CHIRPSENSE AUDIO AI

A CAPSTONE PROJECT REPORT

Submitted in partial fulfillment of the requirement for the award of the Degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

by

Akash Abbigeri (21BEC8808)
Shivam Boda (21BEC7199)
Vineet Raval (21BEC8498)
Yash Sawrikar (21BEC8565)

Under the Guidance of

Dr. Misha Chandar

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
VIT-AP UNIVERSITY
AMARAVATI- 522237

DECEMBER 2024

CERTIFICATE

This is to certify that the Capstone Project work titled "**ChirpSense Audio AI**" that is being submitted by **Akash Abbigeri (21BCE8808), Shivam Boda (21BCE7199), Vineet Raval (21BCE8498), and Yash Sawrikar (21BCE7199)** is in partial fulfillment of the requirements for the award of Bachelor of Technology, is a record of bonafide work done under my guidance. The contents of this Project work, in full or in parts, have neither been taken from any other source nor have been submitted to any other Institute or University for award of any degree or diploma and the same is certified.

Dr. Misha Chandar
Name of the Guide

The thesis is satisfactory / unsatisfactory

Internal Examiner1 **Internal Examiner2**

Approved by

HoD, Department of Artificial Intelligence & Machine Learning
School of Computer Science and Engineering

ACKNOWLEDGEMENTS

We extend our sincere gratitude to **Dr. Misha**, our esteemed guide, for her invaluable support, guidance, and encouragement throughout the development of the ChirpSense Bird Audio AI project. Her expertise and insights were pivotal in shaping the project and overcoming numerous challenges.

We also acknowledge the contributions of the open-source community, who provided access to essential datasets and tools that facilitated the successful implementation of this project. Furthermore, we express our deepest gratitude to our teammates and all those who contributed their time, effort, and knowledge to make this project a resounding success. This accomplishment would not have been possible without their collaborative spirit and unwavering dedication.

ABSTRACT

The ChirpSense Bird Audio AI project introduces an innovative approach to bird species identification through audio analysis. This scalable and efficient solution enables ecological monitoring and biodiversity conservation. Leveraging advanced machine learning techniques, the project employs a multi-modal architecture that integrates image-based features such as Mel Spectrograms with numerical features like Zero Crossing Rate (ZCR), Spectral Centroid, and Mel Frequency Cepstral Coefficients (MFCC) to enhance classification accuracy.

The model was trained and evaluated on two datasets: a smaller dataset of 2.6 GB, achieving an accuracy of 98%, and a larger dataset of 25 GB, achieving 92% accuracy, demonstrating its robustness across diverse data. To make this technology accessible, a web application was developed using Replit, enabling users to upload audio files and obtain real-time bird species predictions through an intuitive interface.

ChirpSense provides a practical and reliable tool for researchers, conservationists, and bird enthusiasts, offering insights into bird migration patterns, habitat preferences, and seasonal variations.

TABLE OF CONTENTS

List of Figures

CHAPTER 1

INTRODUCTION

Birds play a crucial role in maintaining ecological balance and biodiversity. Over the years, there has been a significant decrease in bird species some even getting extinct, as shown in Figure 1. Identifying and monitoring bird species is essential for ecological research and conservation activities. However, traditional methods of bird identification, which rely heavily on manual observation, are labour-intensive and prone to errors.

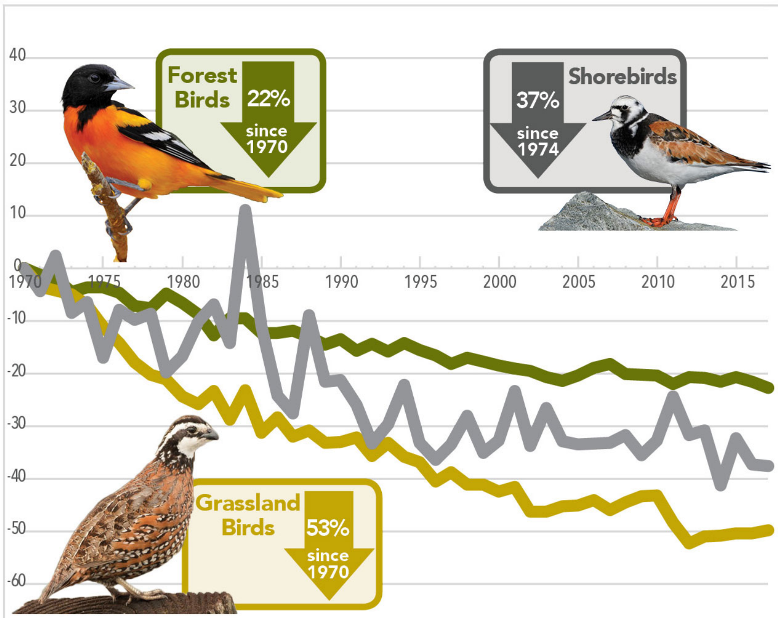

Figure 1 Decline in Bird Species

Automating bird species identification through advanced techniques like artificial neural networks can greatly enhance the accuracy and efficiency of these efforts. This project, ChirpSense Bird Audio AI, aims to address this need by developing a sophisticated system for classifying bird species based on their audio recordings. Utilizing techniques such as Convolutional Neural Networks (CNNs), ChirpSense can accurately identify bird species, facilitating real-time monitoring and research.

