# Department of Artificial Intelligence and Machine Learning

## V Semester – 2022 Scheme

## Lab Experiential Learning – 2024-25

## Course : ARTIFICIAL NEURAL NETWORKS AND DEEP LEARNING (AI253IA)

# Title –Deep Audio Classification

**Team Members**

| | |
|---|---|
| 1RV23AI403 | Omkar Babu Mastamardi |
| 1RV22AI062 | Varun Banda |
| 1RV22AI068 | Labdhi Ranka |

# Introduction



1. Agenda
2. Introduction
3. Literature Survey
4. Summary of LS
5. Requirement analysis – hardware and software specification
6. System architecture ( -- ANN-DL architecture ) Eg .. architecture of CNN
7. Methodology
8. Module specification
   a. Module 1 : data collection and pre processing
      i. Input ii. Process iii output
   b. Module 2 : Implementation of ANN / DL algorithm
      i. Input ii. Process iii output
   c. Module 3 : testing and validation
      i. Input ii. Process iii output

RV College of Engineering®

# Agenda

**Problem:** Monitoring Capuchin bird populations in the Amazon Rainforest is inefficient, labor-intensive, and affected by overlapping sounds and environmental noise.

**Proposed Solution:** A Deep Audio Classifier for Capuchin Bird Density using advanced deep learning techniques.

**Goal:** Automate the detection and classification of Capuchin bird calls from audio recordings.

**System Features:**

- **Noise Reduction:** To minimize interference from environmental sounds.
- **Spectrogram Generation**: To analyze bird calls more effectively.
- **Data Augmentation**: To improve the model's ability to handle noisy datasets.
- **Outcome**: Provide accurate population density estimates and efficient analysis of large, noisy datasets.

**Current Methods**: Lack scalability and effectiveness.

**Motivation**:

- Develop an efficient solution to enhance biodiversity monitoring.
- Aid conservation efforts.
- Showcase AI's potential in ecological research.

# Introduction

- Monitoring biodiversity in tropical forests is crucial for conservation, particularly for species like the Capuchin bird, known for its unique vocalizations.
- Traditional methods for detecting bird calls are labor-intensive and limited in processing large data volumes.
- This research explores the use of deep learning, specifically Convolutional Neural Networks (CNNs), to automate Capuchin bird call detection in audio recordings.
- By analyzing spectrograms of forest sounds, CNNs can identify Capuchin bird calls amidst complex backgrounds.
- The study aims to evaluate CNN-based detection, compare it to existing methods, and highlight its potential to enhance biodiversity monitoring and support conservation efforts.

# Objectives

- To develop a system that can automatically detect Capuchinbird calls from tropical forest audio recordings.
- To leverage Convolutional Neural Networks (CNNs) to analyze spectrograms and improve the accuracy of call detection.
- To create a tool to assist ornithologists and conservationists in monitoring Capuchinbird populations and the health of tropical ecosystems.
- To design a system capable of distinguishing Capuchinbird calls from a noisy background of other species and environmental sounds.

# Literature Survey



| Sl No | Author and Paper title | Details of Publication | Summary of the Paper |
|---|---|---|---|
| 1 | Authors:<br>• Khalid Zaman<br>• Melike Sah<br>• Cem Direkoglu<br>• Masashi Unoki<br><br>• **Paper Title:**A Survey of Audio Classification Using Deep Learning | • Journal: IEEE Access<br>• Volume: 11<br>• Year: 2023<br>• DOI: 10.1109/ACCESS.2023.3318015 | • The paper reviews the use of deep learning models, such as CNNs, RNNs, autoencoders, and transformers, for classifying audio signals into categories like speech, music, and environmental sounds, emphasizing the role of feature extraction through spectrograms. |
| 2 | Authors:<br>• hajin Prince<br>• Justin Jojy Thomas<br>• Sharon Jostana J<br>• Kakarla Preethi Priya<br>• J Joshua Daniel<br><br>• Paper Title:Music Genre Classification using Deep Learning | • Conference: 2022 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)<br>• DOI: 10.1109/CSITSS57437.2022.10026394 | • The paper reviews deep learning methods, especially Convolutional Neural Networks (CNNs), for classifying music into genres like Pop, Jazz, and Rock, using the GTZAN dataset of 1,000 audio clips. |

