IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

AVI-AUDIO NET

Pavan H B*1, Sai Sujan.S*2, Vijay Surya Reddy VV*3, Kiran Kumar.R*4, Mrs Rekha V*5

*1,2,3,4Under Graduate Student, Dept. Of Computer Science & Engineering, Jyothy Institute Of Technology Visvesaraya Technological University, Belagavi Bengaluru-560082, India

*5Assistant Professor, Dept. Of Computer Science & Engineering, Jyothy Institute Of Technology Visvesaraya Technological University, Belagavi Bengaluru-560082, India.

ABSTRACT

The project focuses on developing an automated bird species identification system using audio signal processing and machine learning techniques. By harnessing a dataset of bird vocalizations, the project aims to train a model capable of accurately classifying bird species based on their vocalizations. This involves extracting relevant features from audio recordings, such as Mel-Frequency Cepstral Coefficients (MFCCs), to represent the unique characteristics of each bird species' vocalizations. These features are then used to train an Artificial Neural Network (ANN) model, enabling it to learn and recognize patterns in the audio data. Once trained, the model is integrated into a web application interface, allowing users to upload audio recordings and receive predictions of the bird species present. This system provides a practical tool for researchers and conservationists to quickly and accurately identify bird species in the field, aiding in ecological research and conservation efforts.

I. INTRODUCTION

In the realm of ecological exploration, the endeavor to automate bird species identification stands as a captivating fusion of technology and nature. The intricate calls of birds, forming a vibrant tapestry of sound, serve as the cornerstone for unlocking their identities. Through sophisticated algorithms and machine learning techniques, this study endeavors to bridge the gap between the intricate nuances of avian vocalizations and the computational power of neural networks. By harnessing the capabilities of audio signal processing, we aim to create a seamless system capable of discerning and categorizing bird species based on their distinct vocal signatures.

By leveraging the innate capabilities of neural networks, this project seeks to revolutionize traditional methods of birdwatching and ecological monitoring. In envisioning a future where the melodies of birds not only captivate our senses but also offer profound insights into biodiversity, we aspire to propel the field of avian research into new realms of discovery. Through the amalgamation of cutting-edge technology and the innate beauty of avian vocalizations, this research endeavors to bring us closer to a world where the natural symphony of bird calls serves as a gateway to deeper ecological understanding.

To address this need, the project proposes a solution that utilizes audio recordings of bird vocalizations to train a machine learning model. By extracting relevant features from the audio data and employing machine learning techniques, such as Artificial Neural Networks (ANNs), the system aims to learn patterns in the vocalizations of different bird species. Once trained, the model can accurately classify new audio recordings and identify the corresponding bird species.

The development of such a system has the potential to revolutionize bird species identification, offering a more efficient and reliable alternative to traditional methods. By automating the identification process, researchers and conservationists can gather data more quickly and accurately, leading to better insights into bird populations and habitats. Additionally, the system can be deployed in various environments, including remote areas where visual identification may be challenging or impractical.

This project addresses a critical need within the realms of ecological research and conservation by presenting a pioneering approach to bird species identification. In contrast to traditional methods that heavily rely on visual cues, which may be limited or obscured in certain environments, our approach harnesses the rich auditory landscape of avian communication. By harnessing the power of audio data and machine learning, we can transcend the constraints of visual identification and gain unprecedented insights into avian biodiversity.

In summary, this project represents a transformative leap forward in the field of bird species identification. Through the seamless integration of advanced audio signal processing techniques and the computational prowess of neural networks, we endeavor to create a robust and efficient system that not only enhances our understanding of avian biodiversity but also contributes to broader ecological research and conservation initiatives. As we embark on this journey, we are poised to unlock new frontiers of discovery and deepen our appreciation for the intricate melodies of the natural world.

II. LITERATURE SURVEY

· The Light Weight CNN (LW-CNN) architecture enhances crowd counting accuracy and processing efficiency in public spaces during COVID-19 by providing a point estimate for crowd size, addressing uncertainty. Trained on various scenarios, including full and partial head vision, LW-CNN surpasses pre-trained models in classifying partial head counts, improving accuracy and reducing processing time. Its applicability extends to events, crisis management, and workplace safety. [1]

· The paper introduces Extreme Learning Machine (ELM) techniques to improve feed-forward neural network performance, overcoming slow computation and enhancing gain. Through random enhancement of network components, ELM surpasses traditional methods in accuracy. It proposes a novel feed-forward technique, elucidates ELM variations for diverse applications, and discusses potential future advancements in function approximation. [2]

· In this article, a two-stage automated bird species identification system is introduced. To create spectrograms, the initial step is building the perfect dataset and using wise pre-processing methods. A Convolutional Neural Network (CNN) classifies the spectrograms in real-time for the purpose of identifying bird species in the second stage. [3]

· The study proposes an Animal Species Recognition System utilizing an Artificial Neural Network (ANN) in Matlab for bird sound identification. It entails creating a graphical user interface (GUI), training the ANN for species identification, and acquiring power spectral density data for each bird type. The successful deployment of this system offers a tool for researching climate change effects and endangered animal populations. [4]

· The research introduces an Automated Bird Detection (ABD) system employing Gaussian Mixture Model (GMM) classification and Dual-tree M-band Wavelet transform (DMWT) for feature extraction. Divided into three phases, the system includes DMWT feature extraction, GMM-based classification of bird sounds, and preprocessing using explosion and pre-emphasis filters. It evaluates various classification techniques like neural networks, support vector machines, and extreme learning machines. [5]

· The goal of the research is to address challenges to bird populations and to facilitate bird species identification automatically by utilizing neural networks and sound processing. With CNN (AlexNet) and transfer learning, pre-emphasis is used in this work to obtain 91% accuracy in real-time bird species categorization. A Graphical User Interface (GUI) that is easy to use improves practical applicability. [6]

· The literature review investigates the use of neural networks and signal processing for automated bird species identification. While other research include a variety of methodologies, such as deep convolutional neural networks, audio-visual classification, and feature learning algorithms, illustrating the constantly developing landscape of bird species identification techniques, the main paper obtained 80% accuracy on 138 bird species. [7]

· The research highlights the advantages of machine learning in streamlining the classification process and focuses on employing Convolutional Neural Networks (CNNs) for bird species identification. With an accuracy range of 87%–92%, the research makes recommendations for future developments, such as a mobile application for real-time monitoring, and suggests uses in wildlife monitoring. [8]

· In the study, a two-stage automated method for identifying bird species is introduced. It uses Convolutional Neural Networks (CNN) and reports real-time accuracy of 91%. Although offering benefits including increased scalability and efficiency, drawbacks include inconsistent accuracy rates, dataset restrictions, noise sensitivity, and difficulties managing variability in bird vocalization. [9]

· In order to increase efficiency and scalability, the research article suggests a two-stage method for automated bird species identification utilizing convolutional neural networks (CNNs) and acoustic signal processing. Although the study acknowledges several limits, such as accuracy issues and susceptibility to environmental noise, it highlights the importance of automated systems for ecological research and conservation initiatives. [10]

· The research utilizes Convolutional Neural Networks (CNNs) for swift bird species identification via sound analysis, offering advantages over traditional methods. Despite challenges like background noise, automated systems prove invaluable in environmental research, conservation, and birdwatching, showcasing technology's agility in species identification. [11]

Papers	**Title**	**Authors**	**Year Of Publication**	**Proposed System**	**Drawback's**
[1]	Speedy Image Crowd Counting by Lightweight Convolutional Neural Network	Vivekanandam, B.	2021	Integration of Lightweight CNN (LWCNN) for crowd counting, emphasizing accuracy and efficiency. Trained on various scenes, it categorizes partial head vision, providing higher accuracy during COVID-19.	LWCNN's drawback is sensitivity to diverse crowd conditions, affecting robustness. Optimization is required for consistent performance across dynamic scenarios.
[2]	Study of Variants of Extreme Learning Machine (ELM) Brands and Its Performance Measure on Classification Algorithm	Manoharan, J.Samuel	2021	Investigates variants of ELM for improved neural network performance. Addresses challenges in feed-forward neural networks, proposing ELM algorithms for faster learning and reduced computation time. Future extensions for various applications are discussed.	ELM's drawbacks include sensitivity to parameter tuning, requiring careful optimization, and potential challenges in handling complex data distributions, limiting its applicability in diverse classification tasks.