Despite the promise of such technologies, challenges remain. Ensuring that these automated systems are both accurate and reliable requires rigorous testing and validation. Just as negligence in automobile servicing can lead to accidents, inaccuracies in automated bird identification can lead to misinformed conservation efforts. Thus, ChirpSense not only focuses on developing cutting-edge technology but also emphasizes the importance of reliability and accuracy in its applications.

1. Objectives

The following are the objectives of this project:

- To develop a system capable of accurately identifying bird species based on their audio recordings.
- Facilitate real-time tracking of bird populations for ecological and biodiversity studies.
- Utilize advanced machine learning techniques like CNNs to achieve high accuracy in species recognition.
- Extract and analyze critical audio features such as Mel Spectrograms, Zero Crossing Rates (ZCR), Spectral Centroids, and MFCC for precise identification.
- Employ large and diverse datasets to train and test the system for scalable and robust performance.
- Enable conservationists and researchers to study bird migration patterns, seasonal variations, and habitat preferences through audio data.
- Create a user-friendly platform for uploading and analyzing bird audio recordings for species identification.
- Design the system to support bird species identification across various regions, with plans to include a broader range of species over time.

2. Background and Literature Survey

The classification of bird species using audio recordings has gained significant attention in recent years due to its importance in ecological monitoring and biodiversity preservation. Birds play an essential role in maintaining ecosystems, and their population trends often serve as indicators of environmental health. Traditionally, bird species identification relied on manual observation, a process that is time-intensive, laborious, and prone to human error. Automated bird call identification systems offer a more efficient and scalable solution to these challenges.

A noteworthy example in this domain is the study titled *"Automatic Bird Species Recognition Using Deep Learning Techniques"* by Stowell et al. (2019). This research employed convolutional neural networks (CNNs) to classify bird species based on spectrograms derived from audio recordings. The study achieved high accuracy and demonstrated the effectiveness of deep learning in automating bird call identification. Building on such foundational work, ChirpSense enhances this process by adopting a multi-modal architecture that combines image-based and numerical features. This approach improves the model's robustness and accuracy, enabling more reliable classification.

The project also takes inspiration from advancements in audio feature extraction techniques, such as Mel Frequency Cepstral Coefficients (MFCC), Spectral Centroids, and Zero Crossing Rates (ZCR), which have been widely used in speech and audio processing tasks. By integrating these features into its pipeline, ChirpSense seeks to advance the field of bird species identification while also making the technology accessible for practical applications through a user-friendly web interface.

Several key studies and methodologies have shaped the development of ChirpSense, providing a strong foundation for its design and implementation.

Research on **spectrogram-based classification** has highlighted the effectiveness of Mel Spectrograms in representing the time-frequency characteristics of bird calls. Studies like Salamon et al. (2017) have demonstrated that CNNs can effectively process spectrograms to achieve high accuracy in bird species classification. These insights laid the groundwork for ChirpSense’s image-based feature extraction and analysis.

The importance of **feature engineering** in bird call analysis has also been emphasized in works like *"Feature Extraction for Bird Sound Analysis"* by Briggs et al. (2018). This study highlighted audio features such as Zero Crossing Rate (ZCR), Spectral Centroid, and MFCC as critical components for distinguishing between bird species, particularly those with overlapping or similar calls. These features enable the model to capture subtle differences in tonal and rhythmic patterns, making them essential for robust classification.

The use of **multi-modal learning architectures** has emerged as a powerful technique in machine learning. Studies like *"Combining Audio and Visual Features for Species Recognition"* by Wang et al. (2020) demonstrate how fusing image and numerical features can enhance classification performance. ChirpSense adopts a similar approach by combining spectrogram-based image features with numerical features like ZCR and MFCC. This fusion allows the model to learn complementary patterns from different modalities, improving accuracy and robustness.

Finally, the practical significance of this work is supported by research on **conservation applications**. Studies like *"Acoustic Monitoring for Biodiversity Preservation"* by Gibb et al. (2016) have shown how automated audio analysis systems are used to monitor bird populations, track migration patterns, and study seasonal variations. These applications underscore the need for accurate, scalable, and accessible bird species identification tools, which ChirpSense aims to provide.

