**Table of Contents**

S.NO Page No.

1. Abstract --------------------------------------- 2
2. Introduction ----------------------------------- 3-4
3. Literature review ----------------------------- 5-7
4. Problem Statement -------------------------- 7
5. Explorartory Data Analysis----------------- 7-10
   1. Data Collection and Preparation ------- 8
   2. Feature Extraction and Augmentation- 9
   3. Neural Network Architecture Design--- 9
   4. Dataset Description ----------------------- 9
   5. Model Training and Validation ---------- 9
   6. Testing and Evaluation -------------------- 10
   7. Legal and Ethical Considerations ------- 10
   8. Documentation and Reporting --------- 10
6. Proposed Methodology -------------------- 10-11
7. Result & Discussion ------------------- 11-12
8. Advantages ---------------------------------- 13
9. Limitations------------------------------------ 13
10. Conclusion ---------------------------- 13 11. References --------------------- 14-15
11. **Abstract**

Birds play a vital role in the study of ecosystems and biodiversity. Accurate bird identification helps monitor biodiversity, understand the functions of ecosystems, and develop effective conservation strategies. However, previous bird sound recognition methods often relied on single features and overlooked the spatial information associated with these features, leading to low accuracy.  Recognizing this gap, the present study proposed to create a highly efficient and precise bird sound recognition system by leveraging advanced deep learning techniques, with a specific focus on artificial neural networks (ANN). To achieve this, the project will employ a comprehensive exploratory data analysis (EDA) approach, which involves the systematic examination of the dataset to gain valuable insights into the characteristics of bird sounds and their variations. The EDA phase will help in data preprocessing, feature extraction, and the selection of relevant acoustic features that are crucial for effective bird sound recognition. Artificial neural networks will be designed and trained to process these features, enabling the system to classify and identify different bird species accurately. The utilization of ANNs will also allow for continuous improvement in recognition accuracy through iterative model optimization. In conclusion, this project aims to combine the power of deep learning and data analysis to develop a robust and sophisticated bird sound recognition system with applications in biodiversity monitoring, ecological research, and bird species conservation.

1. **Introduction**

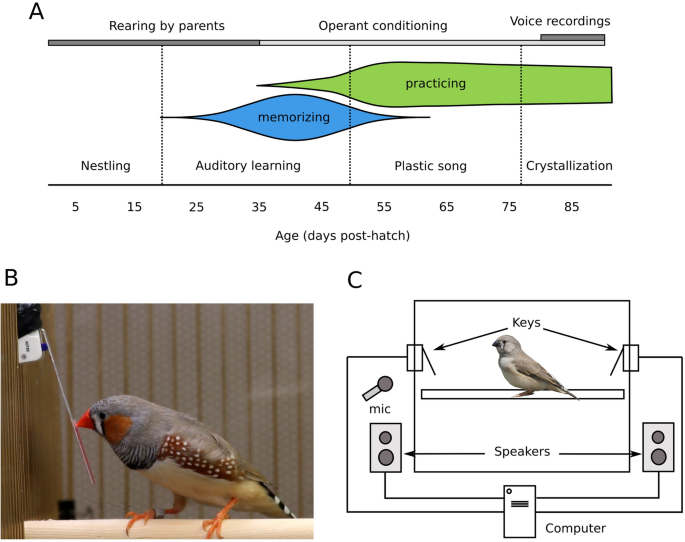
Bird species protection plays a vital role in biodiversity conservation [**1**,**2**]. The abundance of bird populations directly reflects local biodiversity levels, emphasizing the importance of bird recognition in ecological research and biodiversity conservation. Bird sound-based recognition methods offer significant advantages over other bird recognition methods that are commonly used in ornithological studies, such as image-based methods and the marked-recapture method [[**3**](https://www.mdpi.com/1424-8220/23/19/8099#B3-sensors-23-08099)]. They have wider recognition ranges, are immune to forest obstructions, and experience reduced interference from human activities, making them the preferred solutions for bird recognition [[**4**](https://www.mdpi.com/1424-8220/23/19/8099#B4-sensors-23-08099)].