[3]	Automated Bird Species Identification Using Audio	Chandu B, A. M	2020	Two-stage process involving dataset creation and sound preprocessing. Utilizes a Convolutional Neural Network (CNN) to classify bird species based on spectrograms. Real-time implementation is executed for practical application.	Drawbacks include challenges in handling diverse bird vocalizations, potential limitations in accuracy due to environmental noise, and the need for extensive and well-curated datasets for robust performance.
[4]	Bird Sound Identification based on Artificial Neural Network optimization	M. M. Sukri, U. Fadlilah, S. Saon, A. K. Mahamad, M. M. Som, and A. Sidek	2020	Proposes bird sound identification using Artificial Neural Network (ANN) and Matlab. Trains ANN on power spectral density data for accurate bird species recognition with a graphical user interface (GUI).	This bird sound identification system using Artificial Neural Network (ANN) may face challenges in real-world noise conditions, limits identification to one species at a time, and relies on a graphical user interface.
[5]	Automated Bird Detection in Audio Recordings by a Signal Processing Perspective	Raja Shekar Kadurka & Harish Kanakalla	2021	Introduces an Automated Bird Detection (ABD) system using Dual-tree M- band Wavelet transform (DMWT) for feature extraction and Gaussian Mixture Model (GMM) for classification. Promising results in automated bird sound detection are demonstrated.	The ABD system, despite its promising results, may encounter challenges in handling diverse bird vocalizations, and the effectiveness could be influenced by environmental noise, potentially impacting classification accuracy.
[6]	Automated Audio Signal Processing for Bird Species Identification Including Neural Networks	Mohammed Sadiq B1, Pooja H M2	2023	Explores automatic bird species identification using sound processing and a convolutional neural network (CNN) called AlexNet. Achieves 91% accuracy in real-time classification with a user-friendly GUI.	Despite achieving an impressive 91% accuracy, the automated bird species identification system using CNN may face challenges in real-time scenarios with environmental noise, necessitating additional training for robust performance.

[7]	Automatic Bird Recognition Using Signal Processing and Neural Networks	Gitanjali Pote1, Ashwini Sase2, Puja Barwal3, Pooja Nandre4, Prof. Neelam Joshi5	2023	Neural network-based system achieves 80% accuracy in bird recognition based on sound signatures. Reviews various studies on audio signal processing and neural networks for bird species identification.	The comprehensive exploration of automatic bird species identification using signal processing and neural networks presents a neural network-based system achieving 80% accuracy, but potential challenges include diverse avian characteristics and varying feature sets affecting identification reliability.
[8]	Automated Bird Species Identification Using Audio Signal Processing and Neural Networks	Avinash Tatar, Bhushan Chavan, Kashyap Bhamare, Snehal Shirode, Abhay Gaidhani	2022	Proposes a two-stage identification approach using audio signal processing and Convolutional Neural Network (CNN), specifically AlexNet. Reports a 91% accuracy in real-time implementation, emphasizing efficiency and scalability.	The automated bird species identification system using audio signal processing and neural networks achieves 91% accuracy in real-time, but faces limitations such as sensitivity to environmental noise, varying accuracy rates, and challenges in handling bird vocalization variations.
[9]	Bird Species Identification Using Audio Signal Processing and Neural Networks	Dr. Amol Dhakne1, Vaishnav M. Kuduvan2, Aniket Palhade3, Tarun Kanjwani4, Rushikesh Kshirsagar5	2022	Introduces a two-stage identification approach with audio signal processing and CNNs. Highlights challenges, advantages, and limitations in bird species identification. Discusses future scope, including mobile applications and ecological parks.	While proposing an automated bird species identification system using audio signal processing and neural networks, the research acknowledges challenges such as limited accuracy, the need for diverse datasets, sensitivity to environmental noise, and variations in bird vocalizations.

[10]	Automated Bird Species Identification Using Audio Signal Processing and Neural Network	Samruddhi Bhor, Rutuja Ganage, Omkar Domb, Hrushikesh Pathade & Shilpa Khedkar	2022	Utilizes Convolutional Neural Networks (CNNs) for automated bird species identification based on sound analysis. Emphasizes the superiority of CNNs over older methods. Addresses challenges like background noise and proposes potential applications.	Despite the advantages, challenges in automated bird species identification using audio signal processing and CNNs include dealing with background noise and the need for efficient noise removal techniques. Additionally, while the proposed method shows promise, further improvements may be required for real-world applications.
[11]	Bird Species Identification Using Convolutional Neural Network	Dharaniya R a,1, Preetha Mb and Yashmi SC	2023	Focuses on using Convolutional Neural Network (CNN) for bird species identification. Discusses challenges in bird identification and highlights advancements in deep learning. Achieves 87% accuracy and suggests applications in wildlife monitoring.	The paper on bird species identification using CNN highlights its benefits, achieving 87%-92% accuracy, but potential challenges include limited dataset size and applicability, requiring consideration for broader ecological contexts.

III. EXISTING SOLUTION

The literature survey explores various innovative approaches to automated bird species identification and speedy image crowd counting. For crowd counting, the integration of a Light Weight Convolutional Neural Network (LW-CNN) is proposed. This addresses the limitations of existing methods, emphasizing the need for uncertainty indication in crowd counting estimates. On the other hand, in bird species identification, multiple studies showcase the effectiveness of Convolutional Neural Networks (CNNs) and signal processing techniques. Methods involve constructing ideal datasets, employing neural networks like AlexNet for real-time classification, and utilizing audio signal processing for accurate bird species recognition. Despite challenges such as environmental noise, these automated systems demonstrate moderate accuracy and potential applications in wildlife monitoring, conservation, and ecological studies. Future directions include mobile applications and continued improvements in accuracy and datasetdiversity. Overall, technology, especially CNNs, is significantly advancing bird species identification and crowd counting solutions for various practical purposes.

IV. PROPOSED SOLUTION

The proposed solution involves the development of a bird species classification system utilizing audio data and an Artificial Neural Network (ANN) model. The process begins with the preprocessing of audio recordings to extract Mel-Frequency Cepstral Coefficients (MFCCs), a widely used feature representation in audio analysis. These extracted features serve as input to the ANN model, which consists of multiple dense layers with Rectified Linear Unit (ReLU) activation functions. The model is trained using the categorical cross-entropy loss function and the Adam optimizer to classify bird species based on their vocalizations. The Flask web application acts as the interface for users to upload audio files and receive predictions regarding the bird species present in the recordings. Upon uploading an audio file, the application processes the file using the trained ANN model and displays the predicted bird species along with an associated image representing the species. Additionally, the application provides visualizations of training and validation accuracy/loss curves, offering insights into the model's performance over epochs.

