VISVESVARAYA TECHNOLOGICAL UNIVERSITY
Jnana Sangama, Belagavi - 590018

Project Report

on

"BIRD CLASSIFICATION BASED ON IMAGE OR AUDIO USING DEEP LEARNING"

Submitted in partial fulfillment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING
in

COMPUTER SCIENCE & ENGINEERING

by

MAANIKYA	**4MT19CS073**
POOJARY DHEERAJ KUMAR	**4MT19CS106**
RACHANA NAYAK	**4MT19CS117**
RAKSHIT LINGAPPA POOJARI	**4MT19CS122**

Under the Guidance of

Mr. Shreejith K B
Senior Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(*Accredited by NBA*)

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING
Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution
(A Unit of Rajalaxmi Education Trust®, Mangalore - 575001)
Affiliated to VTU, Belagavi, Approved by AICTE, New Delhi
Badaga Mijar, Moodabidri-574225, Karnataka

2022-23

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution
(A Unit of Rajalaxmi Education Trust®, Mangalore - 575001)
Affiliated to VTU, Belagavi, Approved by AICTE, New Delhi
Badaga Mijar, Moodabidri-574225, Karnataka

MITE
Invent Solutions

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(*Accredited by NBA*)

CERTIFICATE

This is to certify that the project work entitled **"BIRD CLASSIFICATION BASED ON IMAGE OR AUDIO USING DEEP LEARNING"** is a bonafide work carried out by **Maanikya (4MT19CS073), Poojary Dheeraj Kumar (4MT19CS106), Rachana Nayak (4MT19CS117), Rakshit Lingappa Poojari (4MT19CS122)** in partial fulfillment for the award of degree of **Bachelor of Engineering in Computer Science & Engineering of the Visvesvaraya Technological University, Belagavi during the year 2022 – 23**. It is certified that all corrections and suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project has been approved as it satisfies the academic requirements in respect of project work prescribed for the Bachelor of Engineering degree.

Mr. Shreejith K B	**Dr. Ravinarayana B.**	**Dr. Prashanth C M**
Project Guide	**Head of the Department**	**Principal**

External Viva

Name of the Examiners **Signature with Date**

1)

2)

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

Accredited by NAAC with A+ Grade, An ISO 9001: 2015 Certified Institution

(A Unit of Rajalaxmi Education Trust)

Affiliated to Visvesvaraya Technological University, Belagavi

Badaga Mijar, Moodabidri – 574 225 , Karnataka

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

(Accredited by NBA)

DECLARATION

We, **Maanikya**(4MT19CS073), **Poojary Dheeraj Kumar**(4MT19CS106), **Rachana Nayak**(4MT19CS117) and **Rakshit Lingappa Poojari**(4MT19CS122) students of 8^{th} semester BE in Computer Science & Engineering, **Mangalore Institute of Technology and Engineering, Moodabidri,** hereby declare that the project work entitled "**Bird Classification based on Image or Audio using Deep Learning",** submitted to the **Visvesvaraya Technological University, Belagavi** during the academic year **2022-23**, is a record of an original work done by us under the guidance of **Mr. Shreejith K B,** Senior Assistant Professor, Department of Computer Science & Engineering , Mangalore Institute of Technology and Engineering Moodabidri. This project work is submitted in partial fulfillment of the requirements for the award of the degree of Bachelor of Engineering in Computer Science & Engineering. The results embodied in this report have not been submitted to any other University or Institute for the award of any degree.

Date:

Place: Moodbidri

Maanikya

Poojary Dheeraj Kumar

Rachana Nayak

Rakshit Lingappa Poojari

ACKNOWLEDGEMENT

The satisfaction and the successful completion of this project would be incomplete without the mention of the people who made it possible, whose constant guidance encouragement crowned our efforts with success.

This project is made under the guidance of **Mr. Shreejith K B, Senior Assistant Professor**, in the Department of Computer Science and Engineering. We would like to express my sincere gratitude to our guide for all the helping hand and guidance in this project.

We would like to thank our project coordinator **Mr. Annappa Swamy D R, Associate Professor,** in the Department of Computer Science and Engineering, for his cordial support, valuable information and guidance which helped us in completing this project through the various stages.

We would like to express appreciation to **Dr. Ravinarayana B., Associate Professor and Head**, Department of Computer Science and Engineering, for his support and guidance.

We would like to thank our Principal **Dr. Prashanth C M**, for encouraging us and giving us an opportunity to accomplish the project.

We also thank our management who helped us directly or indirectly in the completion of this project.

Our special thanks to faculty members and others for their constant help and support.

Above all, we extend our sincere gratitude to our parents and friends for their constant encouragement with moral support.

MAANIKYA
POOJARY DHEERAJ KUMAR
RACHANA NAYAK
RAKSHIT LINGAPPA POOJARI

ABSTRACT

Birds are a vital group of animals that ecologists monitor using autonomous recording units as a crucial indicator ofthe 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 classifyingbirds into different species with high accuracy based on deep learning techniques using either images or audio data.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

CNN	Convolutional Neural Network
YAMNet	Yet another Audio Mobilenet Network
SQA	Software Quality Assurance
UML	Unified Modeling Language
XML	eXtensible Markup Language
IDE	Integrated Development Environment
RAM	Random Access Memory
GPU	Graphics Processing Unit
CPU	Central Processing Unit
SVM	Support Vector Machine
DNN	Deep Neural Network

CHAPTER – 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Introduction

In this chapter, we are going to give the introduction about the Bird classification based on Image or Audio using Deep Learning.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 difference 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.

1.2 Problem Statement

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. Birds are also susceptible to changes in the environment. Therefore, monitoring birds/bird calls in an environment provide vital information about changes in the environment itself. Even a slight decrease in their number endangers the entire ecosystem. They bring plants back to ecosystems through pollination or seed dispersal across the sea to new land masses. Many birds are scavengers and they help in the quick disposal of carcasses and in the recycling of nutrients in the ecosystem to maintain a healthy habitat.

The problem statement for bird classification can be defined as follows:

- Manual bird species identification is time-consuming and error-prone.
- Limited availability of expert knowledge and resources.
- Increasing need for large-scale bird monitoring and conservation.
- Lack of standardized bird classification protocols.
- Advancements in technology and data availability.
- Engaging citizen scientists and the general public in bird identification.

1.3 Objectives

- Develop accurate and robust deep learning models for bird classification.
- Easy identification of bird species.
- Enhanced identification technique using audio clip recorded from the bird at the scene.
- Bird images or audio recordings can vary greatly due to factors such as lighting conditions, background noise, and variations in bird vocalizations.
- Improve the identification accuracy of our model by using Image upscaling.
- Improve ease of recognition & spotting by implementing image and audio as inputs.

1.4 Scope

Bird classification using deep learning has a significant scope, both in terms of image and audio-based classification.

- **Rich and diverse data:** Birds are diverse in terms of their species, appearance, and vocalizations. Image and audio data of birds are readily available through online repositories, such as the Cornell Lab of Ornithology and Xeno-canto. This makes it possible to train and evaluate deep learning models on large-scale datasets, which can improve their accuracy and generalization.

- **Environmental monitoring:** Birds play a crucial role in ecosystem dynamics and biodiversity conservation. Monitoring bird populations and habitats can provide valuable insights into environmental health and conservation efforts. Deep learning models for bird classification can facilitate this monitoring by automating the detection and identification of bird species and their vocalizations.

- **Citizen science:** Citizen science initiatives, such as eBird, encourage bird enthusiasts to contribute to bird monitoring and research by reporting their sightings and recordings. Deep learning models for bird classification can augment these initiatives by automatically verifying and validating citizen-contributed data.

- **Education and outreach:** Bird identification and appreciation are popular among birdwatchers, nature enthusiasts, and students. Deep learning models for bird classification can enhance the learning and outreach opportunities by providing accurate and real-time feedback on bird identification and behaviour.

1.5 Organization of the Report

Chapter 1: Introduction: In this chapter, we are going to introduce Bird classification based on Image or Audio using Deep Learning Summarization starts by introducing the topic, which gives an insight and in-depth analysis of the project work, which was closely followed by the problem statement.

Chapter 2: Literature Survey: This section mainly contains the details of the study that is implemented with various methods.

Chapter 3: Software Requirement Specification: This section contains the details about the product perspective, user characteristics, its assumptions and dependencies, specific requirements, functionality along with resource requirements

Chapter 4: Gantt chart: A Gantt chart is a type of bar chart, developed by Henry Gantt that illustrates a project schedule. Gantt charts illustrate the start and finish of the terminal elements and summary elements of the project.

Chapter 5: System Design which has an Architectural diagram, Use case diagram, Sequence diagram, Activity Diagram, and Data Flow Diagram.

Chapter 6: Implementation: This section contains details about the steps and procedures in the development of the project work.

Chapter 7: System Testing: This chapter contains details about the various types of test cases.

Chapter 8: Results & Snapshots: This section contains the details about the overall outcomes of the project

Chapter 9: Conclusion and Future Work: This section describes the overall outcome of the entire project development and suggests some of the enhancement ideas which could not be covered up due to constraints of time and resources.

References: Enumerates references that are used from various sources for designing and developing this project.

CHAPTER – 2

LITERATURE SURVEY

CHAPTER 2

LITERATURE SURVEY

2.1 Proposed System

Recognition of Endemic Bird Species Using Deep Learning Models by YO-PING HUANG (Fellow, IEEE) AND HAOBIJAM BASANTA [1].

In this study, we 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. Our 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.

Automatic acoustic detection of birds through deep learning by Dan Stowell, Michael D. Wood, Hanna Pamuła, Yannis Stylianou, Hervé Glotin [4].

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.

Bird Call Recognition using Deep Convolutional Neural Network, ResNet-50 by Mangalam Sankupellay and Dmitry Konovalov [2].

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 deep- CNN. The ResNet-50 (a 50 layer deep-CNN architecture), is the first deep-CNN architecture that utilized residual learning.

The ResNet-50 model was trained on 1.28 million training images in 1000 classes and reaching an average of 5.25% of top-5 error.ResNet-50 training accuracy was about 65% when the input was 512 (height (frequency)) x 1024 (length (time)). When the input data is selected at 1024 length (time), in affect the input spectrogram to ResNet-50 is 5.94 sec long (1024 * 32 * 4 / 22050 Hz) call. ResNet-50 training accuracy improved to 72% when the selected input spectrogram was reduced to 512 (height (frequency)) x 512 (length (time)), effectively reducing the input bird call to 2.97 sec long calls. The improvement of ResNet-50 accuracy indicated the CNN learned shorter bird calls at a more accurate level.