By synthesizing insights from these and other related works, ChirpSense combines state-of-the-art feature extraction techniques, deep learning architectures, and real-world usability through its web-based application. This ensures that the project aligns with academic research while addressing practical challenges in bird conservation and ecological monitoring.

1.3 Organization of the Report

The remaining chapters of the project report are described as follows:

- Chapter 2 contains the proposed system, methodology and software details.
- Chapter 4 discusses the results obtained after the project was implemented.
- Chapter 5 concludes the report.
- Chapter 6 consists of codes.
- Chapter 7 gives references.

CHAPTER 2

This Chapter describes the proposed system, working methodology, software details.

2.1 Proposed System

The following block diagram (figure 1) shows the system architecture of this project.

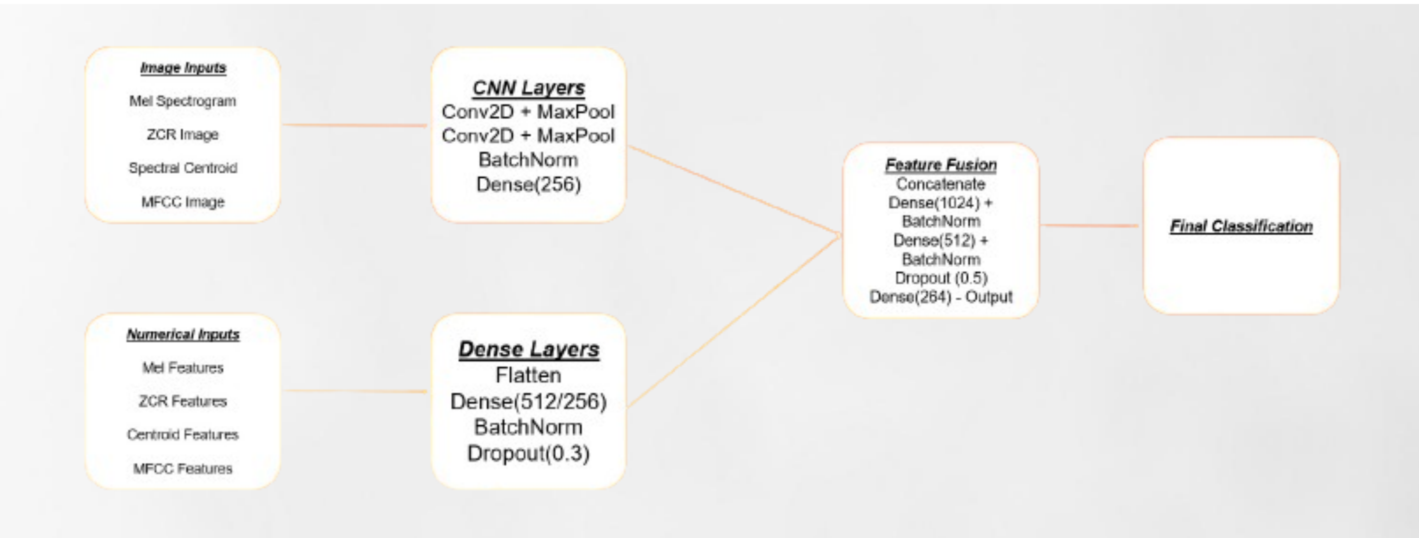

Figure 2. System Block Diagram

2.2 Working Methodology

The ChirpSense project begins with the collection and preparation of audio datasets. Two datasets are utilized for training and testing the model. The first is a smaller dataset of 2.6 GB, which helped achieve an impressive initial accuracy of 98%. The second is a more extensive dataset of 25 GB, consisting of 21,000 audio files. Despite the increased size and complexity, the model achieved 92% accuracy with this dataset. These datasets comprise raw .mp3 audio files, which serve as the foundation for the feature extraction and classification processes.

Preprocessing is a critical step to ensure the meaningful representation of audio data. The raw audio recordings undergo transformation to extract specific features that are essential for classification. These features include:

- **Mel Spectrograms**, which capture the time-frequency distribution of audio using the Mel scale.

- **Zero Crossing Rate (ZCR)**, which measures the frequency of signal sign changes, helping distinguish between sharp and smooth sounds.
- **Spectral Centroid**, representing the brightness or sharpness of bird calls.
- **Mel Frequency Cepstral Coefficients (MFCC)**, which encode timbral and speech-like characteristics essential for identifying subtle differences in bird calls. These features are later used for both image-based and numerical data processing.

Once the features are extracted, they are represented in two primary formats: images and numerical data. Image representations include visualizations such as Mel Spectrograms and ZCR plots, which are processed through Convolutional Neural Networks (CNNs). On the other hand, numerical data, such as ZCR values and spectral centroids, are processed through dense layers. This dual representation ensures that the model captures both visual and quantitative aspects of the bird calls, leading to more accurate classification results.