A bird sitting on a branch

Description automatically generatedIn recent years, bird populations have experienced a noticeable decline due to human and environmental factors. Morrison et al. [[**5**](https://www.mdpi.com/1424-8220/23/19/8099#B5-sensors-23-08099)] conducted a study that presented compelling evidence of a substantial decrease in the diversity and intensity of bird soundscapes across over 200,000 sites in the Northern Hemisphere over the past 25 years. This decline can be attributed to significant decreases in bird species and individual abundance, which have far-reaching consequences on ecosystem health and biodiversity. Therefore, swiftly and accurately recognizing birds through sound analysis is critical for bird population monitoring and ecological conservation efforts.

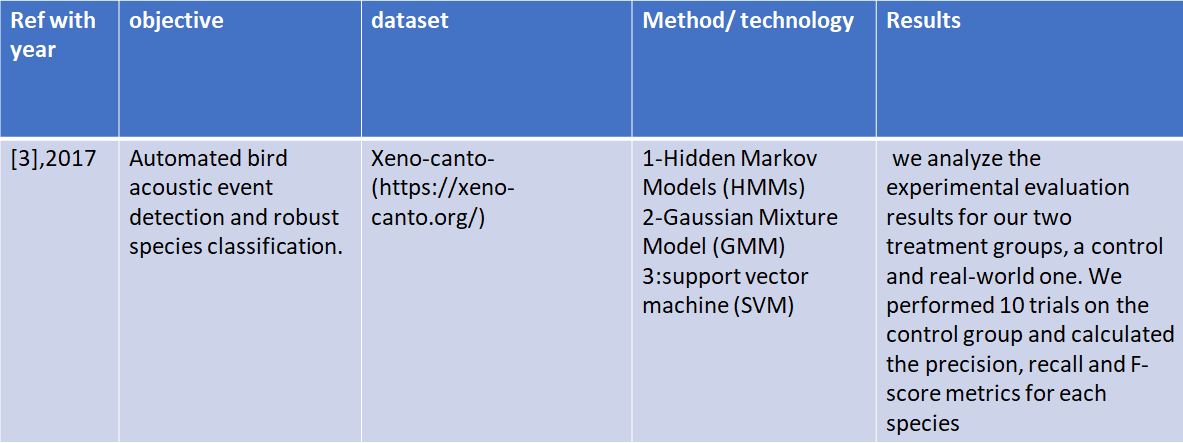
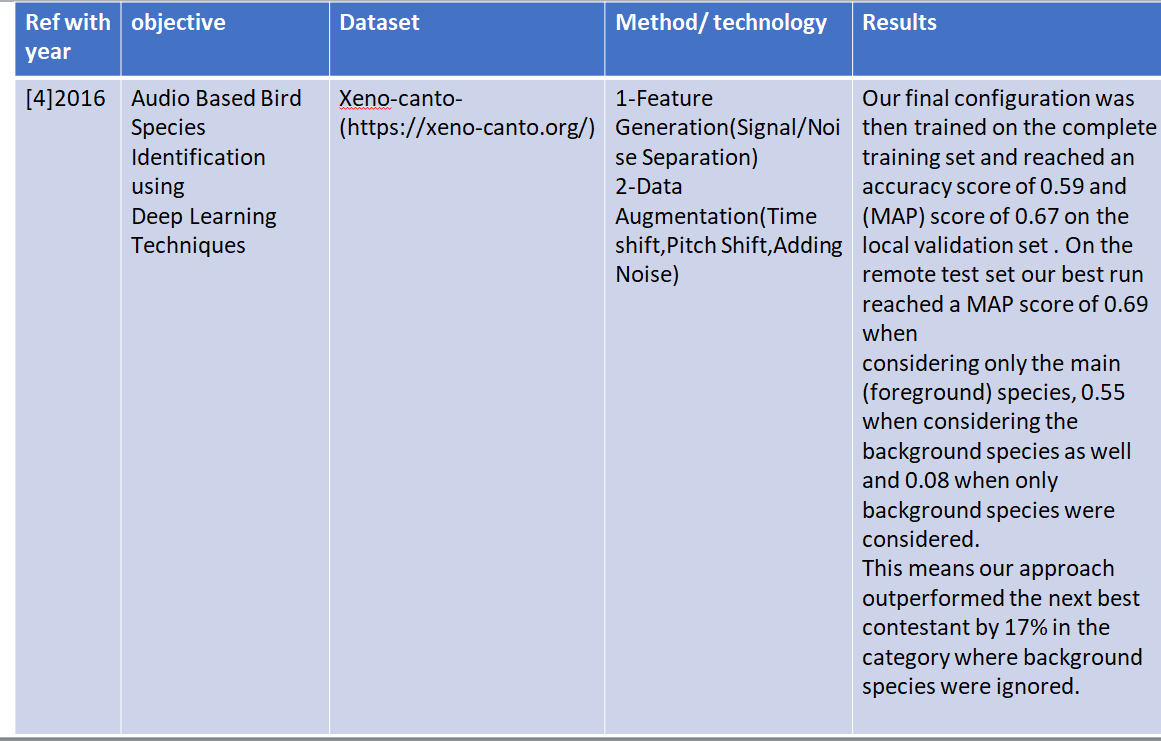
**(Fig 1)Bird species**