# Literature Survey



| Sl No | Author and Paper title | Details of Publication | Summary of the Paper |
| --- | --- | --- | --- |
| 3 | Authors:<br>• Ching Seh Wu<br>• Sasanka Kosuru<br>• Samaikya<br>• Tippareddy<br><br>• **Paper Title:** Bird Species Identification from Audio Data | • 2023 IEEE Ninth International Conference on Big Data Computing Service and Applications (BigDataService),<br><br>• DOI:10.1109/BigDataService58306.2023.00015 | • This paper focuses on identifying bird species using audio recordings. The authors explain that birds are important indicators of environmental changes, and tracking their populations can provide valuable insights into ecosystem health.<br><br>• The study uses K-NN, SVM, and SGD to classify bird species from audio features. |
| 4 | Authors:<br>• Mohammed Safwat<br>• Imran Afia Fahmida Rahman<br>• Sifat Tanvir<br>• Hamim Hassan<br><br>• **Paper Title:**An Analysis of Audio Classification Techniques using Deep Learning Architectures | • Conference: Sixth International Conference on Inventive Computation Technologies (ICICT 2021)<br>• Publisher: IEEE<br>• Year: 2021<br>• DOI:10.1109/ICICT50816.2021.9358774 | • The paper studies audio classification using deep learning, comparing CNNs and RNNs.<br><br>• Preprocessing techniques like normalization, segmentation, and envelope functions improve results. |

# Summary of LS

- To utilize a combination of CNNs for feature extraction and RNNs (like LSTMs) for capturing temporal dependencies in audio classification.
- To emphasize the significance of using appropriate feature extraction techniques, such as Mel-frequency Cepstral Coefficients (MFCCs) and spectrograms, to represent audio signals effectively.
- To highlight the use of deep learning for the detection and analysis of bird calls, enhancing the accuracy of call detection through convolutional neural networks (CNNs) or other architectures.
- To consider the broader implications, such as its potential use in conservation efforts, biodiversity monitoring, and ecological research, helping secure funding or support from environmental organizations.
- To explore machine learning classifiers like K-Nearest Neighbors (KNN), Random Forests, and Artificial Neural Networks (ANNs).

# Analysis – Hardware and Software Specifications

## Hardware Requirements:

**1. GPU System for Training**

- CPU: Modern CPU for preprocessing and operations.
- GPU: CUDA-compatible GPU (e.g., NVIDIA) for model training.
- RAM: 8GB or higher.
- Storage: SSD (256GB or more).
- Network: High-speed Ethernet or Wi-Fi.
- Operating System:Compatible with TensorFlow (Linux, Windows, macOS)

# Analysis – Hardware and Software Specifications

## Software Requirements:

1.Programming Language: Python 3.x

2.Libraries: TensorFlow, Librosa, Matplotlib

3.Development Environment: Jupyter Notebook, PyCharm IDE, Visual Studio code

4.Dataset Sources: ESC-50, UrbanSound8K

# System Design

## Data Flow Diagram Level 0

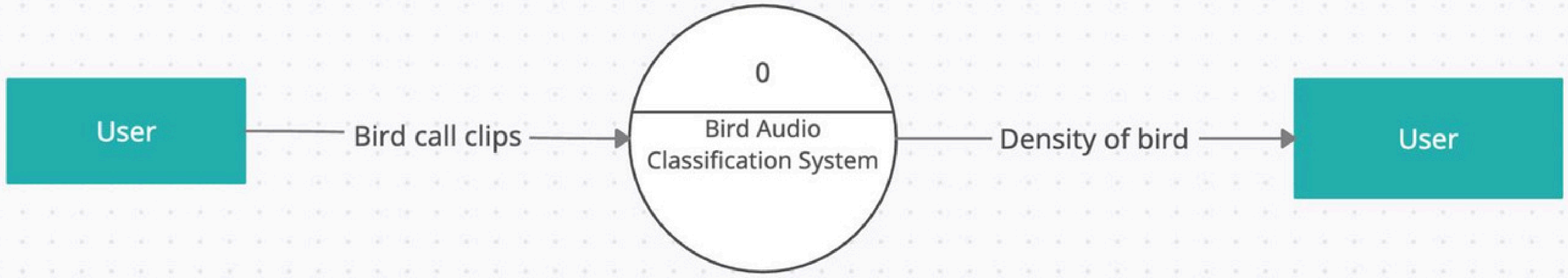



Figure 4.3 Data Flow Diagram Level 0

# System Design

## Data Flow Diagram Level 0

At this level, the system is represented as a single process, "Bird Audio Classification System."