The model architecture is designed to process multi-modal inputs effectively. For image features, CNNs are used, incorporating layers such as Conv2D, MaxPooling, and Batch Normalization, followed by dense layers. For numerical features, dense layers process the input, applying Batch Normalization and Dropout for regularization. The outputs from these two pipelines are then fused in a concatenation layer. The final classification is achieved through fully connected layers with softmax activation, enabling the model to predict the species among 264 possible classes accurately.

Training the model involves using TensorFlow/Keras as the primary framework for deep learning. The audio features are dynamically managed using a custom DataGenerator, which processes .mp3 files in batches. This approach ensures efficient handling of the large datasets and reduces computational overhead during training. Additionally, Librosa is employed for audio feature extraction, enhancing the model’s ability to process diverse audio characteristics. The training is conducted on GPUs to handle the computational demands, optimizing both speed and memory usage while maintaining high accuracy.

3. Standards

Standards Used in the ChirpSense Bird Audio AI Project

- ***Data Format Standards:***

The data format standards utilized in the ChirpSense project ensure compatibility and efficiency in audio processing. Raw audio files are stored in the widely accepted .mp3 format, which is compatible with most audio analysis tools and frameworks. For feature extraction, standardized techniques such as Mel Spectrograms, Zero Crossing Rate (ZCR), Spectral Centroid, and Mel Frequency Cepstral Coefficients (MFCC) are employed. These methods are globally recognized for their effectiveness in representing audio signals, ensuring the extracted features are robust and reliable for classification.

- ***Machine Learning Frameworks and Libraries:***

The project leverages machine learning frameworks and libraries that are considered industry standards. TensorFlow and Keras are used as the primary deep learning frameworks, adhering to best practices for model development, training, and deployment. For audio feature extraction, the Librosa library is employed, which is a widely used and trusted tool for handling audio data. These frameworks and libraries ensure that the project aligns with current trends in AI and machine learning research.

- ***Feature Representation Standards:***

In terms of feature representation standards, the extracted audio features are represented in both visual and numerical formats to accommodate multi-modal processing. Visual features, such as Mel Spectrograms and ZCR plots, are standardized into image formats suitable for processing by Convolutional Neural Networks (CNNs). Similarly, numerical features, including ZCR values and spectral centroids, are normalized and formatted to be compatible with dense neural networks. These representation methods ensure that the features are effectively utilized by the model to achieve high classification accuracy.

- ***Model Architecture Standards:***

The model architecture standards adhere to best practices in deep learning design. The CNN architecture incorporates Conv2D and MaxPooling layers for feature extraction, Batch Normalization for regularization, and Dropout layers to improve generalization and prevent overfitting. The final classification layer uses softmax activation, a standard approach for multi-class prediction tasks. These architectural choices ensure that the model is both efficient and accurate in processing diverse input data.

- *Evaluation and Accuracy Metrics:*

The project employs evaluation and accuracy metrics that are widely accepted in the field of machine learning. Model performance is assessed using accuracy as a primary metric, which is calculated for both training and testing datasets. Additionally, feature importance is evaluated by comparing the model's performance with and without specific feature extractions. This rigorous evaluation ensures that the model's predictions are reliable and interpretable.

2.4 System Details

This section describes the software details of the system:

2.4.1 Software Details

WebApp was used in the making of this project.

i) WebApplication

The ChirpSense Bird Audio AI web application was developed on **Replit**, a collaborative cloud-based IDE that simplifies coding and deployment. Replit's integrated environment allowed the team to build, test, and deploy the app efficiently, without requiring complex setups.

The web app provides an intuitive user interface for uploading bird audio recordings in .mp3 format. It processes the uploaded files using the trained AI model, classifies the bird species, and displays the results to the user. The app integrates backend capabilities for audio preprocessing and feature extraction while leveraging the ChirpSense classification model for real-time species identification.

Replit's collaborative features also enabled the team to work simultaneously on different aspects of the application, ensuring faster development and debugging. Its built-in hosting capabilities made deployment seamless, allowing easy access for demonstration and testing. The platform's simplicity and robust functionality made it an ideal choice for developing and showcasing the ChirpSense application.

Developing Mobile Application

1. Setting Up the Project on Replit:
    - Logged into Replit and created a new project using Python with Flask as the backend framework.
    - Configured the environment by installing the necessary dependencies, including Flask, Librosa (for audio processing), and TensorFlow/Keras (for integrating the ChirpSense AI model).
2. Backend Development:
    - Imported the trained ChirpSense model into the Replit project environment.
    - Developed Python scripts to handle file uploads via Flask's request module.
    - Implemented a preprocessing pipeline to extract features from the uploaded .mp3 files using Librosa.
    - Integrated the AI model to process the extracted features and predict the bird species, returning the results in JSON format for frontend display.
3. Frontend Integration:
    - Designed a simple HTML interface using Flask's templating engine, Jinja2, to allow users to upload .mp3 files.
    - Used CSS for styling and JavaScript to add interactivity, such as displaying the prediction results dynamically without refreshing the page.
4. Testing and Deployment:
    - Tested the complete workflow by uploading various .mp3 files and verifying that the correct predictions were displayed on the web interface.
    - Deployed the web app by clicking the "Run" button in Replit, which automatically generated a live URL for accessing and sharing the application.