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

This paper discuss 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 mobile size 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.

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

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.

CHAPTER – 3

SYSTEM REQUIREMENTS SPECIFICATION

CHAPTER 3

SYSTEM REQUIREMENTS AND SPECIFICATION

3.1 Overall Description

Bird classification based on image or audio using deep learning involves training a machine learning model to accurately identify and classify bird species based on images or audio recordings. The model is typically based on a deep neural network architecture that is trained on a large dataset of labelled bird images or audio samples.

For image-based classification, the model processes an image of a bird and extracts features that are used to classify the bird into a specific species. These features may include colour, texture, shape, and other visual characteristics.

For audio-based classification, the model processes an audio recording of bird vocalizations and extracts features that are used to classify the bird species based on its unique sound signature. These features may include spectral information, frequency patterns, and temporal characteristics of the bird's vocalizations.

The deep learning model is trained using a dataset of labelled bird images or audio samples. This dataset may be collected through various sources, such as online repositories, citizen science projects, or field recordings. The dataset is typically pre-processed to ensure high quality and relevance to the target task.

Once the model is trained, it can be used to classify new bird images or audio recordings into their respective species. This can have various applications, such as monitoring bird populations, studying bird behaviour, or enhancing birdwatching experiences.

3.1.1 Product Perspective

The product perspective of bird classification based on image or audio using deep learning can be viewed from multiple angles, including:

- **User interface:** The product should have a user-friendly interface that allows users to upload bird images or audio recordings and receive accurate species classification results. The interface should be designed to be accessible and intuitive for users with different levels of technical expertise.

- **Performance and accuracy:** The product should have high performance and accuracy in identifying bird species based on the input data. The model should be trained on a large and diverse dataset of labeled bird images or audio recordings to ensure robustness and generalization.

- **Speed and scalability:** The product should be designed to handle large volumes of input data and deliver fast classification results. The model should be optimized for speed and scalability to accommodate different use cases, such as real-time bird monitoring or batch processing of large datasets.

- **Integration and compatibility:** The product should be compatible with different hardware and software environments, including cloud-based platforms, mobile devices, and web browsers. The product should be designed to integrate with other tools and applications, such as bird identification apps, citizen science platforms, or environmental monitoring systems.

- **Privacy and security:** The product should be designed to protect user data and ensure privacy and security. The product should adhere to data protection regulations and implement secure data transmission and storage practices.

3.1.2 Product Functions

- **Data preprocessing:** The system should be able to preprocess the input data, including images or audio recordings, to remove noise, enhance quality, and standardize the format.

- **Feature extraction:** The system should be able to extract relevant features from the input data, such as color, texture, shape, or frequency, using techniques such as convolutional neural networks (CNNs) or spectrogram analysis.

- **Classification:** The system should be able to classify the input data into different bird species using machine learning algorithms, such as support vector machines (SVMs), random forests, or deep neural networks (DNNs).

- **Model training and optimization:** The system should be able to train and optimize the machine learning model using labeled data, such as images or audio recordings of different bird species.

- **Model evaluation and validation**: The system should be able to evaluate and validate the performance of the machine learning model using different metrics, such as accuracy, precision, recall, or F1 score.

- **User interface:** The system should provide a user-friendly interface for users to interact with the system, including uploading images or audio recordings, searching, and browsing bird species, and viewing classification results.

- **Real-time monitoring:** The system should be able to provide real-time monitoring and alerts for specific bird species or habitats, based on the input data and the classification results.

- **Integration with external systems:** The system should be able to integrate with external systems, such as GPS trackers, weather stations, or social media platforms, to enhance the accuracy and reliability of the classification results.

3.1.3 User Classes and Characteristics

- **Bird researchers and ornithologists:** These users are interested in using the classification system to study bird populations, migration patterns, and behavior. They may have expertise in bird identification and classification and require a high level of accuracy and reliability from the system.

- **Conservationists:** These users are interested in using the classification system to monitor bird populations and habitats and to identify potential threats to bird species. They may require real-time monitoring and alerts for specific bird species or habitats.

- **Bird enthusiasts:** These users may be interested in using the classification system to identify birds in their local area or during birdwatching activities. They may require a user-friendly interface and the ability to search and browse bird species based on different criteria.

- **Citizen scientists:** These users may contribute to the development of the classification system by providing labeled data or participating in data collection activities. They may require guidance and training on how to collect and label data effectively.

- **Data scientists and machine learning Engineers:** These users are responsible for designing, developing, and maintaining the classification system. They may require expertise in deep learning algorithms, data preprocessing, and model optimization.

- **IT administrators:** These users are responsible for deploying and maintaining the infrastructure for the classification system. They may require expertise in cloud computing, network security, and data management.

3.1.4 Design and Implementation Constraints

- **Hardware constraints:** The performance of the model can be limited by the hardware used for training and inference. Training deep learning models requires powerful GPUs and large amounts of memory. In addition, the size of the dataset and the complexity of the model can also require high storage capacity.

- **Data availability:** Availability and quality of data is a key constraint for any deep learning application. In the case of bird classification, the availability of labeled images and audio recordings for different bird species can be limited, especially for rare or endangered species.

- **Computational complexity:** The computational complexity of the model can impact its scalability and performance. Deep learning models with many layers and parameters require more computation time and memory, and can be harder to optimize and tune.

- **Algorithmic complexity:** The choice of algorithms and techniques used for feature extraction, dimensionality reduction, and classification can also impact the performance of the model. Some techniques may be more suitable for certain types of data or problems, and may require computation time.

- **Regulatory and ethical considerations:** The use of bird classification systems can have implications for ethical and regulatory compliance. For example, the use of audio recordings for bird classification can raise privacy concerns, and the accuracy of the model should be thoroughly validated before being deployed for conservation or research purposes.

- **Cost:** The cost of developing and deploying a bird classification system can be significant, including the cost of hardware, data collection and labeling, and personnel resources for training and validation.

3.1.5 Assumptions and Dependencies

- **Availability of high-quality and annotated training datasets:** The accuracy and performance of the deep learning model heavily depend on the quality and quantity of the training data. The assumption here is that high-quality annotated datasets are available to train the model.
- **Availability of computing resources:** Deep learning models require significant computing resources such as powerful CPUs or GPUs, large amounts of memory, and storage. The assumption is that the required resources are available to train and run the model.
- **Dependence on external libraries and frameworks:** The development of deep learning models often relies on external libraries and frameworks such as TensorFlow, Keras, or PyTorch. The assumption is that the required libraries and frameworks are available and compatible with the development environment.
- **Dependence on data preprocessing techniques:** The accuracy of the deep learning model depends on how well the input data is preprocessed before training. The assumption is that effective data preprocessing techniques are available to extract relevant features and reduce noise from the input data.
- **Dependence on hyperparameter tuning**: Deep learning models have many hyperparameters that need to be tuned to optimize the model's performance. The assumption is that effective hyperparameter tuning techniques are available to find the optimal values for the model.
- **Limitations of the model:** Deep learning models are not perfect and have limitations. For example, the model may struggle to classify birds in certain lighting conditions or when birds are in similar backgrounds. The assumption is that the limitations of the model are understood and accounted for in the design and implementation of the system

3.2 Specific Requirements

- **Image and audio data collection:** The system should be able to collect a large and diverse set of images and audio recordings of different bird species, from different regions and habitats, to ensure the accuracy and generalizability of the machine learning model.

- **Labeled data:** The system should have access to labeled data, i.e., images and audio recordings with corresponding bird species labels, to train and validate the machine learning model.

- **Data preprocessing and augmentation:** The system should be able to preprocess the input data, such as removing noise, resizing, or normalizing the images, or converting audio recordings to spectrograms. The system should also be able to augment the data by creating new images or audio recordings through techniques such as rotation, scaling, or adding noise.

- **Machine learning algorithms:** The system should be able to use different machine learning algorithms, such as convolutional neural networks (CNNs),

- **Model training and optimization:** The system should be able to train and optimize the machine learning model using different techniques, such as transfer learning, hyperparameter tuning, or ensemble methods.

- **Model evaluation and validation**: The system should be able to evaluate and validate the performance of the machine learning model using different metrics, such as accuracy, precision, recall, or F1 score.

3.2.1 Hardware Requirements

- CPU : 10th Generation Intel® Core™ i5 Processor
- RAM : 8 GB or above
- Hard Disk : 500 GB or above
- GPU : NVIDIA GTX 1650

3.2.2 Software Requirements

- Operating System : Windows 10
- IDE : Visual Studio Code
- Language : Python 3.7–3.10, Java
- Backend Libraries : Keras, TensorFlow
- Frontend : Java, XML
- Software : Jupyter, Android Studio

Python

Python is a popular programming language for deep learning due to its ease of use and a wide range of libraries and frameworks that support deep learning tasks. With libraries such as TensorFlow, PyTorch, Keras, Theano, and Caffe, developers can easily define, optimize, and evaluate mathematical expressions involving multi-dimensional arrays and conduct a range of deep learning tasks, including image classification, segmentation, and object detection. Python's simplicity and flexibility make it an ideal language for both beginners and experienced developers who want to work on deep learning projects.

TensorFlow

TensorFlow is an open-source software library for dataflow and differentiable programming that is widely used in deep learning and machine learning applications. Developed by Google, TensorFlow provides a comprehensive ecosystem of tools, libraries, and community resources that make it easy for developers to build and train machine learning models at scale. TensorFlow supports a range of platforms, including CPUs, GPUs, and TPUs, and is known for its flexibility, scalability, and performance. It provides a high-level API, Keras, that makes it easy for beginners to get started with building deep learning models, while also allowing advanced users to customize models and experiment with new techniques. TensorFlow's rich ecosystem and robust features make it a popular choice for developers working on deep learning projects in both academia and industry.

Keras

Keras is a high-level neural networks API written in Python and designed to make deep learning more accessible and easier to use. It is built on top of lower-level libraries such as TensorFlow, Theano, and CNTK, and provides a user-friendly interface that allows

developers to build and train deep learning models quickly and efficiently. Keras offers a modular and flexible architecture that enables developers to create complex models by combining building blocks called layers. It also includes a range of pre-built layers, including convolutional, recurrent, and pooling layers, which can be easily combined to create various types of models for tasks such as image classification, object detection, and natural language processing. Keras is known for its simplicity, ease of use, and high-level abstraction, making it a popular choice for beginners and experienced developers who want to build deep learning models without getting bogged down in the details of lower-level libraries.