V. SYSTEM DESIGN AND ARCHITECTURE

- ***System Architecture:***

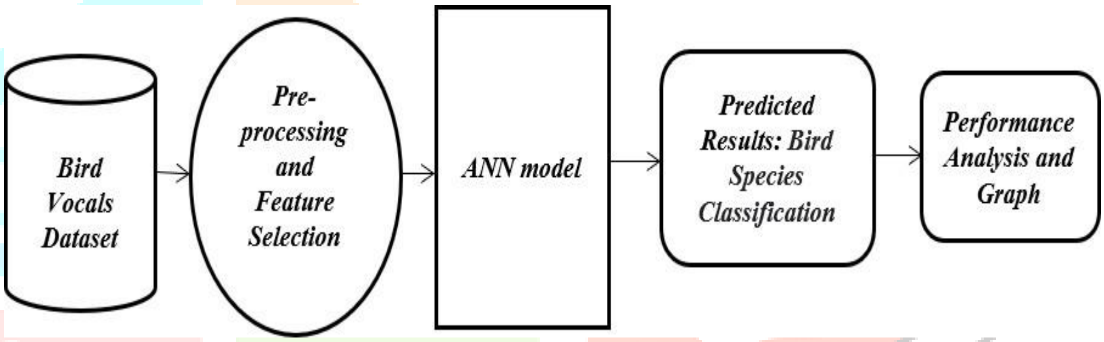


Fig 1: System Architecture Design

- ***Class Diagram:***

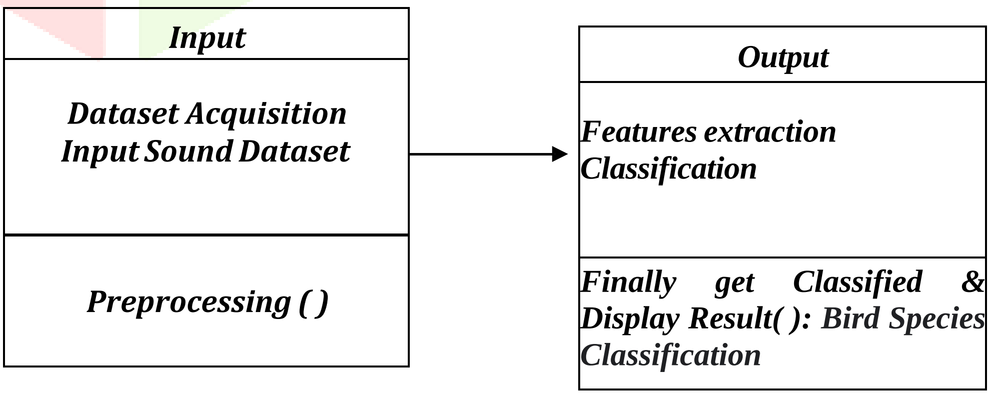


Fig 2: Class Diagram

➢ *Dataflow Diagram:*

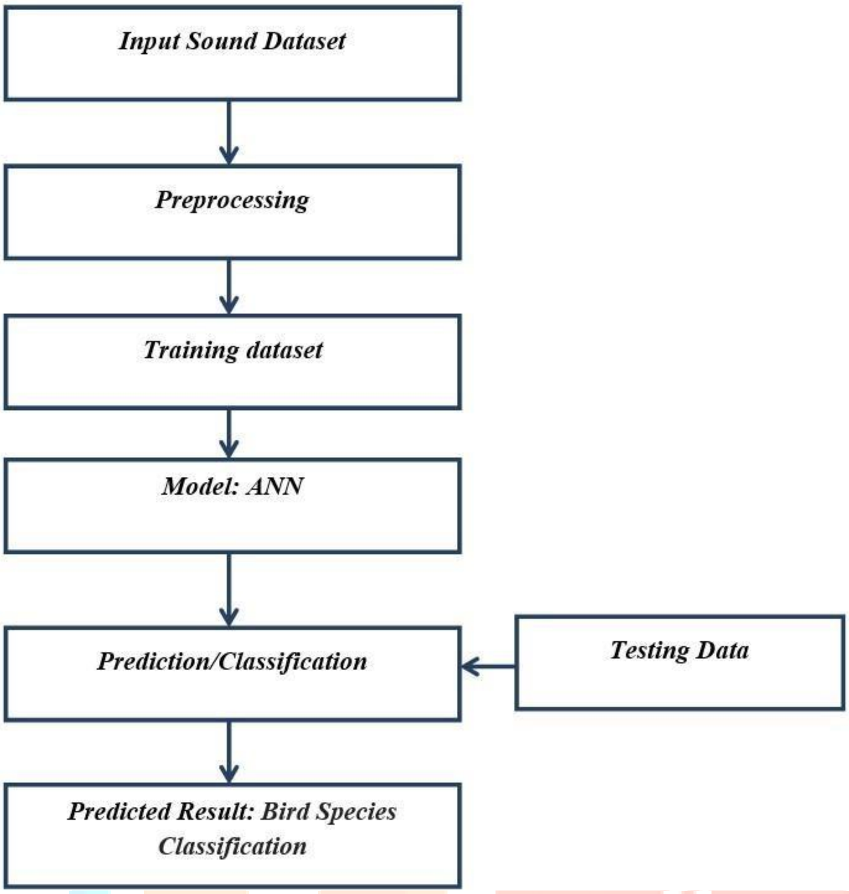


Fig 3: Dataflow Diagram

➢ *Use Case Diagram:*

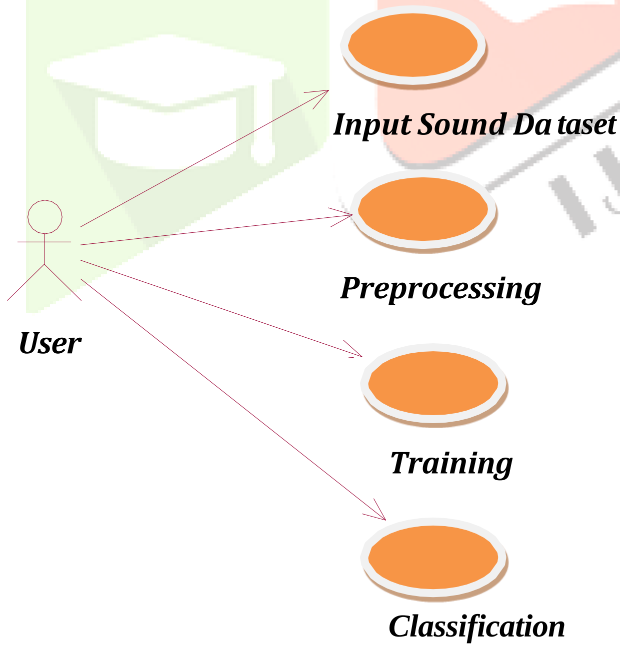


Fig 4: Use Case Diagram

➢ ***Activity Diagram:***

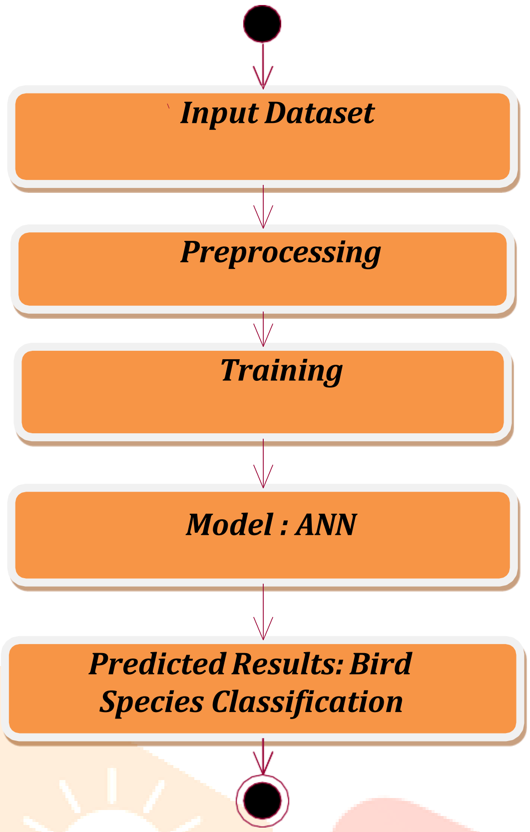


Fig 5: Activity diagram

➢ ***Sequence Diagram:***

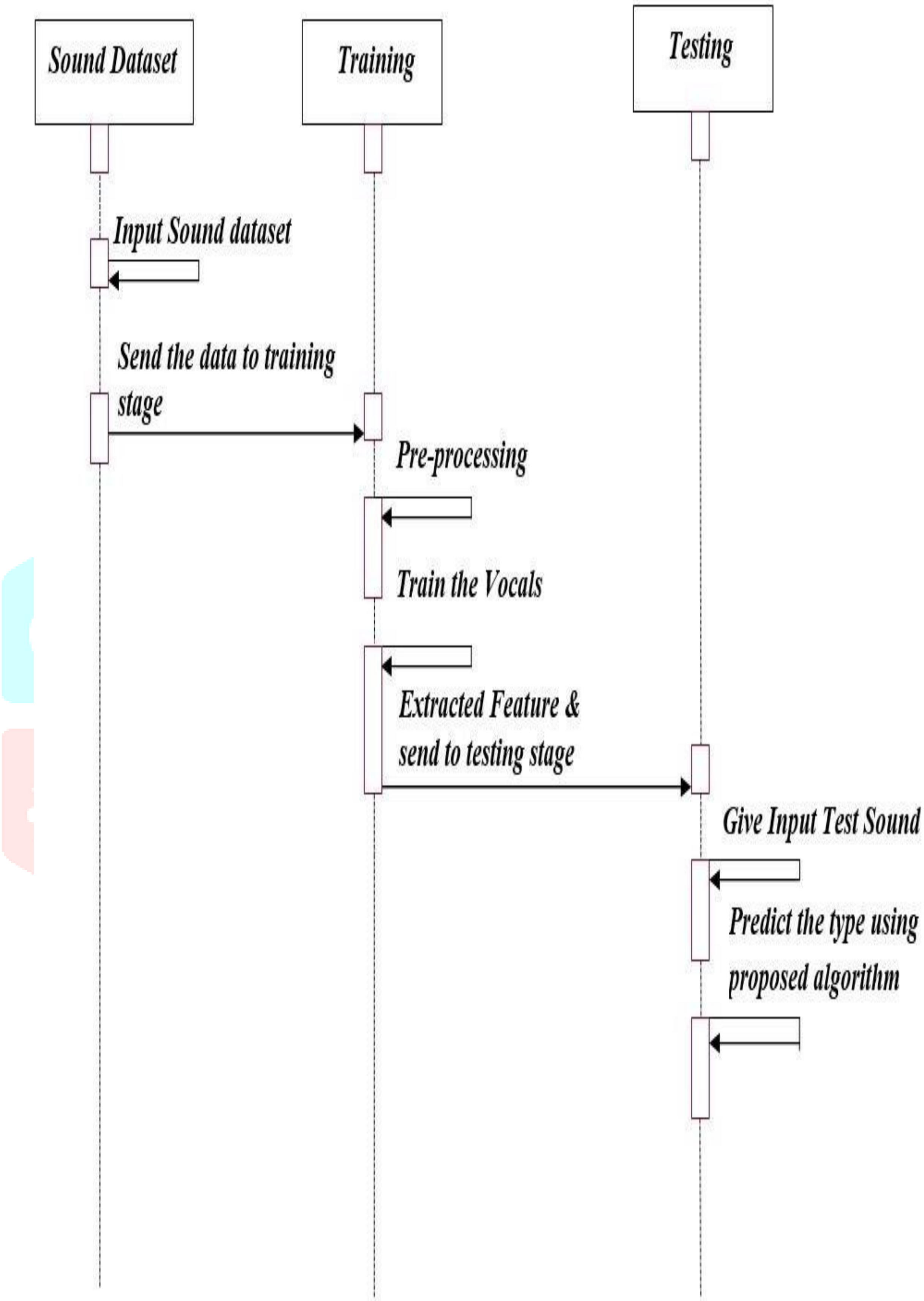


Fig 6: Sequence Diagram

VI. IMPLEMENTATION

➢ ***Introduction***

Avi-Audio Net marks a significant advancement in the fusion of technology and ecology, presenting an automated solution for bird species identification. This project harnesses a rich array of technologies, including Python, HTML, CSS, and JavaScript, to deliver a comprehensive platform enabling users to upload audio recordings of bird calls and receive automated identification outcomes.

Python serves as the cornerstone for backend server development and the implementation of machine learning models. Leveraging the Flask framework, Avi-AudioNet efficiently handles HTTP requests and seamlessly integrates with frontend components. Machine learning algorithms, particularly Artificial Neural Networks (ANNs), are pivotal in the analysis of audio features extracted using signal processing techniques such as Fourier transforms and Mel-frequency cepstral coefficients (MFCCs).

On the frontend, HTML, CSS, and JavaScript synergize to craft an intuitive user interface (UI) accessible via standard web browsers. This UI empowers users to effortlessly upload audio recordings, visualize identification results, and provide feedback for algorithm enhancement. JavaScript enhances the UI's interactivity and responsiveness, guaranteeing a seamless user experience across various devices and screen sizes.

The development process prioritizes scalability, reliability, security, and accessibility. Avi-AudioNet's architecture is meticulously designed to accommodate escalating user traffic, maintain consistent performance, safeguard user data confidentiality, and comply with accessibility standards for users with disabilities. Comprehensive documentation, modular code structure, and well-commented code ensure ease of maintenance and facilitate future updates. Avi-AudioNet emerges as a pivotal step forward in leveraging technology and data-driven methodologies to address pressing ecological challenges and foster biodiversity conservation efforts.

➢ ***Algorithms Used***