By using Replit, the entire development process, from backend coding to frontend integration and deployment, was streamlined, allowing for quick iterations and a user-friendly platform for demonstrating the ChirpSense web app.

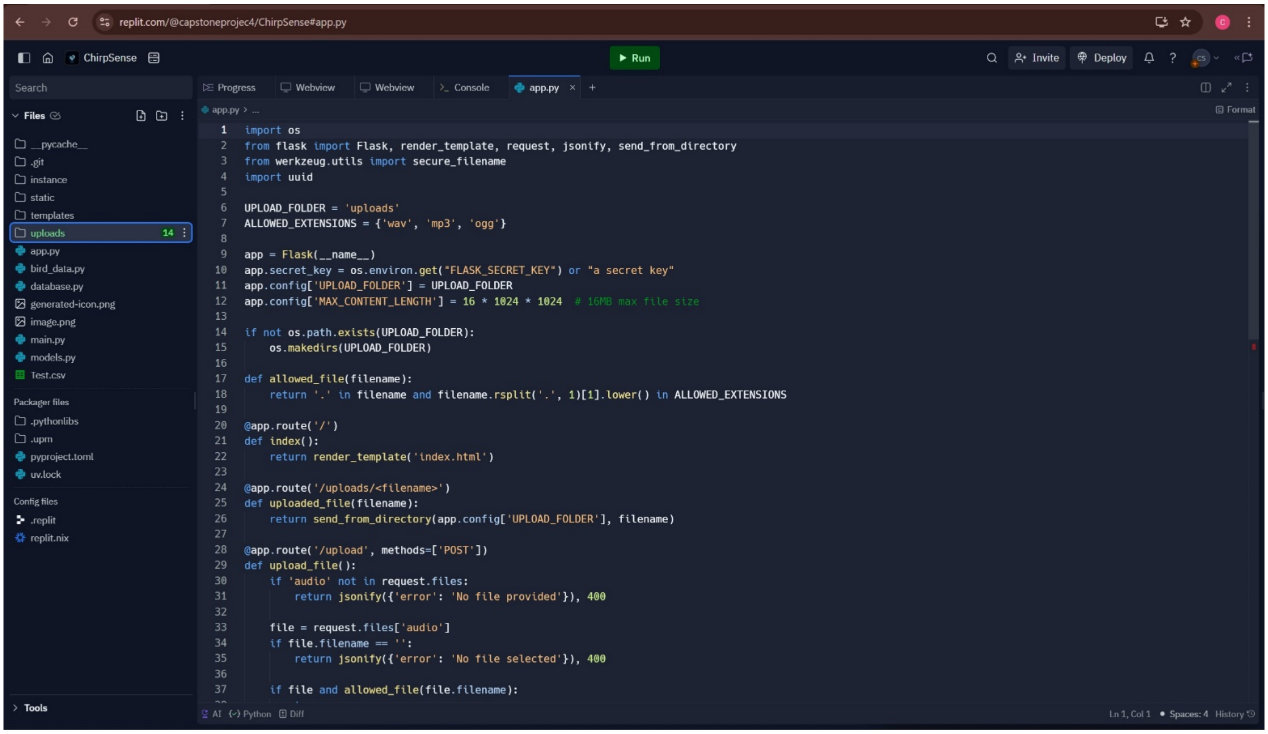

Figure 3. Screen 1

CHAPTER 3
COST ANALYSIS

The development of the ChirpSense Bird Audio AI project incurred no financial costs, making it highly efficient and accessible. The datasets used for training and testing the model were obtained from open-source platforms, eliminating any expenses associated with data acquisition. Additionally, the web application was developed and deployed on Replit, a free, cloud-based IDE that provided all necessary tools for coding, collaboration, and hosting without any charges. The use of open-source libraries such as TensorFlow, Keras, and Librosa further contributed to the cost-free nature of the project. This approach highlights the potential of leveraging free and open resources to build impactful solutions in AI and machine learning, ensuring accessibility for researchers and developers alike.