**1.External Entities:**

- User: Provides the input in the form of bird call audio clips.
- System Output: Sends back the result, which represents the estimated density of birds in the given area.

**2.Data Flows:**

- The user uploads a bird call audio clip.
- The system processes the audio and returns the bird density estimation to the user.

RV College of Engineering®

# System Design

## Data Flow Diagram Level 1

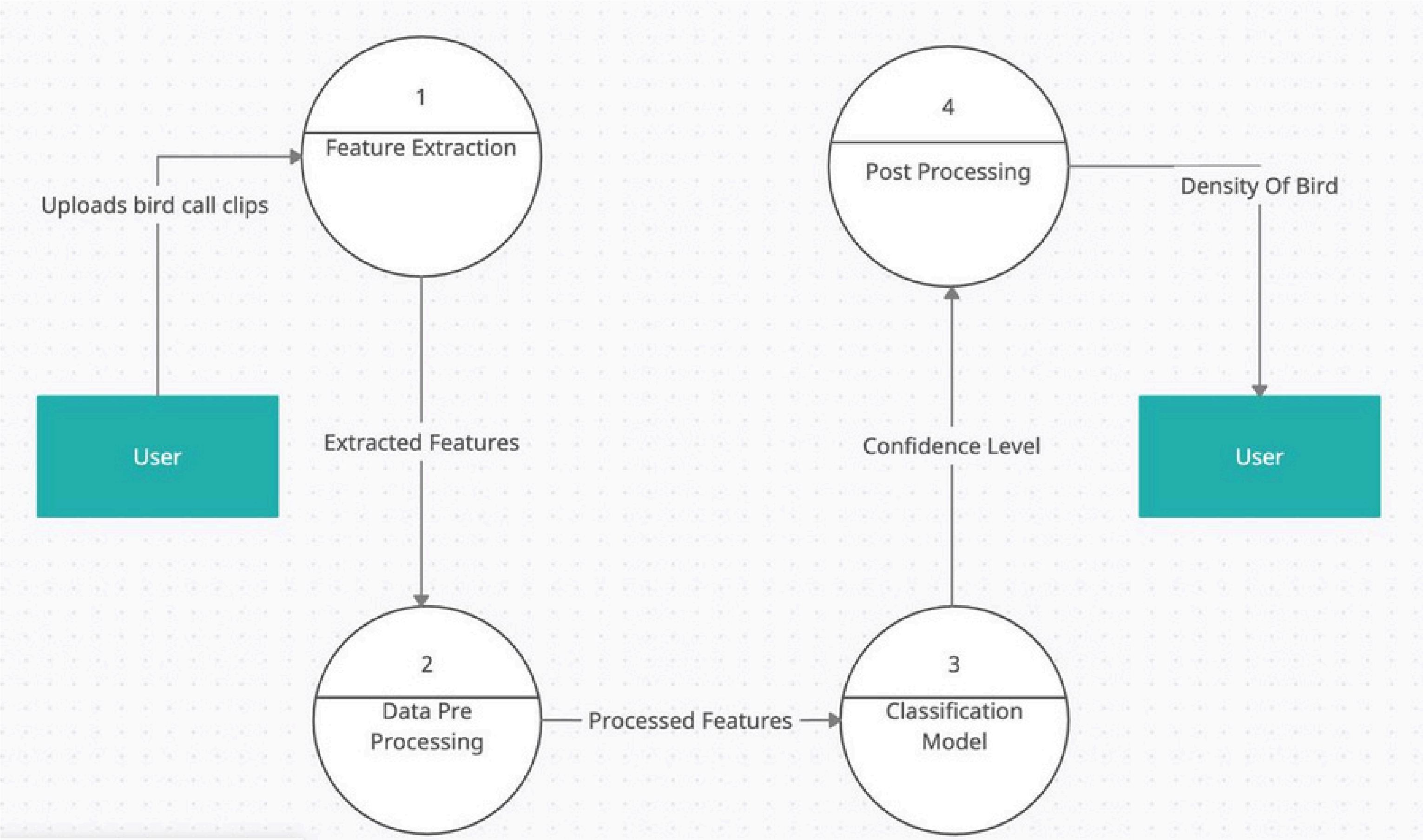



Figure 4.4 Data Flow Diagram Level 1

# System Design

## Data Flow Diagram Level 1

**1.Feature Extraction:** Extracts key acoustic features from the uploaded bird call audio to prepare for analysis.

**2.Data Preprocessing**: Cleans and normalizes the extracted features, removing noise and irrelevant sounds.

**3.Classification Model:** Uses a trained model to analyze the preprocessed features and classify bird species or detect activity levels.

