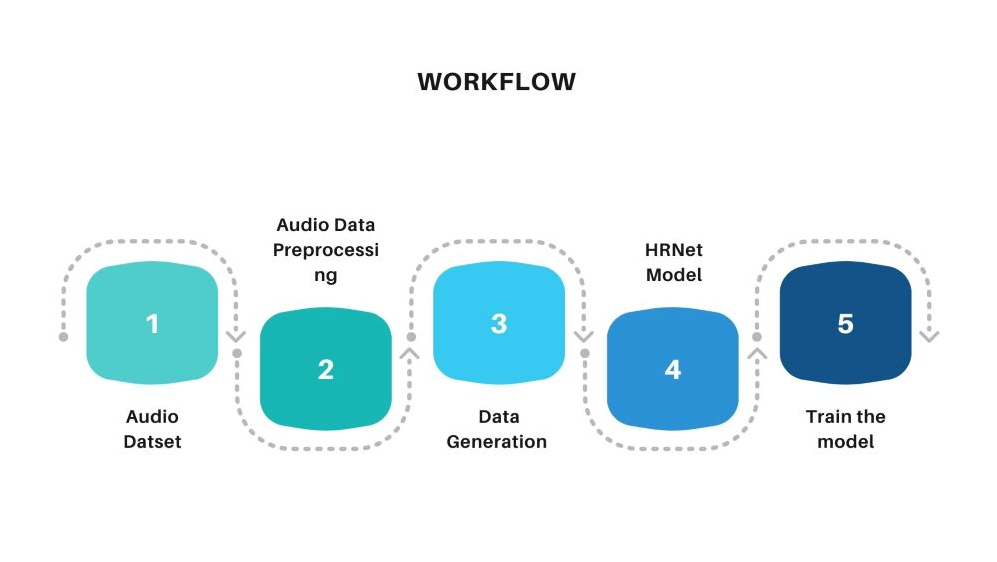


Figure 1 Workflow

Data Collection

For this project, we collected a diverse dataset from popular dataset sources like

Kaggle. Given the variation in each dataset, we focus our initial efforts on collecting different datasets and merging them to create a strong and balanced dataset for our further development.

Our final Dataset has two classes labelled as Real and Forged, each class containing   
32339 and 31160 respectively. The dataset was further split into training and   
validation.

Audio Data Processing

The image data was processed before sending to the network. The preprocessing

includes different augmentation approaches to ensure compatibility, consistency and quality of the data.

Data Generation

Three different CNN models were built for training to extract the image features and further used for detection. The models are ResNet50, VGG16, and MesoInception4. The detailed description of the models is given below:

10

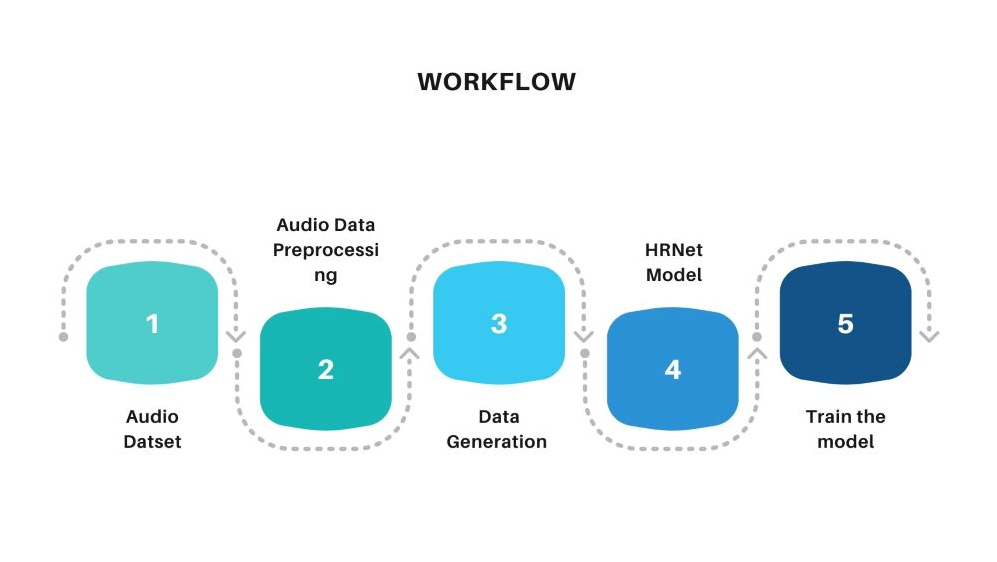


Figure 2

2 Model Building

● The pre-trained VGG16 model is loaded without its top layers   
 (include\_top=False) to use it as a feature extractor. This model is trained on   
 the ImageNet dataset and includes powerful convolutional layers for feature   
 extraction.