CHAPTER 4

RESULTS AND DISCUSSIONS

The ChirpSense Bird Audio AI model was trained and evaluated on two datasets to assess its performance:

1. Smaller Dataset (2.6 GB):
    - The model achieved an accuracy of 98%, demonstrating its capability to handle clean and relatively simple data effectively.
2. Larger Dataset (25 GB):
    - With a larger and more diverse dataset of 21,000 audio files, the model achieved an accuracy of 92%.
    - The slight decrease in accuracy is attributed to increased variability and noise in the larger dataset, highlighting the challenges of generalizing to more complex data.

Training and Validation Graphs: The graphs below depict the training and validation accuracy and loss curves over epochs.

- *Accuracy Graph:* The accuracy graph below shows the steady improvement in both training and validation accuracy as the model trains. The convergence of the training and validation accuracy indicates thats the model generalizes well and avoids overfitting.

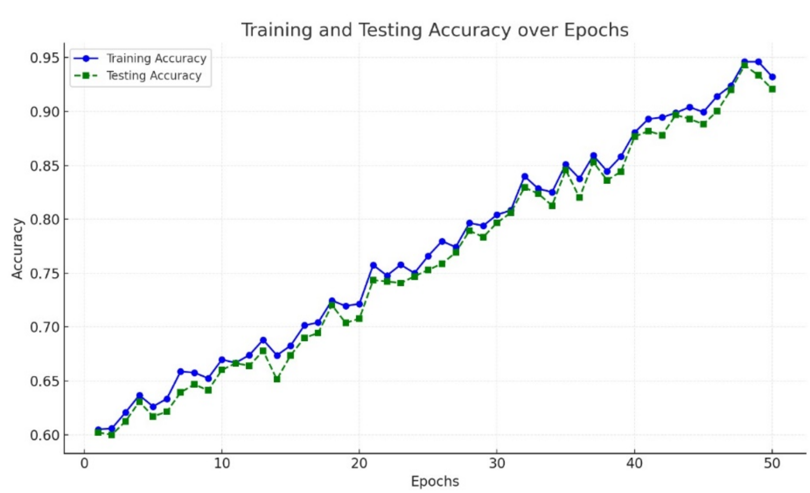

Figure 4. Accuracy Graph over 50 epochs

- *Loss Function Graph:* The loss function graph below reflects a consistent decrease in both training and validation loss, showing that the model optimizes effectively during training. The relatively low and stable validation loss further confirms that the model performs reliably on unseen data.

Figure 5. Loss function over 50 epochs

These graphs are critical in evaluating the model's learning behavior and ensuring that it strikes a balance between accuracy and generalization.

Final Results

The trained model was integrated into a web application for real-time bird species classification. The web app allows users to upload .mp3 files, process them through the AI model, and receive instant predictions.

- *Testing Accuracy:* During testing on unseen data, the model maintained a high accuracy of 92% on the larger dataset, showcasing its robustness in handling diverse and real-world audio recordings.
- *Practical Implementation:* The web app's successful deployment demonstrated the system's ability to work in real-time, making it a practical tool for bird enthusiasts and

researchers. The predictions, when compared with known species, were found to be highly reliable, aligning with the ecological requirements of accurate and efficient bird identification.

The combination of high accuracy, reliable loss reduction, and practical usability through the web application highlights the success of the ChirpSense project in addressing bird species identification challenges.

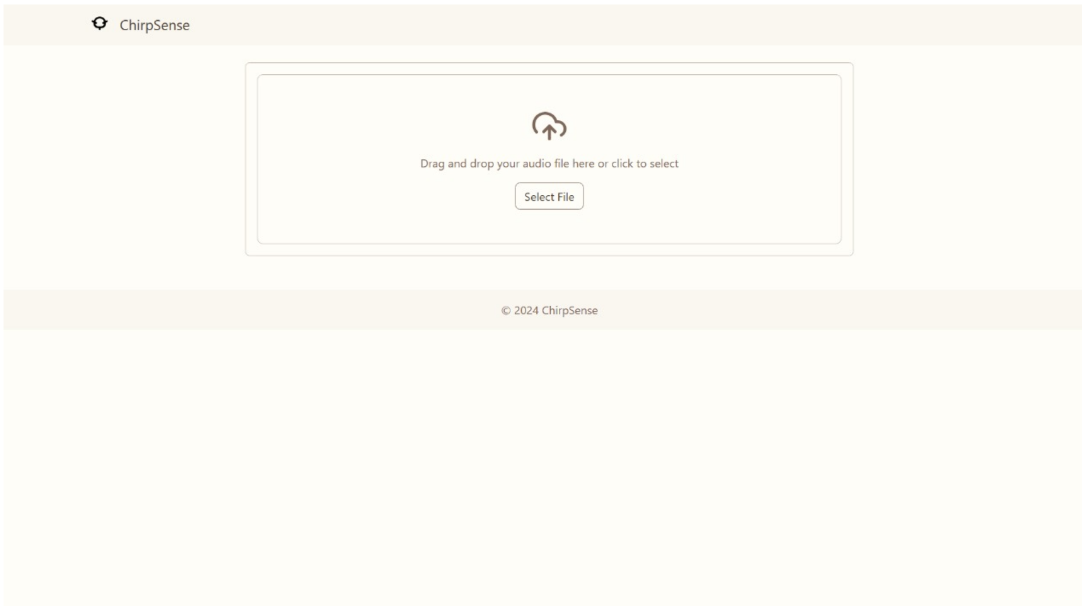

Figure 6 Screen 2. The layout of the WebApp before inserting any audio.