**4.Density Estimation:** Uses classification results to estimate bird population density in the recorded area.

# System Design

## Data Flow Diagram Level 1

**Data Flows:**

- The audio clip flows from the user to the feature extraction process.
- Extracted features move to the data preprocessing stage.
- The preprocessed data is passed to the classification model, which predicts bird species or activity levels.
- The classification output is used in the density estimation process.
- The final bird density result is sent to the user, completing the process.

# System architecture ( ANN-DL Architecture )

**We will start with sound files, convert them into spectrograms, input them into a CNN plus Linear Classifier model, and produce predictions about the class to which the sound belongs, in our model it will be 1 in case of if we hear the bird sound and 0 in case of we hear other sound.**

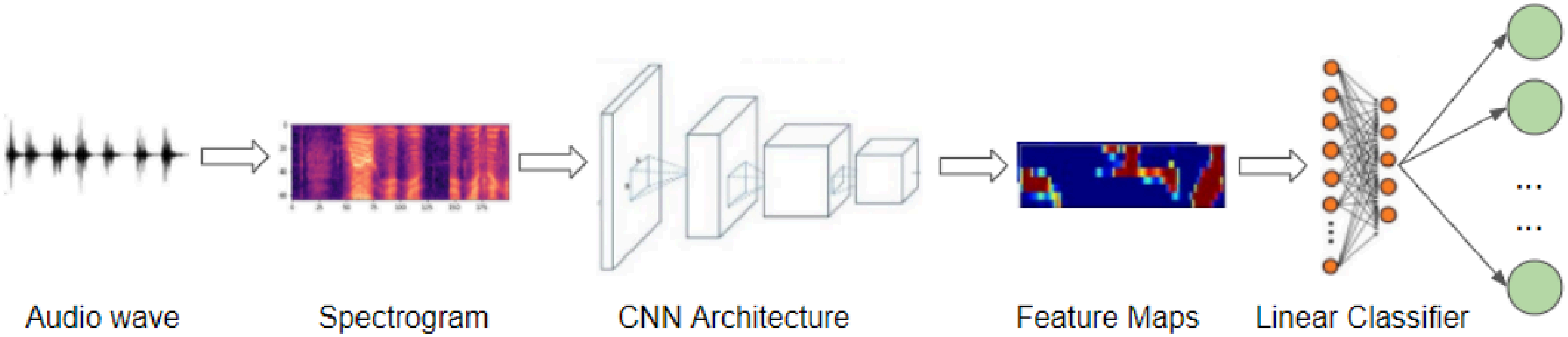



**Figure 1.1 Sound Classification System Using Deep Neural Networks**

# System architecture ( ANN-DL Architecture )

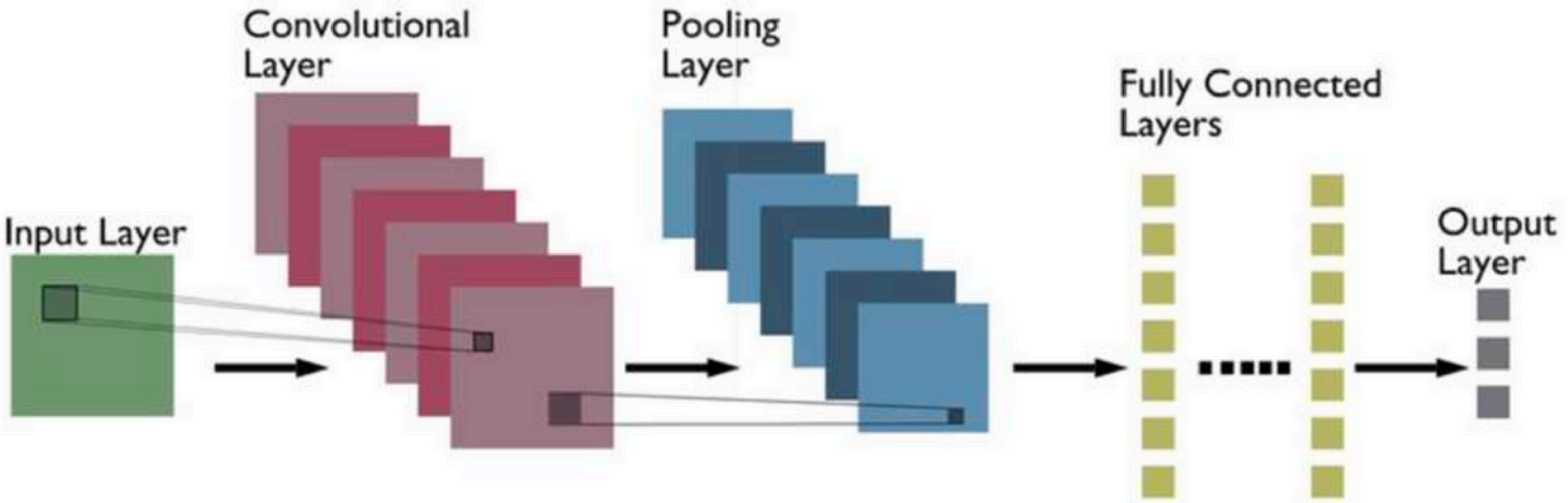