● All layers in the VGG16 model are frozen to prevent their weights from being   
 updated during training. This ensures that the pre-trained features are retained. ● A new model is constructed on top of the VGG16 base adding Dense layers and   
 pooling layers.

● A custom classification head is added on top of the VGG16 base. This includes   
 dense layer with 128 units and ReLU activation. And, a final dense layer with

1 unit and sigmoid activation for binary classification.

3 Model compilation and training

● Compile the model using the Adam optimizer and binary cross-entropy loss.   
● Train the model using the ‘fit’ method on training and validation generators.   
● Implements a ‘ModelCheckPoint’ callback to save the best weights based on

validation loss.

11

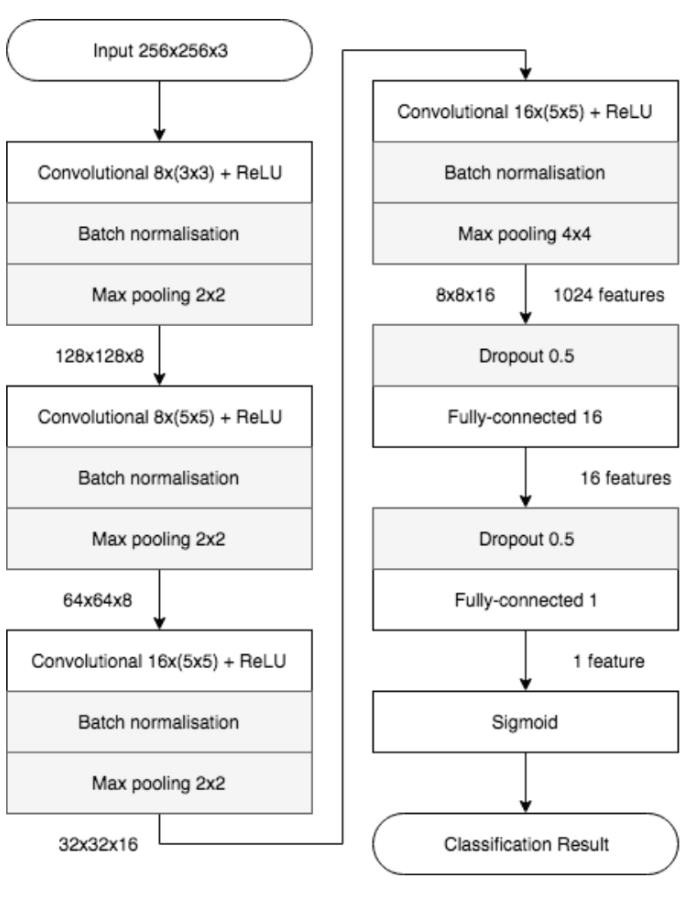


Figure 3

1 Data Preparation

● It uses an “ImageDataGenerator” for data augmentation and preprocessing.

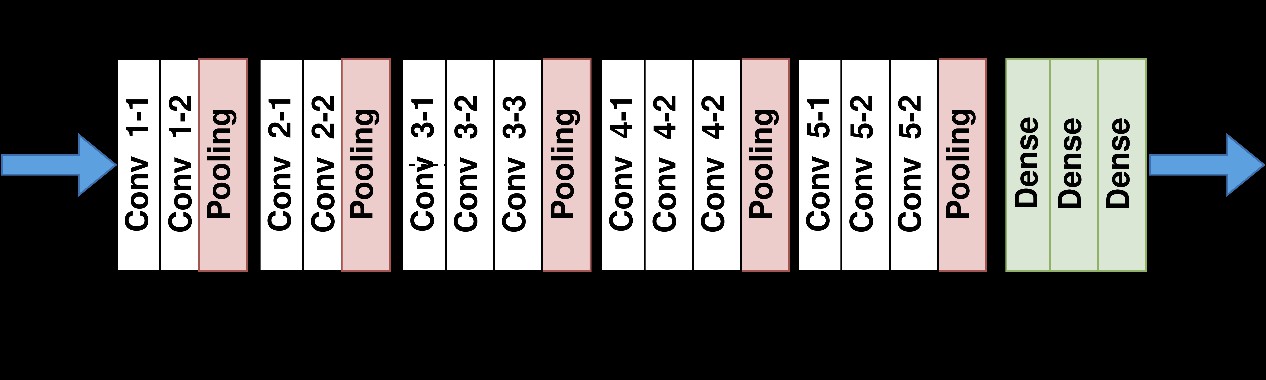
● The dataset is split into training and validation sets and loaded from the   
 specified paths.