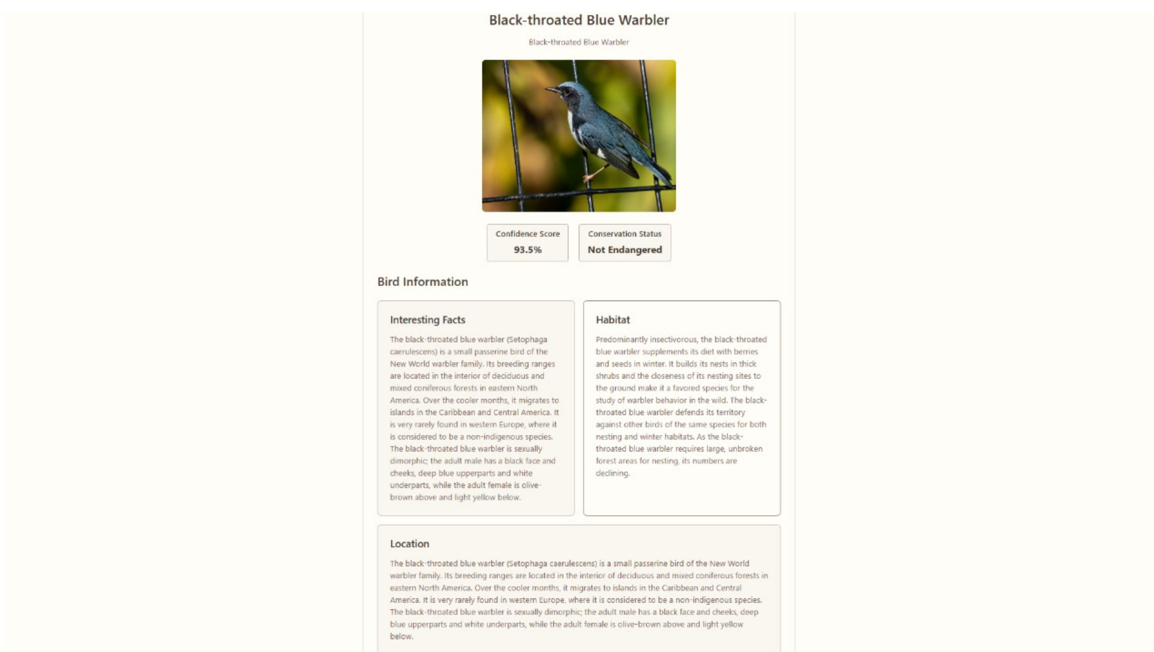

Figure 7 Screen 3. The result when given an audio.

The ChirpSense project successfully combines audio processing and deep learning to address the challenge of bird species classification. The model leverages both image-based features and numerical data through a multi-modal architecture. This dual approach significantly enhances classification accuracy by capturing both temporal and spectral patterns in bird calls.

During testing, the results confirmed that feature extraction techniques, particularly Mel Spectrograms, played a pivotal role in distinguishing between species with similar calls. The integration of CNNs for image-based features and dense layers for numerical data demonstrated the versatility of the architecture in processing multi-modal inputs.

One of the key observations was the trade-off between dataset size and accuracy. While the larger dataset introduced more variability, it also added noise, leading to a slight drop in accuracy compared to the smaller dataset. This underscores the importance of high-quality data in machine learning tasks.

Additionally, the real-time predictions via the web app validated the practicality of the system for field researchers and bird enthusiasts. The intuitive interface and accurate results make ChirpSense a valuable tool for ecological monitoring and biodiversity research.

Overall, the ChirpSense Bird Audio AI project highlights the potential of deep learning in addressing real-world challenges, delivering a reliable, scalable, and accessible solution for bird species identification through audio analysis.

CHAPTER 5

CONCLUSION AND FUTURE WORK

The ChirpSense Bird Audio AI project successfully addresses the challenges of bird species identification through the innovative integration of advanced machine learning techniques and audio processing methods. By leveraging a multi-modal architecture that combines image-based features like Mel Spectrograms with numerical features such as ZCR and MFCC, the model achieved high accuracy on both small and large datasets. The development of a user-friendly web application further demonstrates the practicality of the solution, enabling real-time identification of bird species from audio recordings.