The core of Avi-Audio Net's functionality lies in the machine learning algorithms employed for bird species identification. The following algorithms were utilized:

- ***Artificial Neural Networks (ANNs):*** ANNs are a class of machine learning models inspired by the structure and function of the human brain. In Avi-Audio Net, ANNs were employed for classification tasks, learning patterns in audio features extracted from bird recordings to predict the corresponding bird species.

- ***Signal Processing Techniques:*** Signal processing techniques, such as Fourier transforms and Mel-frequency cepstral coefficients (MFCCs), were utilized for extracting relevant features from audio recordings. These techniques enable the conversion of raw audio data into a format suitable for input to machine learning models, capturing essential characteristics of bird vocalizations.

1. ***Peak Detection Algorithms:*** Peak detection algorithms, such as find_peaks() from the SciPy library, were used to identify segments of interest in audio recordings corresponding to bird calls. These algorithms help isolate bird vocalizations from background noise, facilitating more accurate species identification.

By leveraging these algorithms, Avi-Audio Net achieved robust performance in automatically identifying bird species from audio recordings, contributing to advancements in ecological research and conservation efforts.

➢ ***How does The Project Work?***

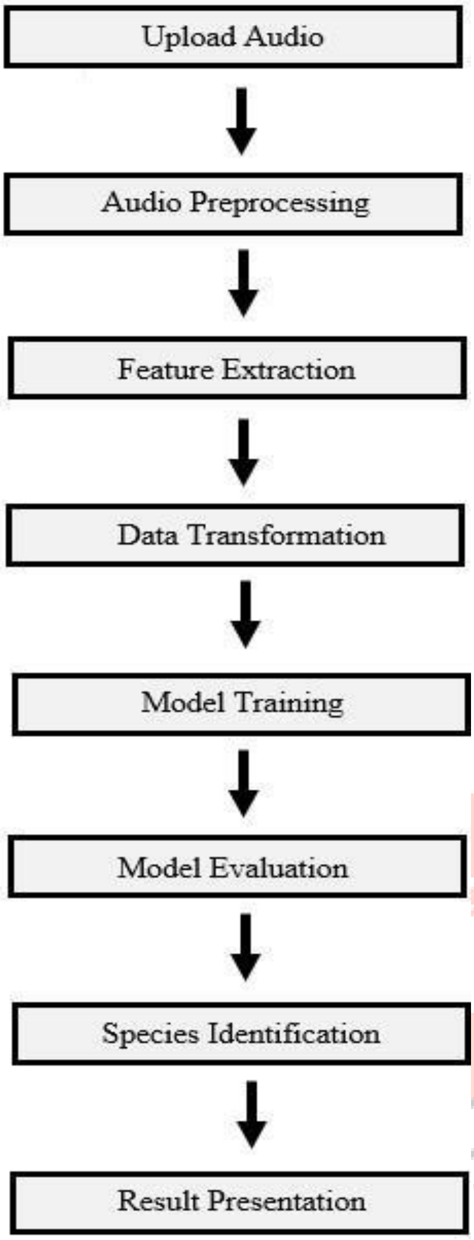


Fig 7: Sound Acquisition Process

- ***Upload Audio:***
  In this step, users upload an audio recording of a bird call. The technology used here depends on the implementation, but common options include HTML forms for web interfaces or command-line tools for terminal-based interfaces.
- ***Audio Preprocessing:***
  After upload, the audio recording undergoes preprocessing to enhance its quality. Technologies such as PyDub in Python or FFmpeg command-line tool can be used for tasks like noise reduction and normalization.
- ***Feature Extraction:***
  Relevant features are extracted from the preprocessed audio recording using signal processing techniques. Libraries like Librosa in Python are commonly used for this purpose.
- ***Data Transformation:***
  The extracted features are transformed into a format suitable for input into the machine learning model. This transformation can be done using NumPy and Pandas in Python for data manipulation and formatting.
- ***Model Training:***
  Machine learning models are trained using the transformed features and corresponding labels. Technologies such as Keras or TensorFlow in Python are often used for building and training neural network models.