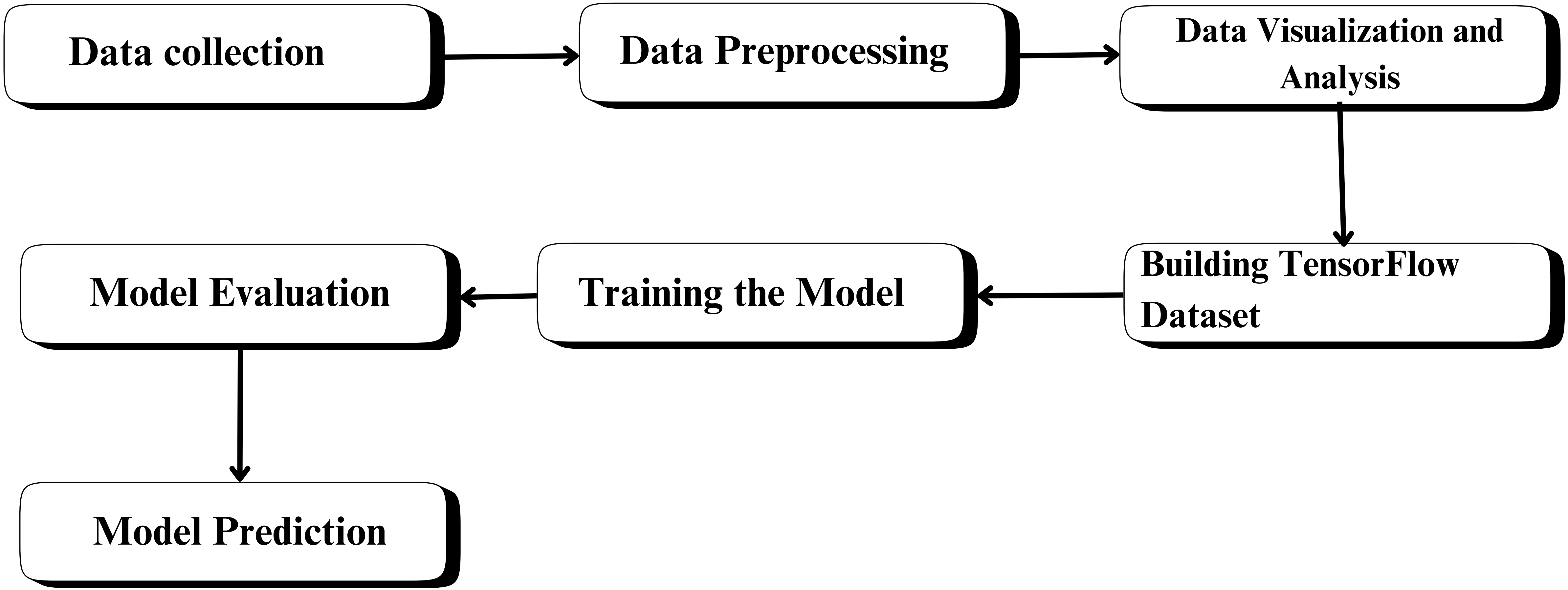



Fig.4 CNN Architecture

## Mel-Frequency Cepstral Coefficients (MFCCs): Key Features for Audio Classification

- **Feature Extraction** – MFCCs convert audio signals into a set of features that represent the spectral properties of sound, making them suitable for speech and audio classification tasks.
- **Mel-Scale Representation** – They use the Mel scale, which mimics human auditory perception by emphasizing lower frequencies and reducing sensitivity to higher frequencies.
- **Cepstral Analysis** – MFCCs transform the frequency domain representation of an audio signal into the cepstral domain, helping to distinguish different sound patterns effectively.
- **Robust in Speech & Audio Recognition** – Widely used in speech recognition, music classification, and bioacoustic monitoring, as they capture timbral characteristics crucial for distinguishing sounds.
- **Feature Set Components** – Typically include the first 12–13 coefficients along with their derivatives (delta and delta-delta) to capture temporal variations and improve classification accuracy.