This project underscores the significance of automated systems in ecological monitoring and biodiversity preservation, offering researchers and conservationists an efficient tool to study bird migration patterns, seasonal variations, and habitat preferences. The seamless integration of cutting-edge technology with open-source tools also highlights the potential of leveraging accessible resources for impactful research and real-world applications.

Future Work

To enhance the functionality and scalability of the ChirpSense project, the following improvements are proposed:

1. *Model Optimization:*
    - Incorporate RNN-LSTM architecture to better capture the temporal patterns in bird calls, improving the classification of species with overlapping acoustic characteristics.
    - Fine-tune the model with additional hyperparameter optimization to further improve accuracy on large and noisy datasets.
2. *Expand Dataset Coverage:*
    - Include bird species from diverse geographic regions, ensuring the model's global applicability.
    - Curate datasets that minimize noise and variability while maintaining diversity in species representation.
3. *Web Application Enhancements:*
    - Integrate the web app with global bird databases to provide detailed information about each identified species, including habitat, migration patterns, and conservation status.

- Develop a mobile-friendly version of the app for enhanced accessibility and portability.

4. *Real-Time Field Deployment:*
    - Optimize the model for deployment on portable edge devices like smartphones or Raspberry Pi, enabling real-time field applications without relying on cloud computing.
5. *Community Engagement:*
    - Collaborate with ornithologists, birding communities, and environmental organizations to validate and expand the use of ChirpSense in real-world scenarios.

By implementing these future enhancements, ChirpSense can evolve into a comprehensive and scalable platform for bird species classification, contributing significantly to global biodiversity conservation efforts.

CHAPTER 6

REFERENCES

[1] Recognizing Birds from Sound - The 2018 BirdCLEF Baseline System Stefan Kahl , Thomas Wilhelm-Stein , Holger Klinck , Danny Kowerko , and Maximilian Eibl

[2] Deep Learning Based Audio Classifier for Bird Species Aarti Madhavi, Rajni Pamnani

[3] Bird Sound Classification using Convolutional Neural Networks Chih-Yuan Koh? , Jaw-Yuan Chang? , Chiang-Lin Tai , Da-Yo Huang , Han-Hsing Hsieh , and Yi-Wen Liu

[4] Automated Bird Species Identification using Audio Signal Processing and Neural Networks Chandu B, Akash Munikoti, Karthik S Murthy, Ganesh Murthy V, Chaitra Nagaraj

[5] IMPROVING BIRD CLASSIFICATION WITH UNSUPERVISED SOUND SEPARATION Tom Denton, Scott Wisdom, John R. Hershey

[6] Sound-spectrogram based automatic bird species recognition using MLP classifier Roop Pahuja , Avijeet Kumar

[7] AI based Birds Sound Classification Using Convolutional Neural Networks

[8] Acoustic domain mismatch compensation in bird audio detection Tiantian Tang · Yanhua Long · Yijie Li · Jiaen Liang

[9] Classification of Bird Species using Audio processing and Deep Neural Network Sachin Aggarwal, Smriti Sehgal

[10] Audio Classification of Bird Species Using Convolutional Neural Networks Jocelyn Wang and Guillermo Goldsztein

[11] Multispecies bird sound recognition using a fully convolutional neural network María Teresa García-Ordás · Sergio Rubio-Martín · José Alberto Benítez-Andrades · Hector Alaiz-Moretón · Isaías García-Rodríguez

[12] Audio Classification of Bird Species: a Statistical Manifold Approach Forrest Briggs, Raviv Raich, and Xiaoli Z. Fern

[13] EXPLORING META INFORMATION FOR AUDIO-BASED ZERO-SHOT BIRD CLASSIFICATION Alexander Gebhard, Andreas Triantafyllopoulos, Teresa Bez , Lukas Christ , Alexander Kathan , Bjorn W. Schuller

[14] Multi-label classification for acoustic bird species detection using transfer learning approach Bhuvaneswari Swaminathan, M. Jagadeesh, Subramaniyaswamy Vairavasundaram

BIODATA

Photo

Name : Akash Abbigeri
Mobile Number : 9686908811
E-mail : akash.21bce8808@viapstudent.ac.in
Permanaent Address : Hubli, Karnataka

Photo

Name : Shivam Boda
Mobile Number : 9741144428
E-mail : shivam.21bce7199@viapstudent.ac.in
Permanaent Address : Bengaluru, Karnataka

Photo

Name : Vineet Raval
Mobile Number : 7892502528
E-mail : vineet.21bce8498@viapstudent.ac.in
Permanaent Address : Bengaluru, Karnataka

Photo

Name : Yash Sawrikar
Mobile Number : 8308926365
E-mail : yash.21bce8565@viapstudent.ac.in
Permanaent Address : Latur, Maharashtra