- ***Model Evaluation:***
  The trained model's performance is evaluated using validation data. This evaluation can be done using Scikit-learn in Python, which provides tools for computing various evaluation metrics.
- ***Species Identification:***
  The trained model is used to predict the bird species present in the audio recording. This prediction step typically involves running inference with the trained model using libraries like Keras or TensorFlow.
- ***Result Presentation:***
  Finally, the identification results are presented to the user. This presentation can be done using HTML and CSS for web interfaces, or terminal output for command-line interfaces. Additionally, visualization libraries like Matplotlib in Python can be used to generate informative plots and charts for result visualization.

- ***Importing Necessary Libraries***

During the implementation phase, each technology played a pivotal role in realizing the vision of Avi- Audio Net. Here's a closer look at how each technology was leveraged:

- **Python:** As the backbone of the project, Python facilitated various aspects of Avi-Audio Net's development. Its ease of use, extensive library ecosystem, and versatility made it the ideal choice for backend server development, machine learning model implementation, and data processing tasks.
- **Flask:** Flask, a lightweight and powerful web framework, formed the foundation of Avi-Audio Net's backend infrastructure. It enabled the creation of robust APIs for handling HTTP requests, managing file uploads, and orchestrating interactions between the frontend and backend components.
- **PyDub:** PyDub emerged as a crucial tool for audio manipulation and processing within Avi-Audio Net. This Python library provided essential functionality for tasks such as trimming audio clips, adjusting volume levels, and converting between different audio file formats, ensuring compatibility with various recording sources.
- **Keras:** Keras, a high-level neural networks API, was instrumental in building and training the machine learning models at the heart of Avi-Audio Net's bird species identification system. Its user-friendly interface allowed for rapid prototyping of deep learning architectures, enabling experimentation with different model configurations and hyperparameters.
- **Librosa:** Librosa, a Python library for audio and music analysis, facilitated the extraction of meaningful features from audio recordings in Avi-Audio Net. By computing Mel-frequency cepstral coefficients (MFCCs), spectral contrast, and other audio descriptors, Librosa provided valuable insights into the acoustic characteristics of bird vocalizations.
- **NumPy and Pandas:** NumPy and Pandas formed the backbone of Avi-Audio Net's data processing pipeline. NumPy's efficient array operations and mathematical functions were used for numerical computations and data manipulation tasks. Meanwhile, Pandas provided a powerful framework for working with structured data, enabling seamless integration with CSV files containing metadata.
- **SciPy:** SciPy, a comprehensive library for scientific computing, offered a diverse range of signal processing functionalities critical to Avi-Audio Net's success. From Fourier transforms and spectral analysis to peak detection and filtering, SciPy provided the tools necessary to preprocess and analyze audio data with precision
- **Scikit-learn:** Scikit-learn played a central role in Avi-Audio Net's machine learning workflow, providing a vast array of tools and algorithms for data preprocessing, model training, and evaluation. By leveraging Scikit-learn's implementation of various classification algorithms, including support vector machines (SVMs) and random forests, Avi-Audio Net achieved robust performance in bird species identification.
- **Matplotlib:** Matplotlib enabled the creation of informative visualizations to aid in the analysis and interpretation of Avi-Audio Net's results. By generating plots depicting accuracy metrics, loss curves, and feature distributions, Matplotlib facilitated deeper insights into the performance and behavior of the machine learning models deployed in Avi-Audio Net.

Through the effective integration of these technologies, Avi-Audio Net emerged as a sophisticated platform capable of automating the identification of bird species from audio recordings. This implementation underscores the transformative potential of technology in addressing real-world ecological challenges and advancing biodiversity conservation efforts.

- ***Exploratory Data Analysis of Audio data :***

We have 5 different folders under the urban dataset folder. Before applying any preprocessing, we will try to understand how to load audio files and how to visualize them in form of the waveform. If you want to load the audio file and listen to it, then you can use the IPython library and directly give it an audio file path. We have taken the first audio file in the folder 1 that belongs to the bird vocals category.

Now we will use Librosa to load audio data. So when we load any audio file with Librosa, it gives us 2 things. One is sample rate, and the other is a two-dimensional array. Let us load the above audio file with Librosa and plot the waveform using Librosa.

- ***Sample rate*** – It represents how many samples are recorded per second. The default sampling rate with which librosa reads the file is 22050. The sample rate differs by the library you choose.
- ***2-D Array*** – The first axis represents recorded samples of amplitude. And the second axis represents the number of channels. There are different types of channels – Monophonic(audio that has one channel) and stereo(audio that has two channels).

We load the data with librosa, then it normalizes the entire data and tries to give it in a single sample rate. The same we can achieve using scipy python library also. It will also give us two pieces of information – one is sample rate, and the other is data.

When you print the sample rate using scipy-it is different than librosa. Now let us visualize the wave audio data. One important thing to understand between both is- when we print the data retrieved from librosa, it can be normalized, but when we try to read an audio file using scipy, it can't be normalized. Librosa is now getting popular for audio signal processing because of the following three reasons.

1. It tries to converge the signal into mono(one channel).
2. It can represent the audio signal between -1 to +1(in normalized form), so a regular pattern is observed.
3. It is also able to see the sample rate, and by default, it converts it to 22 kHz, while in the case of other libraries, we see it according to a different value.

- ***Imbalance Dataset Check :***

Now we know about the audio files and how to visualize them in audio format. Moving format to data exploration we will load the CSV data file provided for each audio file and check how many records we have for each class.The data we have is a filename and where it is present so let us explore 1st file, so it is present in fold 5 with category as a bird vocals. Now use the value counts function to check records of each class.When you see the output so data is not imbalanced, and most of the classes have an approximately equal number of records

- ***Data Preprocessing :***

Some audios are getting recorded at a different rate-like 44KHz or 22KHz. Using librosa, it will be at 22KHz, and then, we can see the data in a normalized pattern. Now, our task is to extract some important information, and keep our data in the form of independent(Extracted features from the audio signal) and dependent features(class labels). We will use Mel Frequency Cepstral coefficients to extract independent features from audio signals.

MFCCs – The MFCC summarizes the frequency distribution across the window size. So, it is possible to analyze both the frequency and time characteristics of the sound. This audio representation will allow us to identify features for classification. So, it will try to convert audio into some kind of features based on time and frequency characteristics that will help us to do classification. To demonstrate how we apply MFCC in practice, first, we will apply it on a single audio file that we are already using.

Now, we have to extract features from all the audio files and prepare the dataframe. So, we will create a function that takes the filename(file path where it is present). It loads the file using librosa, where we get 2 information. First, we'll find MFCC for the audio data, And to find out scaled features, we'll find the mean of the transpose of an array.

Now, to extract all the features for each audio file, we have to use a loop over each row in the dataframe. We also use the TQDM python library to track the progress. Inside the loop, we'll prepare a customized file path for each file and call the function to extract MFCC features and append features and corresponding labels in a newly formed dataframe.

➢ ***Splitting the dataset :***

Split the dataset into train and test. 80% train data and 20% test data.

➢ ***Audio Classification Model Creation :***

We have extracted features from the audio sample and splitter in the train and test set. Now we will implement an ANN model using Keras sequential API. The number of classes is 10, which is our output shape(number of classes), and we will create ANN with 3 dense layers and architecture is explained below.