2 Model Building

● A model is constructed based on the given architecture.

● The final dense layer has 1 node for output with Sigmoid activation.

12



3 Save the model in .h5 format.

● The model was saved in the Hierarchical Data Format 5 file that is further   
 used for prediction.

13

14

Chapter 4 System Requirements and Specifications

4.1. Software Specifications

To ensure the smooth execution of our project, we used a wide range of software tools and libraries. Below are detailed software specifications.

Programming languages

● Python 3.9

Python is the primary language used for data preprocessing, model   
implementation, and training due to its simplicity and extensive library support.

Libraries and frameworks

● OpenCV

OpenCV is a highly popular open-source library that provides a comprehensive set of tools and functions for computer vision tasks. It was used for video processing tasks, such as frame extraction, resizing etc.

● TensorFlow or PyTorch

Used for building and training deep learning models (ResNet50, VGG16, MesoInception4).

● Streamlit 1.35

A web application framework used for building the detection interface.

● IDEs

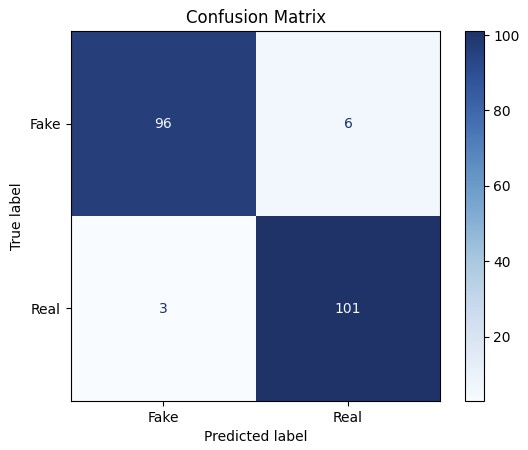
VS code, Jupyter Notebook, Google Colab

4.2 Hardware Specifications

● GPU (GTX 1660Ti 6GB) ● RAM (16 GB)

● CPU (core i5 or equivalent)

15



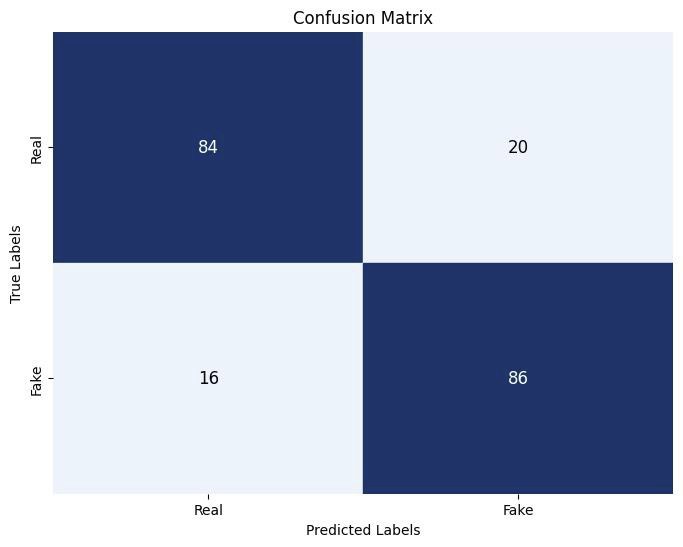
Chapter 5 Discussion on the achievements

Evaluation Metrics

The model is evaluated on a distinct test dataset and the results are below:

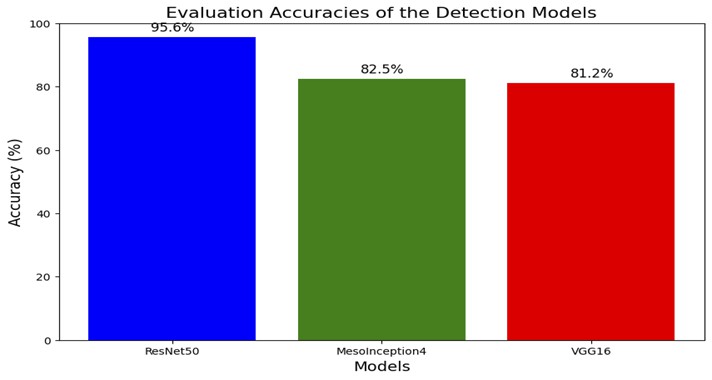
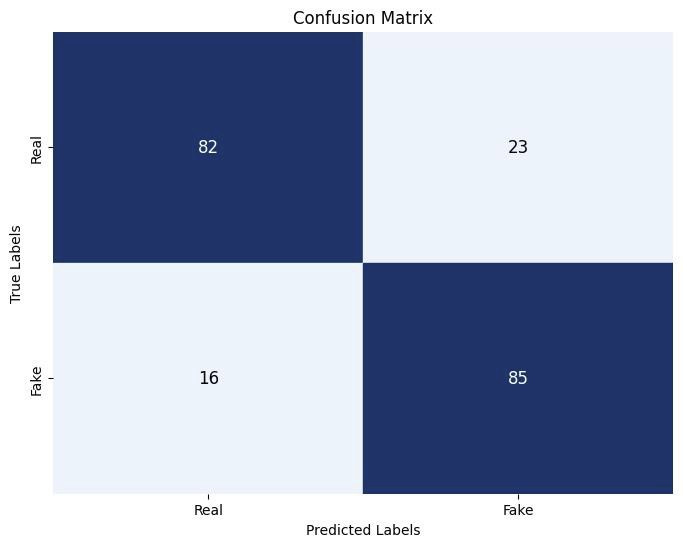
For ResNet50

16



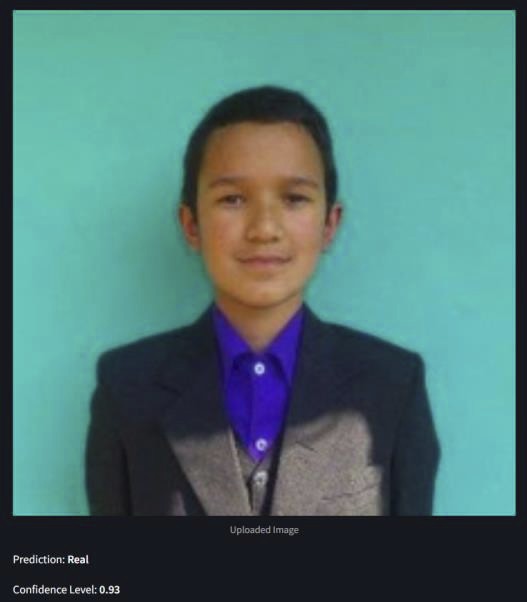
For MesoInception4

17



For VGG16

18



Some outcomes of the Project

19

Chapter 6 Conclusion and Recommendation

Conclusion

In this project, we developed a HRNet model capable of classifying the birds species based on the features extracted from audio using mel spectrograms.

While our approach showed significant improvements, challenges remain such as

model complexity and computational demands. Future works will focus on expanding the dataset, selecting the next best feasible approaches for better end results, and   
exploring additional applications.

Limitations

Despite the strengths and approaches utilized in our project, there are several

limitations to our model that should be acknowledged. These limitations provide insights into potential areas of improvement and future research directions. The limitations include limited dataset diversity, complexity and computational   
requirements, and more.

Future Enhancements

To further improve the performance and applicability of our project, several

enhancements can be considered. These enhancements aim to address the current   
limitations, explore new opportunities for advancing the research, and extend the   
utility of our project. Exploration of other model architectures can potentially enhance the model ability to make better results. Other potential work can be expanded for more different audio classification domains.

20

References

“"Investigation of Different CNN-Based Models for Improved Bird Sound Classification," in *IEEE Access*, vol. 7, pp. 175353-175361, 2019, doi:” by J. Xie, K. Hu, M. Zhu, J. Yu and Q. Zhu <https://arxiv.org/abs/1809.00888>

“Bird Sounds Classification Using Linear Discriminant Analysis” by M. Ramashini, P. E. Abas, U. Grafe and L. C. De Silva [arxiv.org/abs/2303.15823](http://arxiv.org/abs/2303.15823)

“Active Learning for Bird Sounds Classification” by S. Hirzel Verlag · EAA - <https://www.warse.org/IJATCSE/static/pdf/file/ijatcse62922020.pdf>

21