Java

Java is the primary language used for developing Android applications. With its vast community and robust libraries, Java provides a powerful and flexible platform for building high-quality, user-friendly, and scalable Android applications. Java's object-oriented programming model and garbage collection features make it easy for developers to write complex applications while minimizing memory management and other low-level programming details. Additionally, the Android Software Development Kit (SDK) provides a comprehensive set of tools and libraries that allow developers to easily create, test, and deploy applications for various Android devices. Whether you are building a simple app or a complex enterprise-level application, Java provides the tools and resources you need to create high-quality Android applications that meet the needs of your users.

XML

XML, or eXtensible Markup Language, is a widely used data format in Android development. XML is used to define user interfaces, layouts, and other resources in Android applications. With XML, developers can create layouts that are easily scalable and provide a consistent user experience across various screen sizes and resolutions. Android uses a special flavor of XML called Android XML to define layout resources, which are then compiled into a binary format that can be efficiently loaded and displayed by the application. XML is also used for storing and retrieving data, such as preferences, configurations, and user data. The flexibility and extensibility of XML make it an ideal format for developers to define and manage data and resources in Android applications.

Visual Studio code

Visual Studio Code (often abbreviated as VS Code) is a free and open-source code editor developed by Microsoft. It is available for Windows, macOS, and Linux and is designed to be lightweight and highly customizable, making it a popular choice for developers across different platforms. VS Code comes with a range of features that make it easy for developers to write and debug code, including syntax highlighting, code completion, debugging, and Git integration. It also has a wide range of extensions and plugins that can be used to customize the editor to suit your specific needs, such as support for different programming languages, code formatting, and themes. Additionally, VS Code has a rich ecosystem of community resources, including a vast library of extensions, a marketplace for themes and plugins, and a range of tutorials and documentation.

3.2.3 Functional Requirements

- **Input data handling:** The system should be able to handle input data in the form of bird images or audio recordings. It should be able to accept different file formats and sizes and preprocess the data to ensure quality and consistency.
- **Model training and testing:** The system should be able to train a deep learning model using a dataset of labeled bird images or audio recordings. It should be able to test the model's accuracy and performance on a validation dataset and fine-tune the model parameters as needed.
- **Feature extraction and classification:** The system should be able to extract features from the input data and use them to classify the bird species using a deep neural network model. It should be able to handle different types of features, such as color, texture, shape, and sound patterns.
- **Output generation and visualization:** The system should be able to generate output results in the form of a list of bird species and their corresponding probabilities based on the input data. It should be able to visualize the results using different formats, such as charts, tables, or maps.
- **Model evaluation and optimization:** The system should be able to evaluate the model's accuracy and performance on a test dataset and optimize the model parameters to improve its performance. It should be able to handle different types of evaluation metrics, such as precision, recall, F1-score, or confusion matrix.
- **Integration and deployment:** The system should be able to integrate with other tools and applications, such as bird identification apps, citizen science platforms, or environmental monitoring systems. It should be able to deploy the model on different hardware and software environments, such as cloud-based platforms, mobile devices, or web browsers.

3.2.4 Non-Functional Requirements

- **Performance:** The system should be designed to deliver high performance in terms of speed, accuracy, and scalability. The model should be optimized for efficient computation and memory usage to handle large volumes of input data and deliver fast classification results.

- **Usability:** The system should be designed to be user-friendly and accessible to users with different levels of technical expertise. The user interface should be intuitive, responsive, and easy to navigate, and the system should provide clear and informative feedback to users.

- **Reliability:** The system should be reliable and consistent in delivering accurate classification results. The model should be trained on a large and diverse dataset of labeled bird images or audio recordings to ensure robustness and generalization.

- **Security and privacy:** The system should be designed to protect user data and ensure privacy and security. The system should adhere to data protection regulations and implement secure data transmission and storage practices.

- **Maintainability:** The system should be designed to be maintainable and easy to update and enhance. The code should be well-organized, documented, and tested, and the system should have a version control system to track changes and updates.

CHAPTER – 4

GANTT CHART

CHAPTER 4

GANTT CHART

A Gantt chart is a type of bar chart, developed by Henry Gantt that illustrates a project schedule. Gantt charts illustrate the start and finish of the terminal elements and summary elements of the project. Terminal elements and summary elements comprise the work breakdown structure of the project.

The following is the Gantt chart of the project “Bird classification based on Image or Audio using Deep Learning”.

Number	Task	Start	End	Duration(days)
1	Synopsis	19-Aug-2022	2-Sep-2022	15
2	Presentation on idea	29-Oct-2022	29-Oct-2022	1
3	Software Requirement Specification	03-Nov-2022	13-Nov-2022	10
4	System Design	05-Dec-2022	15-Dec-2022	10
5	Implementation	02-Jan-2023	01-Apr-2023	90
6	Presentation on work progress	09-Apr-2023	17-Apr-2023	09
7	Testing	20-Apr-2023	27-Apr-2023	07
8	Result and Report	29-Apr-2023	08-May-2023	10

Table 4.1 - Gantt chart of planning and scheduling of project

ACTIVITY/ MONTH	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY
SYNOPSIS										
PRESENTATION ON IDEA										
SRS										
DESIGN										
IMPLEMENTATION										
TESTING										
REPORT										

Figure 4.1 - Gantt chart

CHAPTER – 5

SYSTEM DESIGN

CHAPTER 5

SYSTEM DESIGN

System design is the process of designing the architecture, components, modules, interfaces, and data for a system to meet specific requirements. It involves determining the software and hardware infrastructure required for the system, as well as defining the relationships and interactions between the components of the system.

The purpose of system design is to create a blueprint that can guide the development of a system that meets the desired functionality, performance, scalability, security, and reliability requirements. It typically involves breaking down a complex system into smaller, more manageable components that can be developed and tested separately before being integrated into the larger system.

System design can involve a variety of techniques, including modeling and simulation, prototyping, and testing. The goal is to create a system that is efficient, maintainable, and flexible, and that can be easily modified or adapted as requirements change over time.

5.1 Architectural Diagram

An architecture diagram is a visual representation of the structure and components of a system or model. It provides a high-level overview of how the system or model is organized and how its different components interact with each other. Architecture diagrams are commonly used in the field of computer science, machine learning, and deep learning to illustrate the design and structure of complex systems or models.

Architectural design can also involve considering non-functional requirements such as performance, reliability, scalability, and security, and determining the best way to meet those requirements. This can involve selecting appropriate technologies, defining standards and protocols, and designing for fault tolerance, load balancing, and other aspects of system behavior.

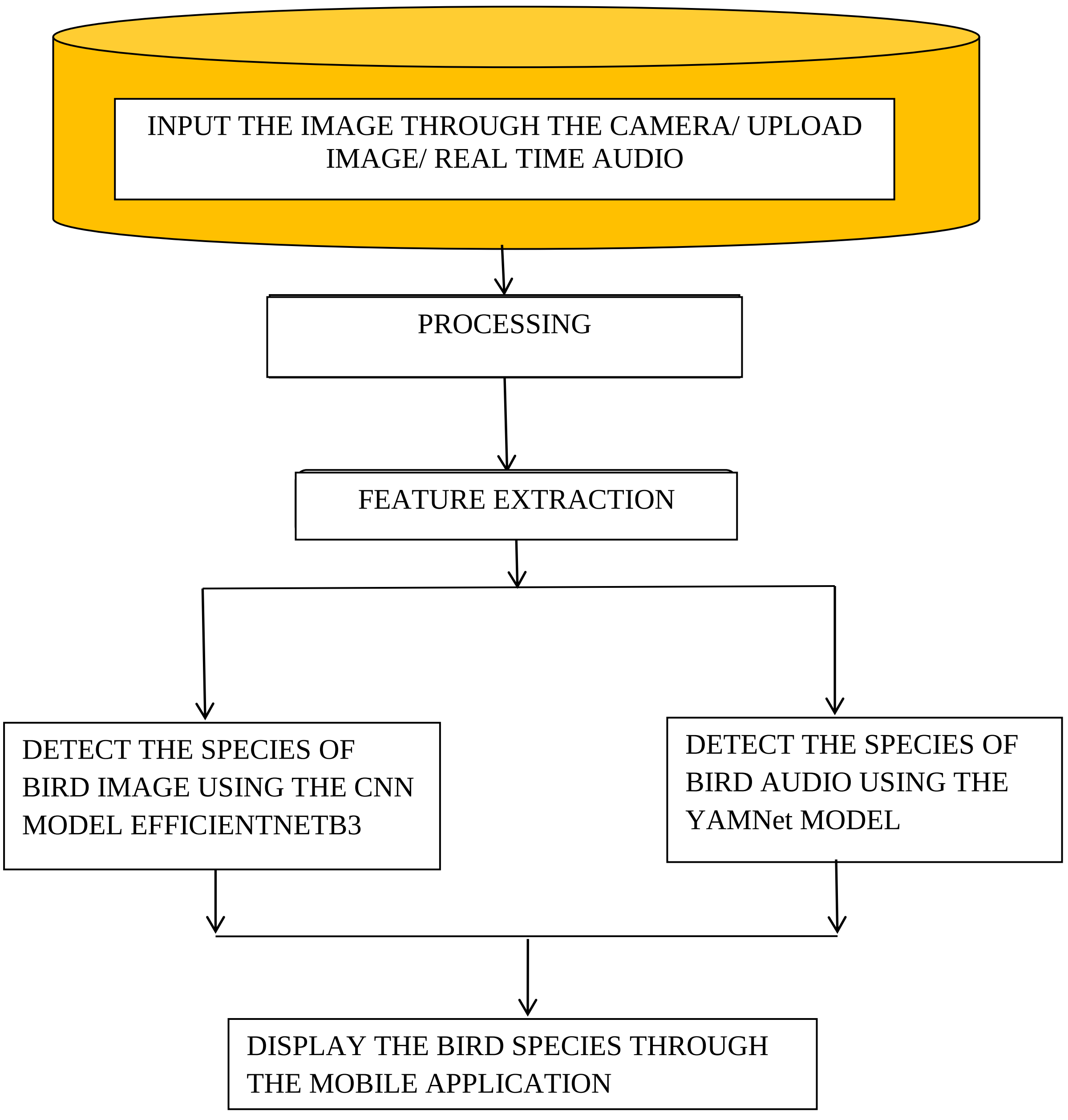


Fig 5.1 - Architectural Diagram

5.2 Use Case Diagram and Description

A use case diagram is a visual representation of the interactions between actors (users or external entities) and the system (software application or system under consideration). It is used to model the functional requirements of a system from the perspective of its users, illustrating the various use cases or scenarios that the system must support.