1. The first layer has 100 neurons. Input shape is 40 according to the number of features with activation function as Relu, and to avoid any overfitting, we'll use the Dropout layer at a rate of 0.5.
2. The second layer has 200 neurons with activation function as Relu and the drop out at a rate of 0.5.
3. The third layer again has 100 neurons with activation as Relu and the drop out at a rate of 0.5.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
dense (Dense)                (None, 1000)              41000
_________________________________________________________________
dense_1 (Dense)              (None, 750)               750750
_________________________________________________________________
dense_2 (Dense)              (None, 500)               375500
_________________________________________________________________
dense_3 (Dense)              (None, 250)               125250
_________________________________________________________________
dense_4 (Dense)              (None, 100)               25100
_________________________________________________________________
dense_5 (Dense)              (None, 50)                5050
_________________________________________________________________
dense_6 (Dense)              (None, 5)                 255
=================================================================
Total params: 1,322,905
Trainable params: 1,322,905
Non-trainable params: 0
_________________________________________________________________
```

Fig 8 : Training Layers

➢ ***Compile the Model :***

To compile the model we need to define loss function which is categorical cross-entropy, accuracy metrics which is accuracy score, and an optimizer which is Adam.

➢ ***Train the Model :***

We will train the model and save the model in HDF5 format. We will train a model for 250 epochs and batch size as 32. We'll use callback, which is a checkpoint to know how much time it took to train over data.

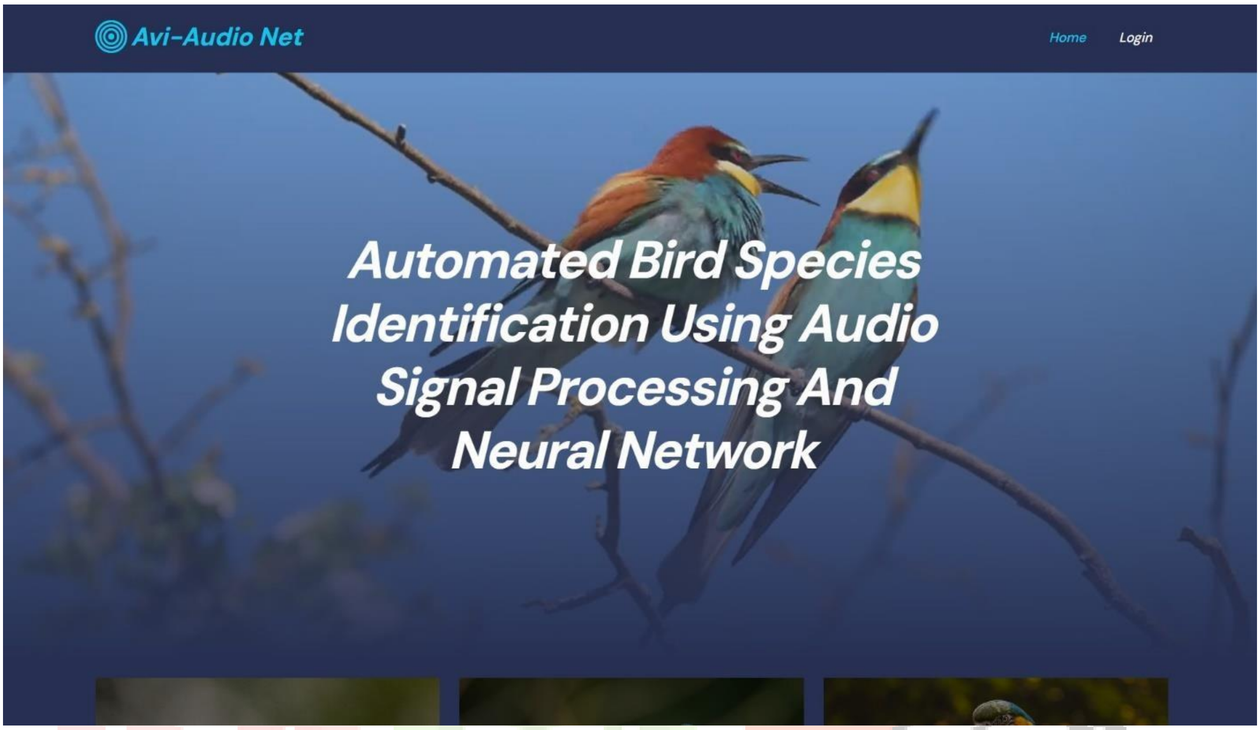
- ***Check the Test Accuracy :***

Now we will evaluate the model on test data. We got near about 97 percent accuracy on the training dataset and 100 percent on test data.

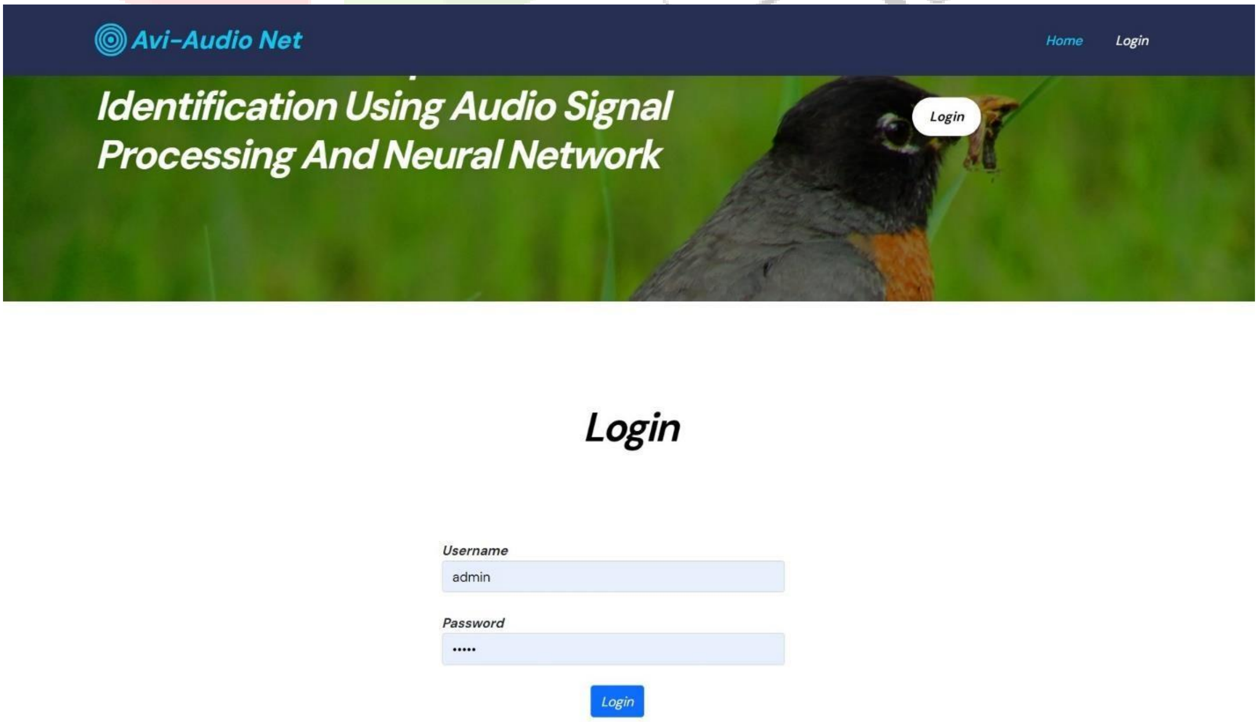