# Methodology



# Methodology

**1.Input Data Collection:**

- The dataset comprises two main categories of audio recordings:
  - **Capuchinbird calls:** Audio clips containing the vocalizations of Capuchinbirds, recorded in their natural habitat.
  - **Non-Capuchinbird recordings:** Audio clips with other tropical forest sounds, including background noise and calls from other species.

**2.Preprocessing:**

- preprocessing is being done for all the audio files.
  - Spectrogram Conversion:
    - The Short-Time Fourier Transform (STFT) is applied to convert audio signals from the time domain into the frequency domain.

**3.Data Visualization and Analysis :**

The audio files will be initially visualized by plotting their waveforms using Matplotlib to ensure data integrity. This step will confirm that the audio files are correctly loaded and processed.

# Methodology

**4.Building TensorFlow Dataset:**

To build the TensorFlow dataset, the TensorFlow Dataset API will be used to manage the dataset. Initially, the audio files will be divided into two categories: positive samples (Capuchinbird calls) and negative samples (non-Capuchinbird recordings).

**5.Training the Deep Learning Model:**

A Convolutional Neural Network (CNN) will be employed as the model architecture for this study.

**6.Model Evaluation and prediction:**

Accuracy score,precision score will be calculated to see how the model is working.

# Module Specification

## Module 1: Data Collection and Preprocessing

Raw audio files (Capuchin bird calls and non-Capuchin bird sounds)

**input**

→

Data cleaning: Remove any corrupt or incomplete files.

**process**

→

Preprocessed and labeled audio dataset ready for model training.

**output**

## Module 2: Implementation of ANN / DL Algorithm

Preprocessed audio data (spectrograms,

**input**

→

Model Selection, Training

**process**

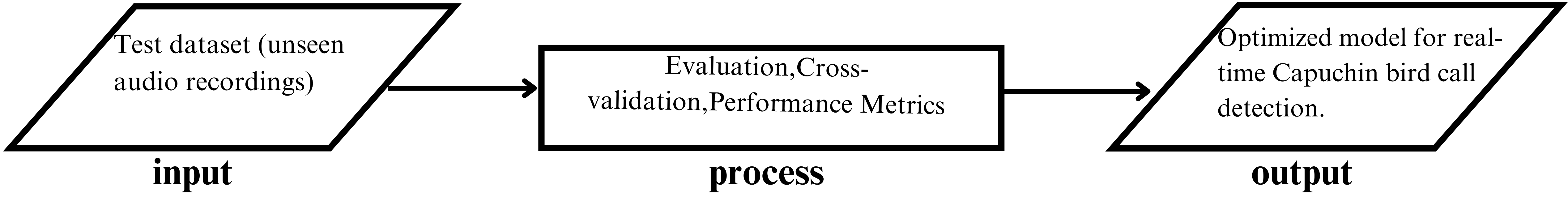
→

Trained model capable of classifying Capuchin bird calls and non-Capuchin sounds.

**output**

# Module Specification

## Module 3: Testing and Validation



# Conclusion

- The audio classification project showcases the effectiveness of deep learning in bioacoustic monitoring, specifically for detecting Capuchinbird calls. Using a CNN trained on MFCCs extracted from audio clips, the model achieved high accuracy (98.1% on training and 96.7% on validation). Preprocessing techniques like feature extraction, data augmentation, and normalization enhanced robustness and reduced overfitting. The integration of ModelCheckpoint in TensorFlow ensured efficient training, making the system a scalable and accurate tool for species identification, significantly improving upon traditional manual methods.

- Beyond classification, this AI-driven approach aids in biodiversity tracking, endangered species monitoring, and ecological research with minimal human intervention. The potential integration of real-time acoustic sensors could enable continuous, non-intrusive observation of avian populations, crucial for conservation efforts. Future advancements, such as real-time data processing and cloud-based deployment, could expand the system's capabilities to monitor a wider range of species and habitats, revolutionizing wildlife conservation.

# THANK YOU