It includes actors, use cases, and relationships such as association, include, and extend. It helps to identify the functional requirements of the system and provide a high-level overview of the system's functionality and its interactions with external systems, making it an important tool in software development during the requirements gathering phase.

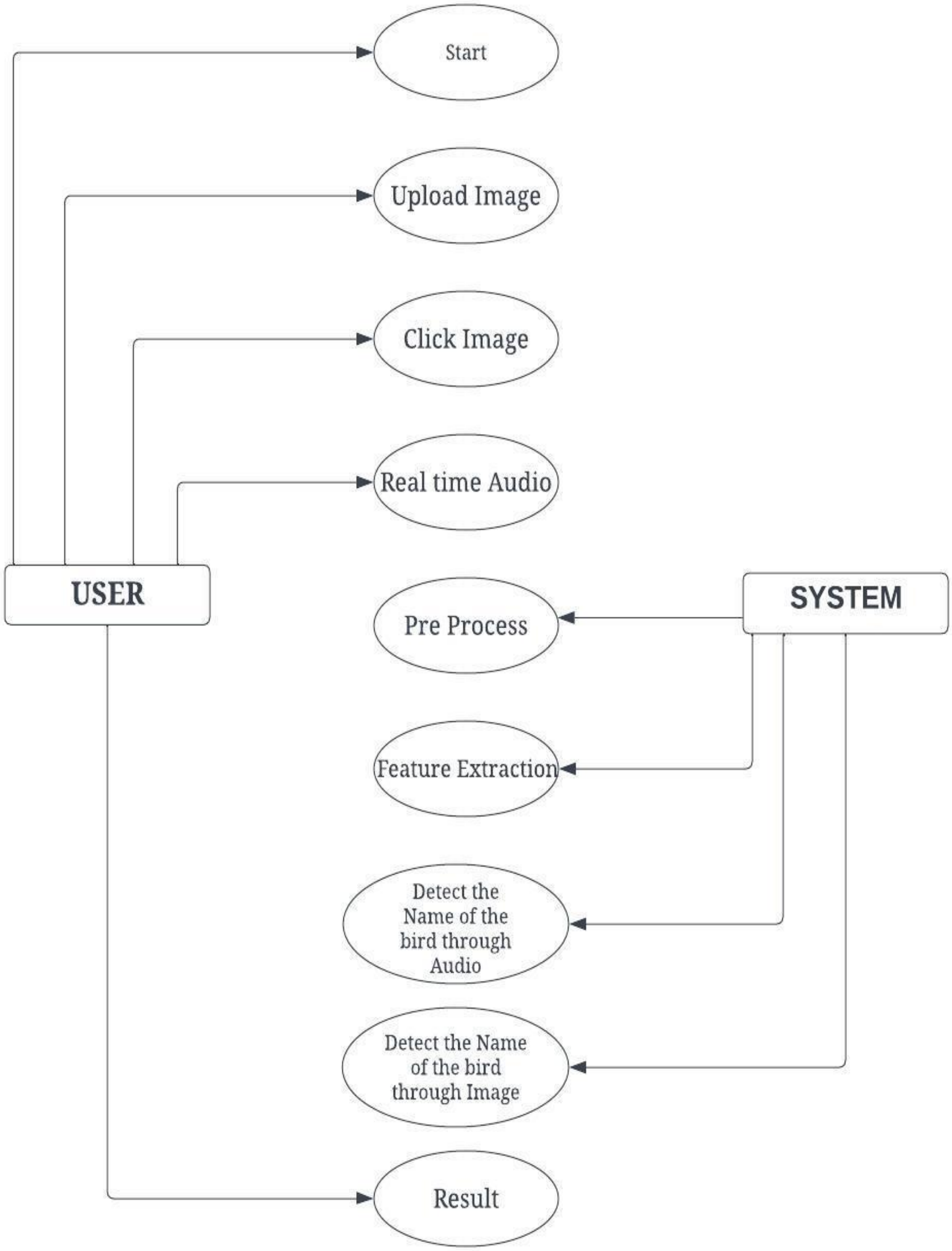


Fig 5.2 - Use Case Diagram

5.3 Sequence Diagram

A sequence diagram is a type of interaction diagram that shows the interactions or messages exchanged between objects or components in a system over time. It provides a visual representation of the dynamic behavior of a system, showing the order of interactions between objects or components in a sequence.

Sequence diagrams are useful for visualizing the flow of interactions between objects or components in a system, and can help identify potential issues or improvements in the system's behavior.

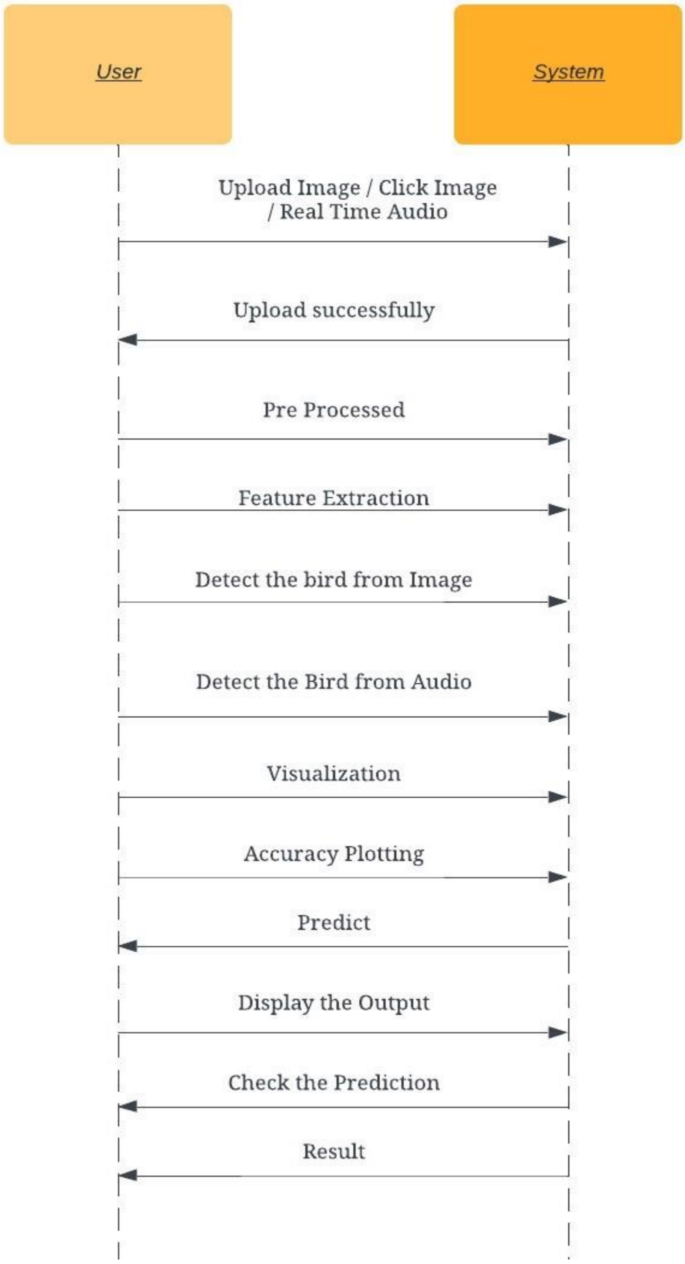


Fig 5.3 - Sequence Diagram

5.4 Activity Diagram

An activity diagram is a type of behavior diagram in UML (Unified Modeling Language) that models the flow of activities or actions within a system or process. It is used to represent the steps, decisions, and flows of activities or actions in a graphical and intuitive way, providing a visual representation of the dynamic behavior of a system or process.

Activity diagrams are commonly used during the analysis and design phases of software development to model the flow of activities or actions in a system or process. They can help visualize the dynamic behavior of a system, identify potential issues or bottlenecks, and ensure the correct flow of activities to achieve the desired outcomes.

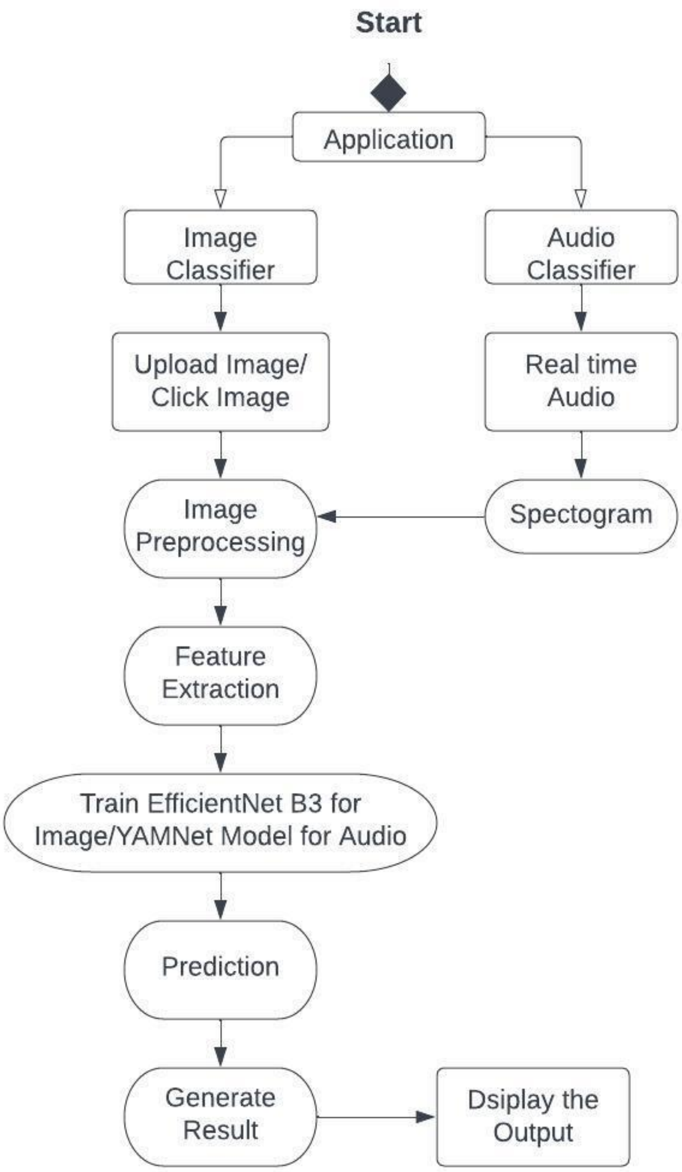


Fig 5.4 - Activity Diagram

5.5 Data Flow Diagram

A Data Flow Diagram (DFD) is a graphical representation of the flow of data or information within a system. It is used to model how data is input, processed, stored, and output in a system or process, showing the flow of data between different components or processes. DFDs are commonly used in system analysis and design to understand and document the data flow and interactions in a system.