- ***Saving the Trained Model :***

Once you're confident enough to take your trained and tested model into the production-ready environment, the first step is to save it into a .h5 or .pkl file using a library like pickle. Make sure you have pickle installed in your environment, then import the module and dump the model into .h5 file.

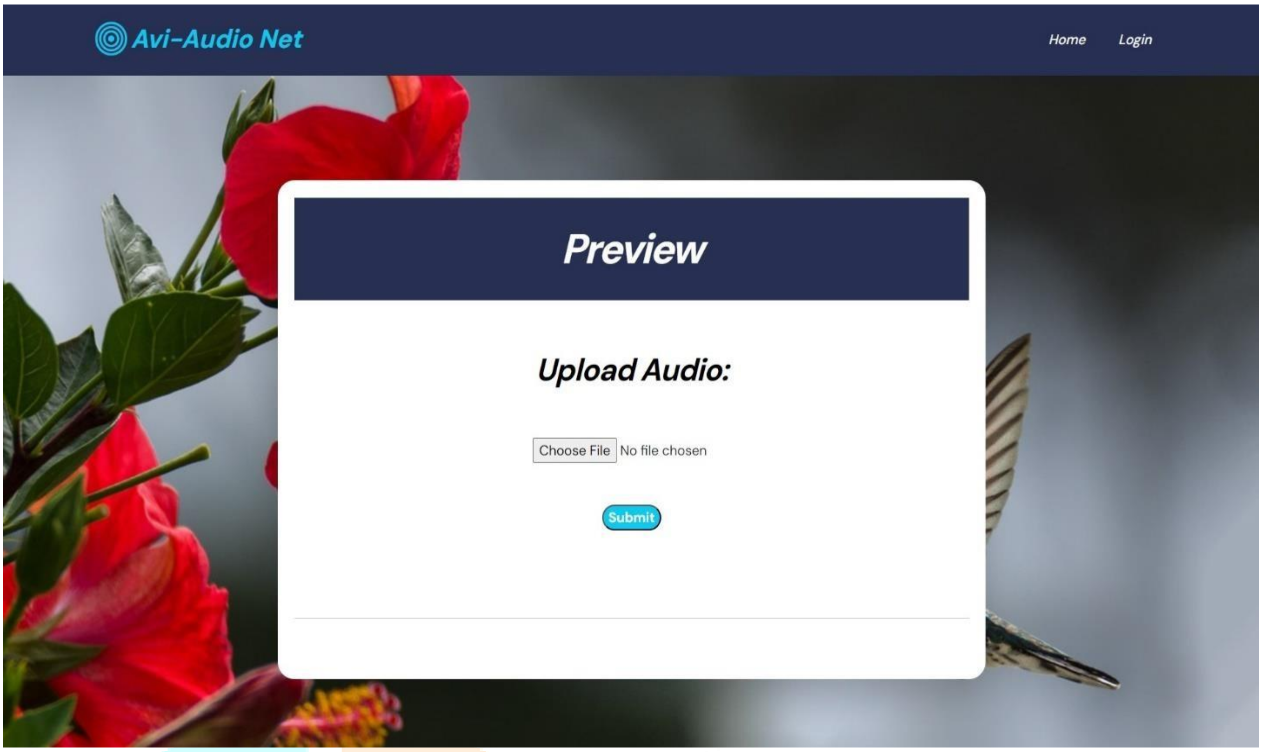
VII. RESULTS & SCREENSHOTS



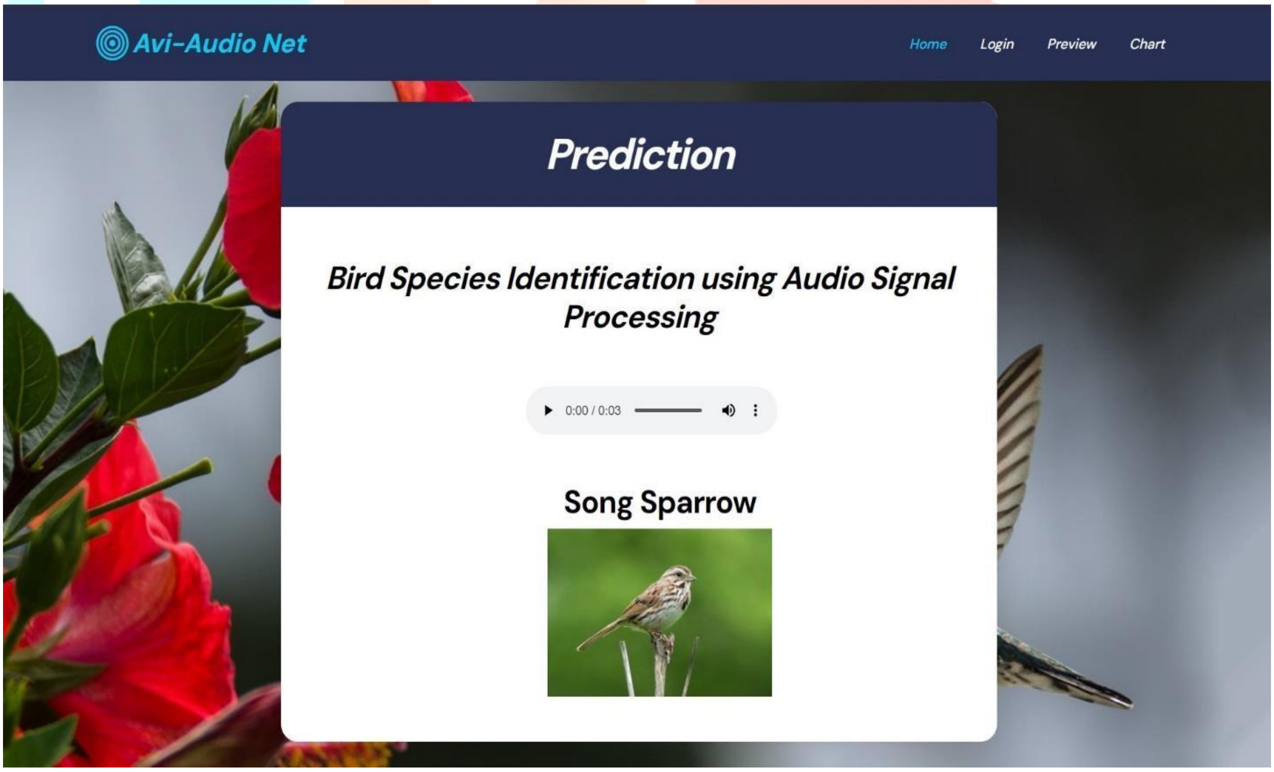
Screenshot 1 : *Home Page*



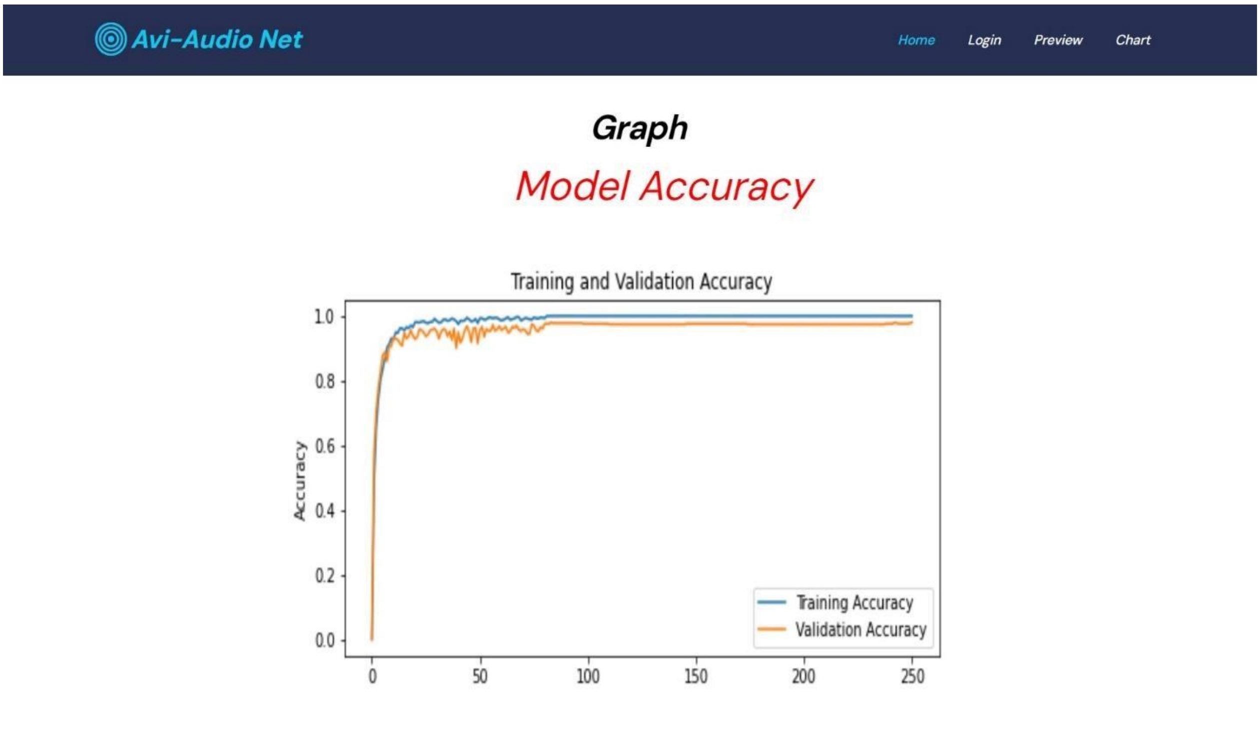
Screenshot 2 : *Login Page*



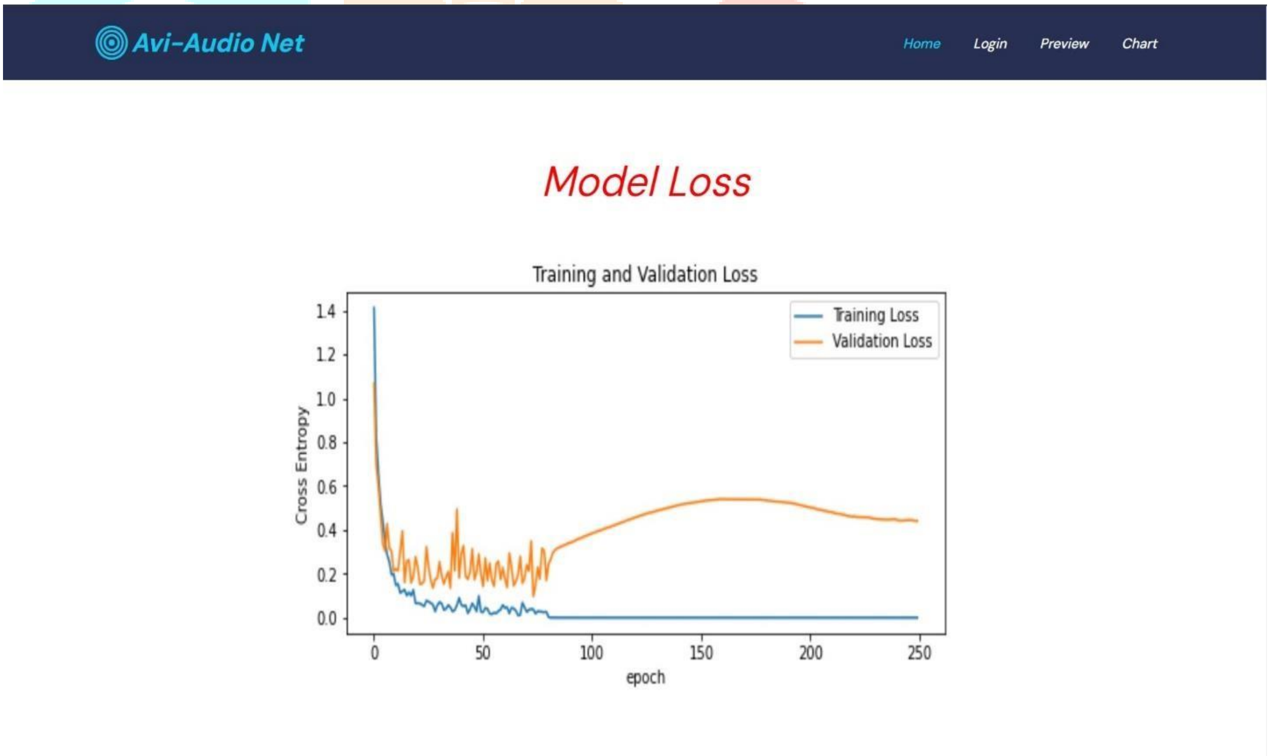
Screenshot 3 : *Preview Page*



Screenshot 4 : *Prediction Page*



Screenshot 5 : *Model Accuracy*



Screenshot 6 : *Model Loss*

VIII. CONCLUSION

In conclusion, the developed bird species classification system demonstrates the potential of Artificial Neural Networks (ANNs) in analyzing bird vocalizations and providing automated solutions for bird species identification. The system offers a user-friendly interface for researchers, ornithologists, and enthusiasts to analyze audio recordings and identify bird species accurately. While the system shows promising results, further refinements and optimizations are necessary to enhance its scalability, robustness, and accuracy. Future directions may include experimenting with different ANN architectures, hyperparameter tuning, and incorporating transfer learning techniques to improve classification performance. Overall, the project contributes to ornithology, ecological research, and conservation efforts by facilitating the monitoring of bird populations and biodiversity.

IX. FUTURE WORKS

Future work involves optimizing the current ANN model for better performance by refining its architecture and adjusting hyperparameters. Data augmentation techniques will be explored to diversify training data, while transfer learning and ensemble methods may be implemented to enhance classification accuracy. Deployment on cloud platforms, usability enhancements, and integration with existing bird monitoring systems are also priorities. Continuous evaluation and user feedback will drive further improvements to ensure relevance and effectiveness in practical scenarios.