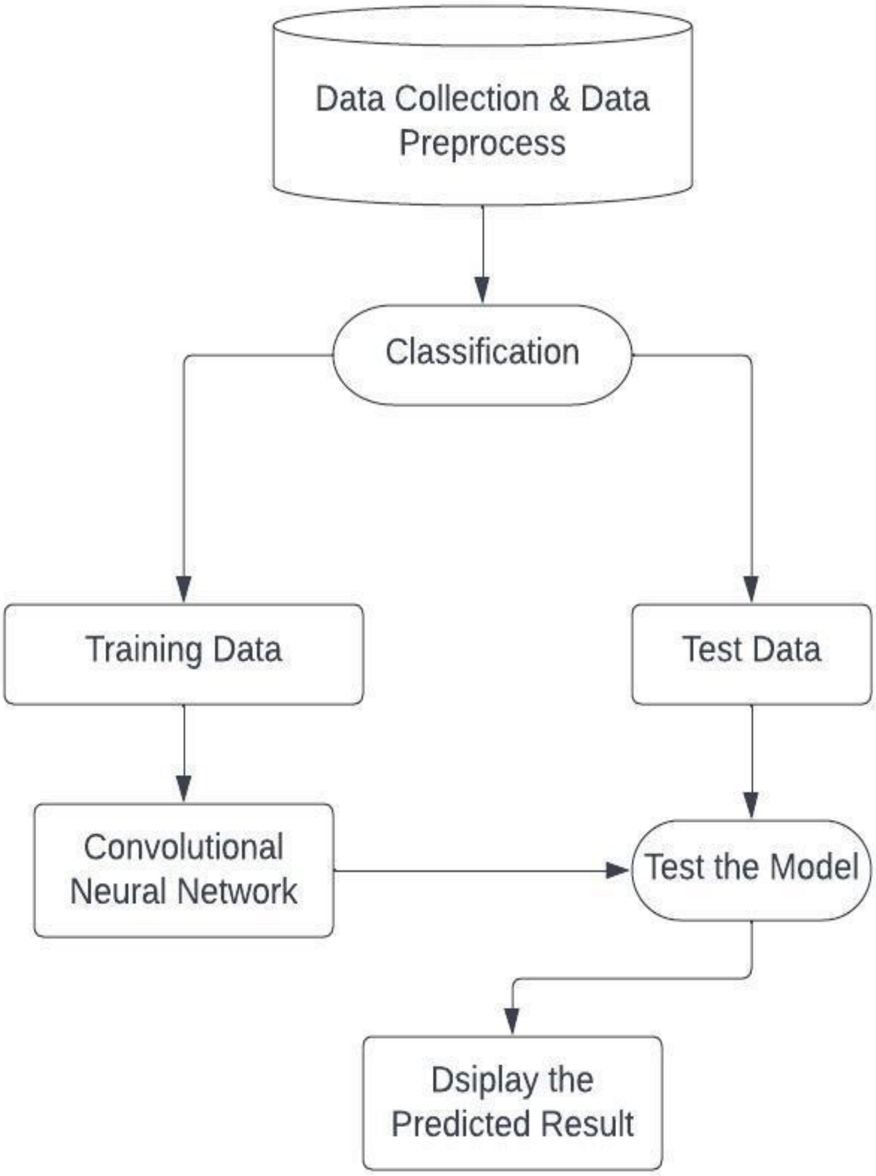


Fig 5.5 - Data Flow Diagram

CHAPTER – 6

IMPLEMENTATION

CHAPTER 6

IMPLEMENTATION

6.1 Module Implementation

Implementing a bird classification system using deep learning requires several key steps, including data collection and preprocessing, model selection and adaptation, training and evaluation, and deployment. The use of models, such as EfficientNet-B3 for image classification and YAMNet for audio classification, can greatly simplify the process by leveraging existing knowledge and expertise. However, it is important to fine-tune these models for the specific classification task and carefully evaluate their performance on testing data. By following these steps, developers can create accurate and efficient bird classification systems based on either image or audio inputs, which can be useful for a variety of real-world applications.

6.1.1 Image Classification Model

EfficientNet-B3 is a convolutional neural network (CNN) model that is part of the EfficientNet family of models, which are known for their efficiency in terms of both computational resources and model size. EfficientNet-B3 is larger and more powerful compared to EfficientNet-B0 and EfficientNet-B1, but still relatively efficient compared to many other CNN models.

The basic building block of the EfficientNet-v3 architecture is the EfficientNet-v2 block, which includes a combination of 3x3 and 1x1 convolutional layers along with a squeeze-and-excitation block to capture feature interdependencies. EfficientNet-v3 also includes a new activation function called Swish, which has been shown to improve model performance.

EfficientNet-v3 includes a total of 18 EfficientNet-v2 blocks arranged in a hierarchical structure, with features from each block being passed to the next block through a skip connection. The network also includes several feature fusion blocks that combine features from different blocks to capture both high-level and low-level features. In addition to the main convolutional layers.

Bird classification using the EfficientNet-B3 model would involve the following steps:

- **Data preparation:** Gather a dataset of bird images for training, validation, and testing the model. This dataset should be diverse and representative of the different bird species you want to classify.

- **Data pre-processing:** Resize the bird images to a consistent size that the EfficientNet-B3 model expects as input.
- **Model training:** Split your pre-processed bird image dataset into training, validation, and test sets.
- **Model evaluation:** After training, evaluate the performance of the trained EfficientNet-B3 model on the validation and test sets.
- **Model fine-tuning:** If the model's performance is not satisfactory, you can perform model fine-tuning to further improve its accuracy.

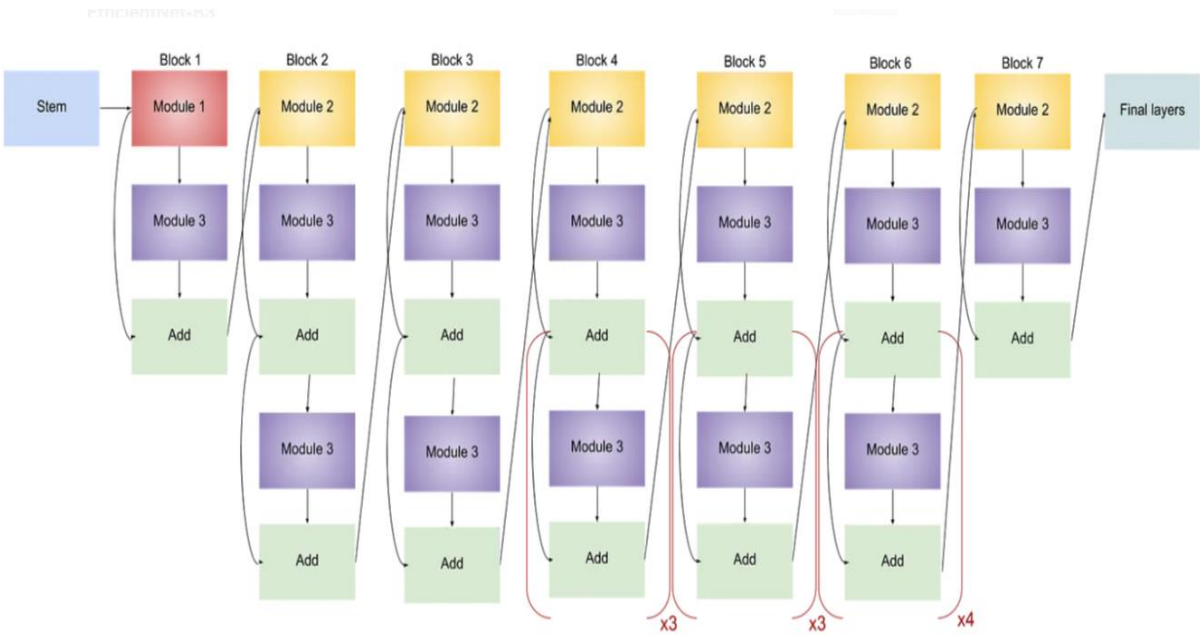


Fig 6.1 - Block Diagram of EfficientNet-B3

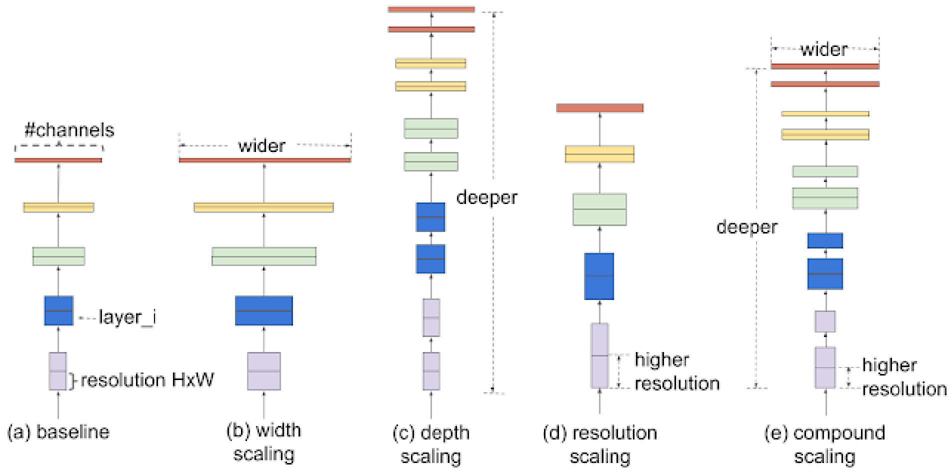


Fig 6.2 - EfficientNet-B3 Model Scaling

6.1.2 Audio Classification Model

YAMNet (Yet Another Multi-Scale Convolutional Neural Network) is a deep learning model designed for environmental sound classification. It is pre-trained on the AudioSet dataset, which contains over 2 million 10-second audio clips across a diverse range of sound classes. The YAMNet model consists of a stack of convolutional layers, followed by global average pooling, a fully connected layer, and a softmax activation layer. The convolutional layers are designed to capture different time-frequency patterns at multiple scales, which allows the model to effectively capture both local and global features of the input audio signal.

The YAMNet model can classify sounds into 521 audio event classes, including various types of human activities, animal vocalizations, musical instruments, and environmental sounds. The model outputs a probability distribution over the 521 classes, indicating the likelihood of the input audio signal belonging to each class.

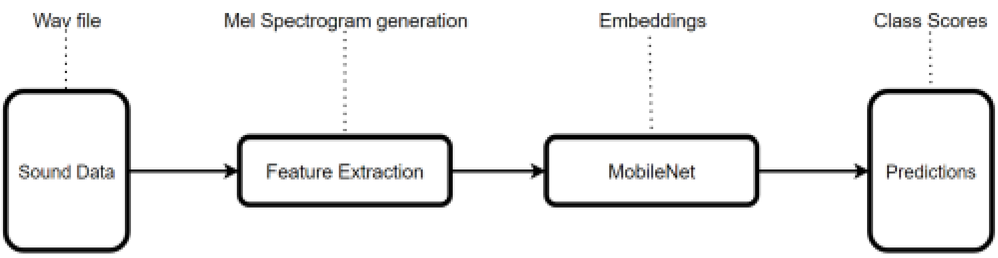


Fig 6.3 - Block Diagram of YAMNet Model

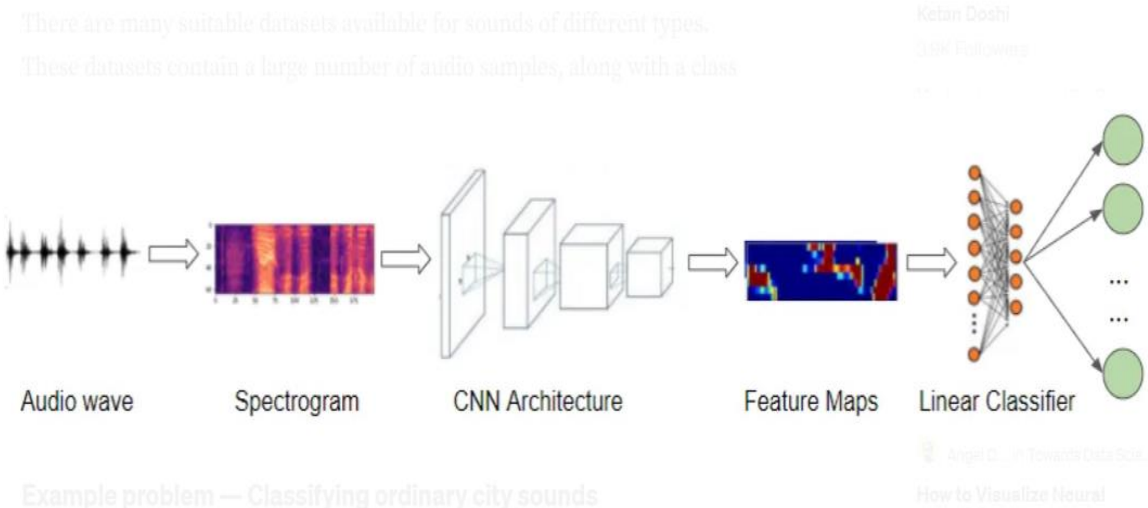


Fig 6.4 - Architectural Diagram of YAMNet Model

CHAPTER – 7

TESTING

CHAPTER 7

TESTING

Testing is an important phase in the development life cycle of the product. During the testing, the program to be tested was executed with a set of test cases and the output of the program for the test cases was evaluated to determine whether the program is performing as expected. Errors were found and corrected by using the following testing steps and correction was recorded for future references. Thus, a series of testing was performed on the system before it was ready for implementation. An important point is that software testing should be distinguished from the separate discipline of Software Quality Assurance (SQA), which encompasses all business process areas, not just testing.

7.1 Testing Levels

Testing is part of Verification and Validation. Testing plays a very critical role for quality assurance and for ensuring the reliability of the software.

The objective of testing can be stated in the following ways.

- A successful test is one that uncovers as-yet-undiscovered bugs.
- A better test case has high probability of finding un-noticed bugs.
- A pessimistic approach of running the software with the intent of finding errors.

Testing can be performed in various levels like unit test, integration test and system test.

7.1.1 Unit Testing

Unit testing tests the individual components to ensure that they operate correctly. Each component is tested independently, without other system component. This system was tested with the set of proper test data for each module and the results were checked with the expected output. Unit testing focuses on verification effort on the smallest unit of the software design module.

7.1.2 Integration Testing

Integration testing is another aspect of testing that is generally done in order to uncover errors associated with the flow of data across interfaces. The unit-tested modules are grouped together and tested in small segment, which makes it easier to isolate and correct errors. This approach is continued until we have integrated all modules to from the system.

7.1.3 System Testing

System testing tests a completely integrated system to verify that it meets it requirements. The proposed system is used to classify the given bird image or audio through app.

7.1.4 Acceptance Testing

After the system tests for its accuracy, efficiency, and reliability in Classifying the bird species. The test cases will be designed to cover all possible scenarios and edge cases to ensure that the model performs well in real-world scenarios. The test results will be evaluated against the predetermined acceptance criteria to determine whether the model meets the specified requirements and is fit for deployment.

7.2 Test Cases

A test case is a software testing document, which consists of events, action, input, output, expected result and actual result. Technically a test case includes test description, procedure, expected result and remarks. Test cases should be based primarily on the software requirements and developed to verify correct functionality and to establish conditions that reveal potential errors.

Individual PASS/FAIL criteria are written for each test case. All the tests need to get a PASS result for proper working of an application.

Test No.	Test Case	Expected Results	Actual Results	Status
1	Open Bird Classifier Image application.	The application should launch successfully without any errors.	Application processed the image and displayed the predicted bird species.	Pass
2	Open Bird Classifier Audio application.	The application should launch successfully without any errors.	Application processed the sound and displayed the predicted bird species.	Pass

3	Click on Predict in Bird Classifier application for Image.	The user should be able to choose an image or record audio of a bird. The "Predict" button should initiate the classification process and provide the predicted bird species.	Predicted the Bird Species	Pass
4	Click on Predict in Bird Classifier application for Audio.	The application should accurately classify the bird species based on the recorded sound using the YAMNet model.	The Model Successfully Predicted the Bird Species.	Pass
5	Click on Predicted Bird’s name in the Classifier application.	The user should be able to choose an image of a bird and click on the "Predict" button to initiate the classification process.	When user click on Bird name, then it will navigate to Google chrome and user can see the information of the Predicted Bird	Pass
6	Predict Bird using camera image	The "Choose Image" button should allow the user to select an image from the device gallery or capture an image using the device camera.	The "Choose Image" button will allow the user to select an image from the device gallery or capture an image using the device camera.	Pass

Table 7.1 - Test Cases

CHAPTER – 8

RESULTS AND SNAPSHOTS

CHAPTER 8

RESULTS AND SNAPSHOTS

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 below.