X. REFERENCES

[1]Vivekanandam, B. "***Speedy Image Crowd Counting by Light Weight Convolutional Neural Network***“. Journal of Innovative Image Processing 3, no. 3 (2021): 208-222.

[2]Manoharan, J. Samuel. "***Study of Variants of Extreme Learning Machine (ELM) Brands and its Performance Measure on Classification Algorithm***“. Journal of Soft Computing Paradigm (JSCP) 3, no. 02 (2021): 83-95.

[3]Chandu B, A. M. (2020). “***Automated Bird Species Identificat ion using Audio***”. 2020 International Conference on Artificial Intelligence and Signal Processing (AISP).

[4]M. M. M. Sukri, U. Fadlilah, S. Saon, A. K. Mahamad, M. M. Som and A. Sidek, "***Bird Sound Identification based on Artificial Neural***" IEEE 2020 IEEE Student Conference on Research and Development (SCOReD),
pp. 342-345, 2020.

[5]Raja Shekar Kadurka & Harish Kanakalla. “***Automated Bird Detection in Audio Recordings by a Signal Processing Perspective***” (2021).

[6]Mohammed Sadiq B1, Pooja H M2. “***Automated Audio Signal Processing for Bird Species Identification Including Neural Networks***” (2023).

[7]Gitanjali Pote1, Ashwini Sase2, Puja Barwal3, Pooja Nandre4, Prof. Neelam Joshi5. “***Automatic Bird Recognition Using Signal Processing and Neural Networks***” (2023).

[8]Avinash Tatar,Bhushan Chavan,Kashyap Bhamare,Snehal Shirode, Abhay Gaidhani. “***Automated Bird Species Identification using Audio Signal Processing and Neural Networks***”. International Journal of Creative Research Thoughts (IJCRT) ,vol 11, no. 03,2023.

[9]Dr. Amol Dhakne1, Vaishnav M. Kuduvan2, Aniket Palhade3, Tarun Kanjwani4, Rushikesh Kshirsagar5. “***Bird Species Identification Using Audio Signal Processing and Neural Networks***” (2022).

[10] Samruddhi Bhor, Rutuja Ganage, Omkar Domb, Hrushikesh Pathade & Shilpa Khedkar. “***Automated Bird Species Identification Using Audio Signal Processing and Neural Network***” (2022).

[11] Dharaniya, R., Preetha, M., & Yashmi, S. C. (2022). “***Bird Species Identification Using Convolutional Neural Network***”. In D.J. Hemanth et al. (Eds.), Advances in Parallel Computing Algorithms, Tools and Paradigms , IOS Press. (2022).