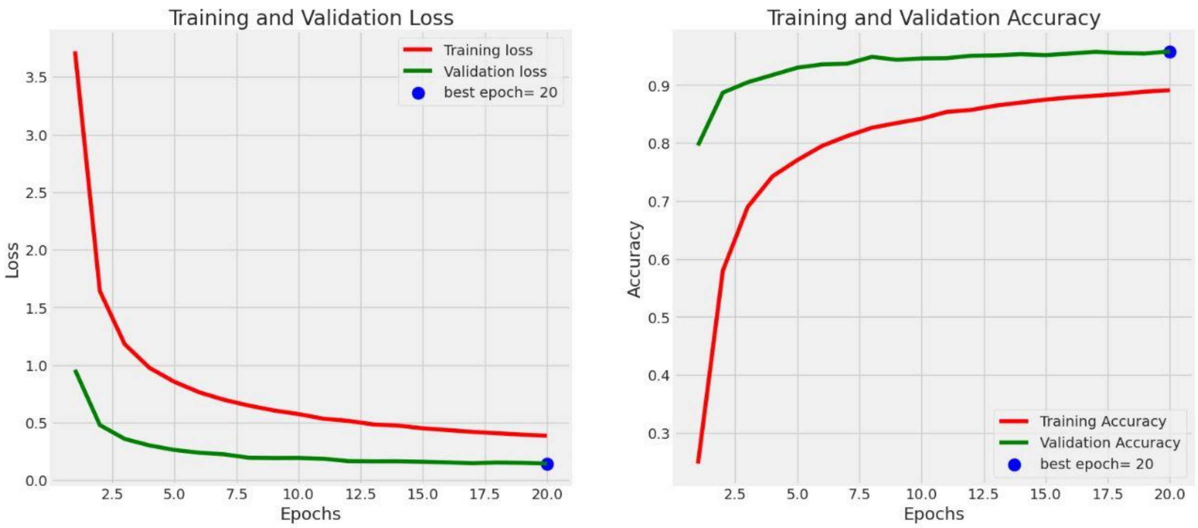


Fig 8.1 - EfficientNet B3 Model Accuracy and Loss

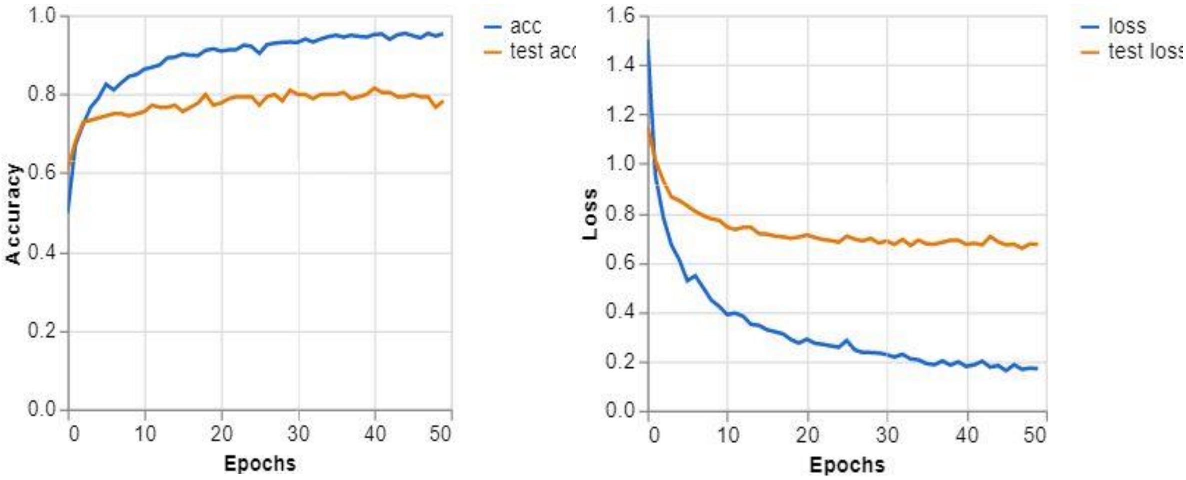


Fig 8.2 - YAMNet Model Accuracy and Loss

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 8.3 -Application Interface

Fig 8.4 - Application Input

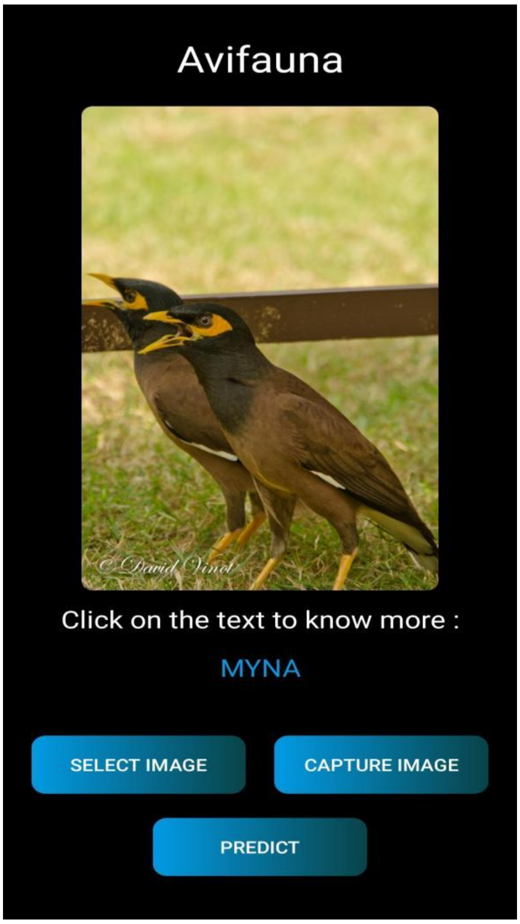


Fig 8.5 – Image Result Sample – 1

Fig 8.6 - Image Result Sample – 2

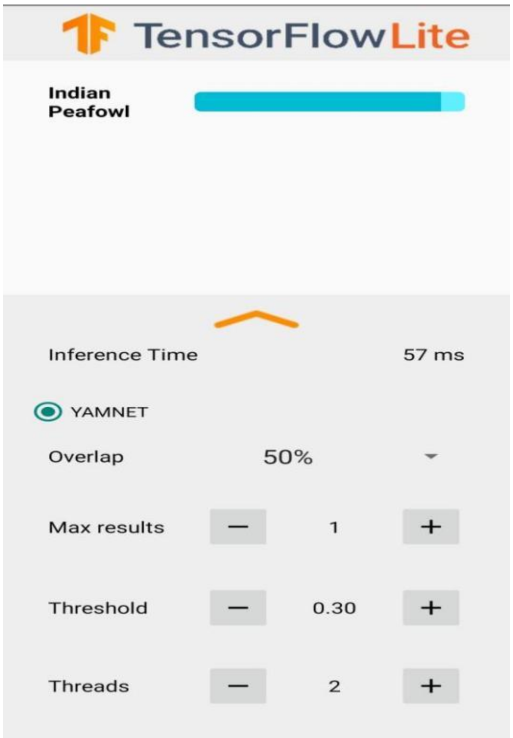


Fig 8.7 - Audio Result Sample - 1

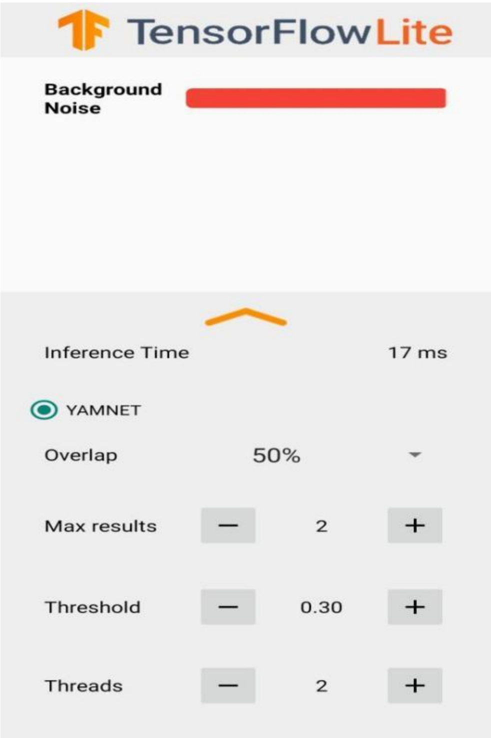


Fig 8.8 - Audio Result Sample - 2

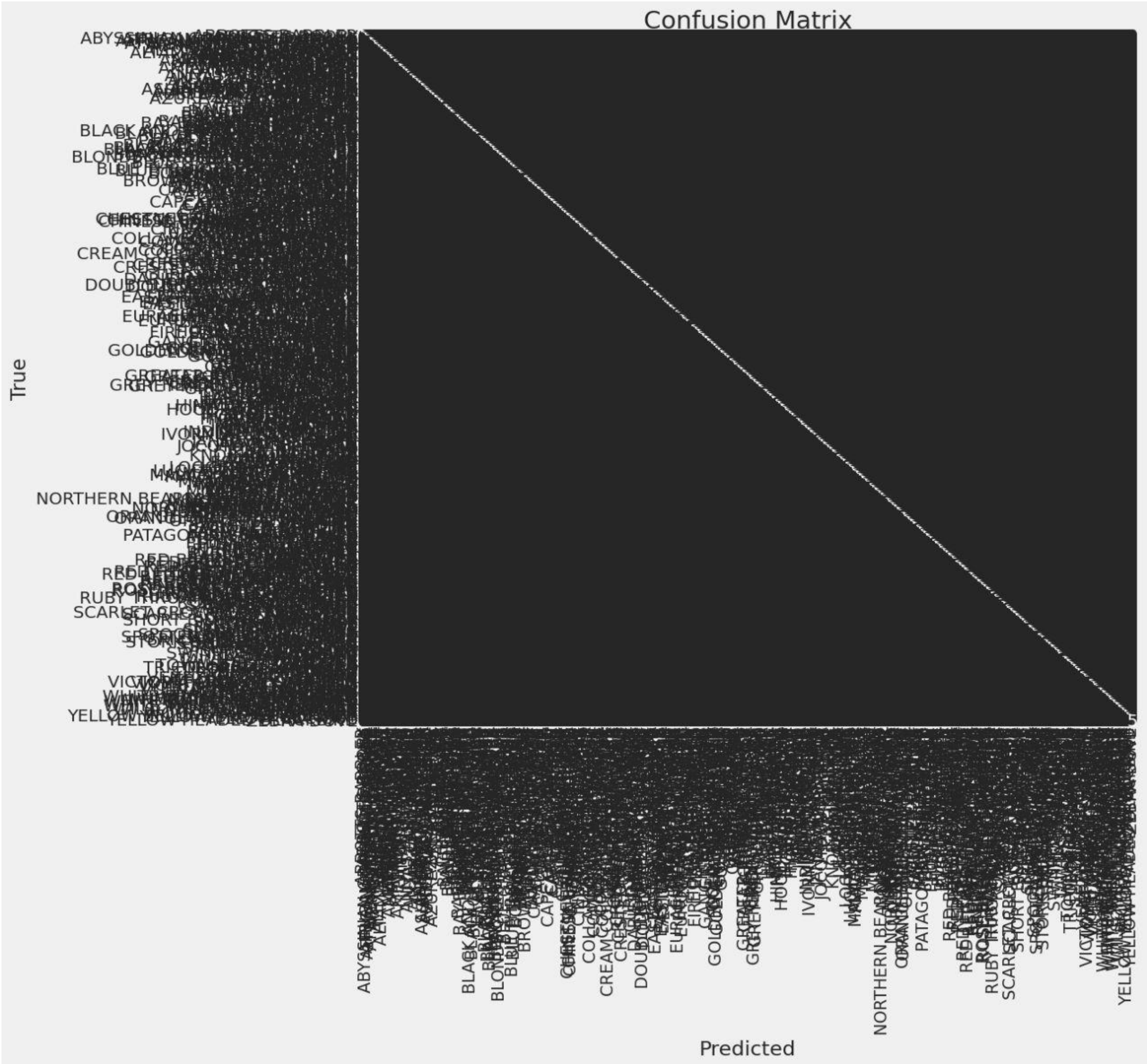


Fig 8.9 - Confusion Matrix

A confusion matrix is a table that is used to evaluate the performance of a machine learning model by comparing the predicted and actual values of a classification problem. It is a matrix of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

The confusion matrix is a valuable tool for measuring the accuracy, precision, recall, F1-score, and other performance metrics of a classification model. It helps to understand how well the model is performing on each class and how it is making errors**.**

CHAPTER – 9

CONCLUSION AND FUTURE WORK

CHAPTER 9

CONCLUSION AND FUTURE WORK

Deep Learning convolutional neural network (CNN) models for bird classification based on image data holds great promise. These models have demonstrated high accuracy and robustness in identifying bird species from images, leveraging their ability to learn complex patterns and features such as coloration, morphology, and structural characteristics.

Deep learning CNN models, trained on large and diverse datasets of bird images, can generalize well to unseen data and can handle variations in environmental conditions, making them suitable for real-world bird classification tasks. Transfer learning, which involves leveraging pre-trained models, can further enhance their performance even with limited data.

However, challenges remain, including the limited availability of annotated data for rare or understudied bird species, potential biases in training data, and the need for interpretability and explain ability of deep learning models. Additionally, addressing issues such as data quality, class imbalance, and model robustness against adversarial attacks is crucial for ensuring reliable bird classification results.

Despite these challenges, the use of deep learning CNN models for bird classification based on image data has the potential to revolutionize bird monitoring, conservation, and research efforts. These models can provide valuable insights into avian biodiversity, habitat assessment, and species distribution mapping, which can inform conservation strategies and management decisions.

Further advancements in deep learning techniques, coupled with the availability of larger and more diverse datasets, are expected to continue improving the accuracy, robustness, and interpretability of deep learning CNN models for bird classification. These advancements have the potential to significantly contribute to our understanding of avian ecology, conservation, and the sustainable management of bird populations.

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 Technology 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 Liu1:00000000 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

PAPER PUBLICATION

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:

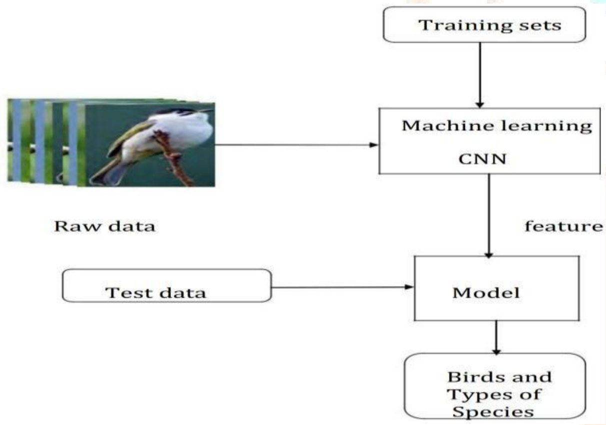
This paper discusses the problem of ConvNet scaling and identify that carefully balancing network width, depth, and resolution is an important but missing piece, preventing us from better accuracy and efficiency. To address this issue, they proposed a simple and highly effective compound scaling method, 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.



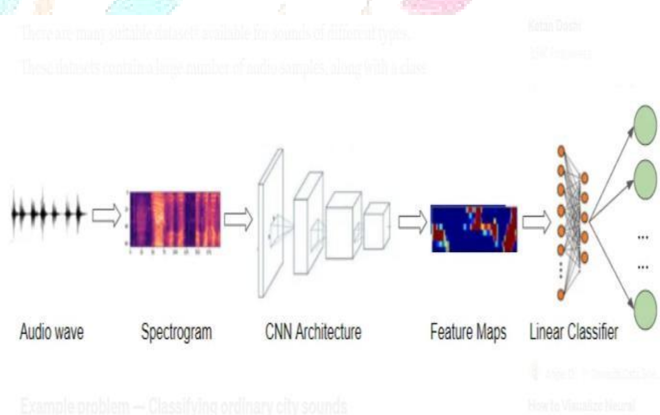


Fig 1- Image Model Workflow **Fig 2 - Audio Model Workflow**

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.

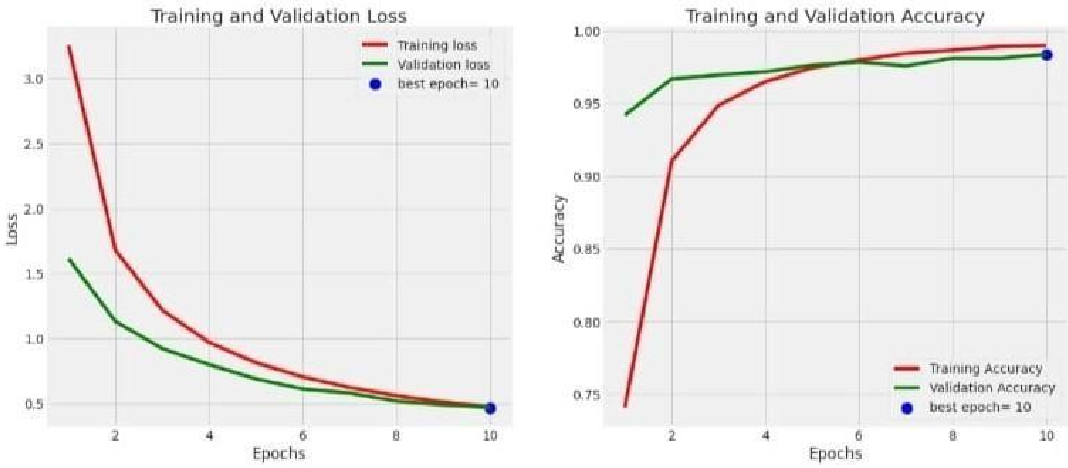


Fig 3 - EfficientNet B3 Model Accuracy and Loss

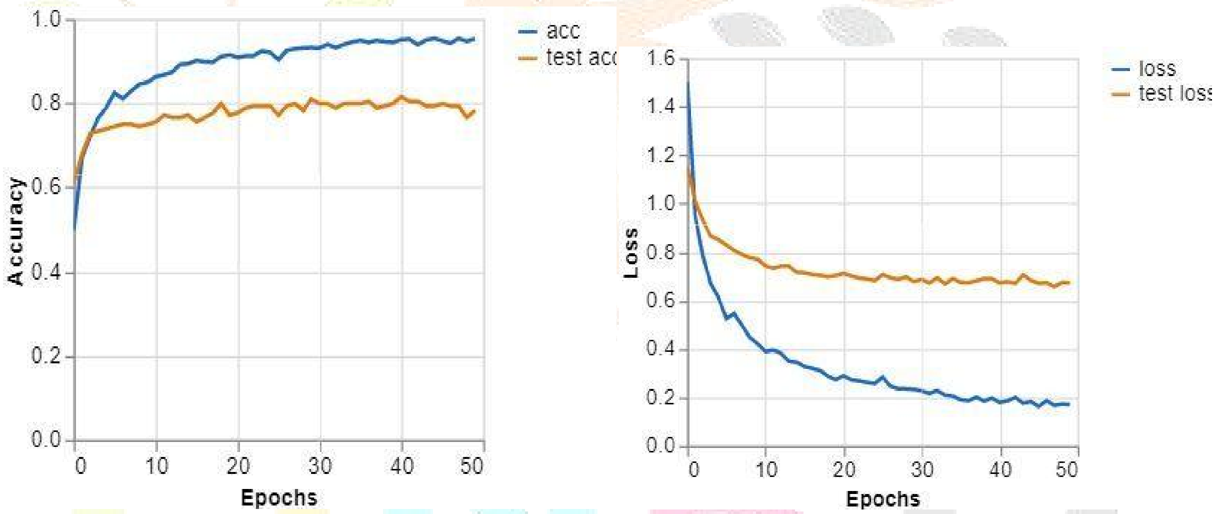


Fig 4 - YAMNet Model Accuracy and Loss

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

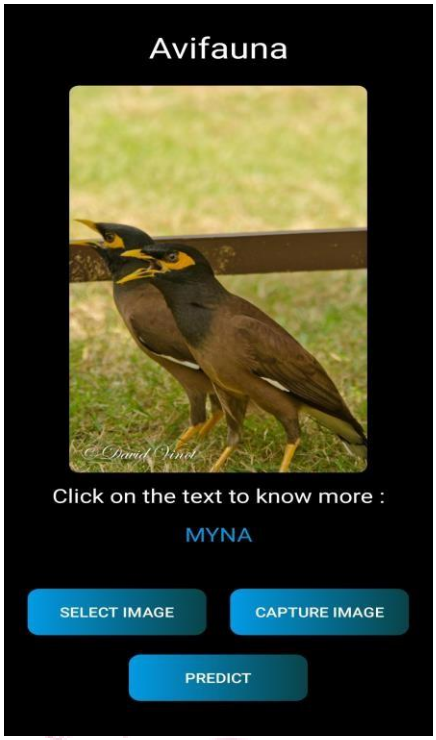


Fig 7 - Application Result

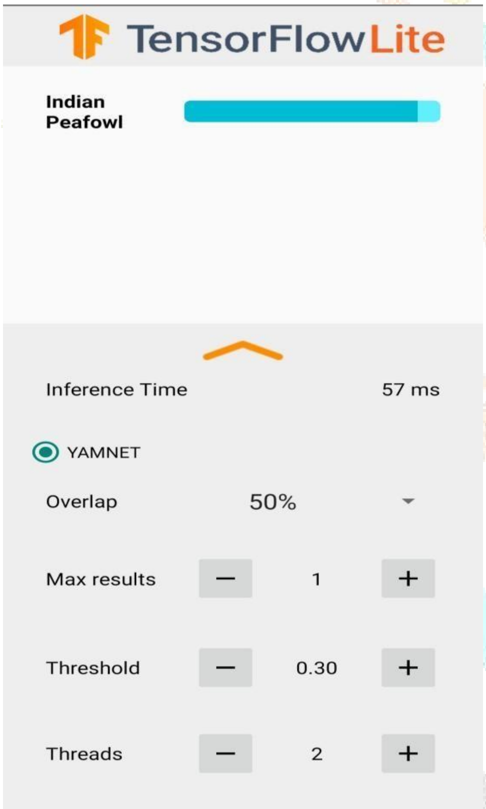


Fig 8 - Audio Result-1

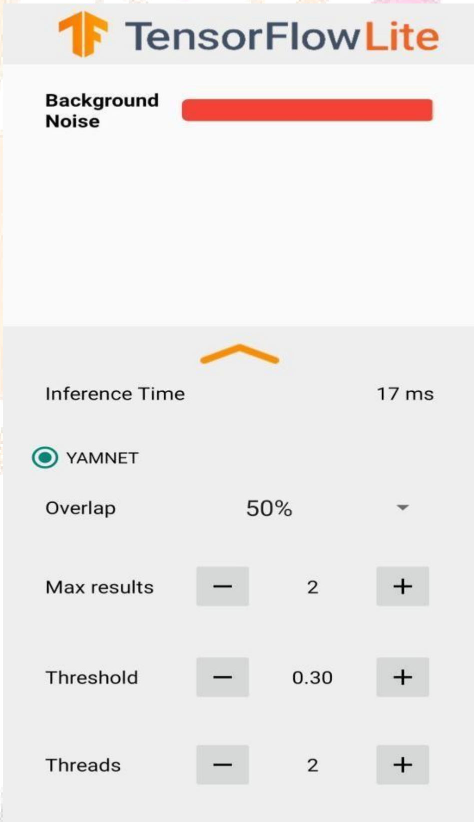


Fig 9 - Audio Result-2

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 Technology 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