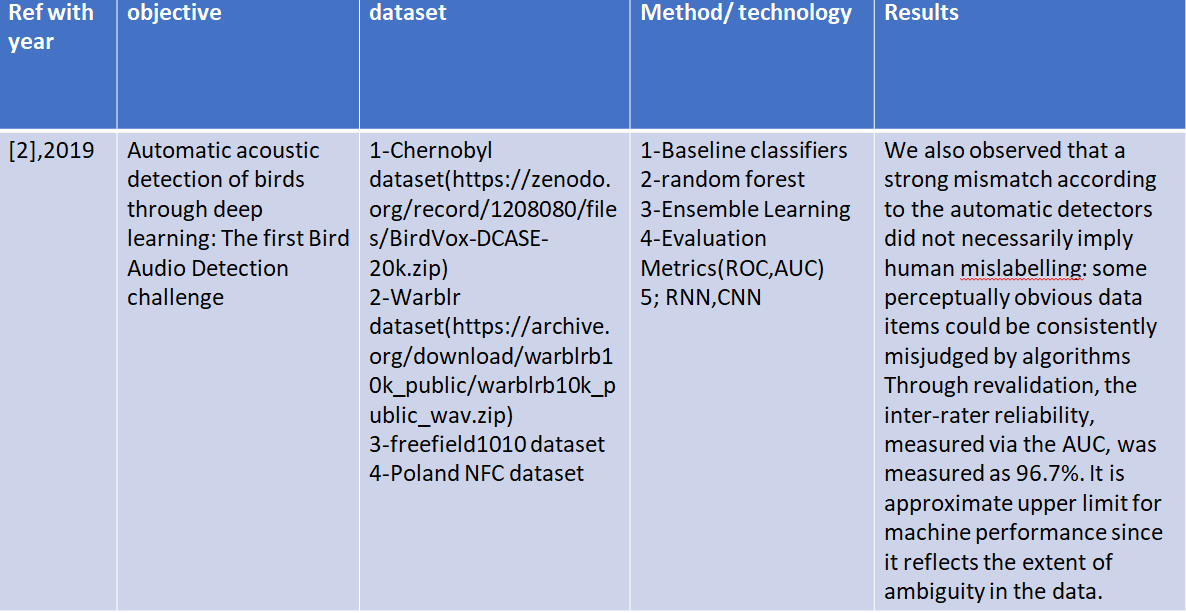
Bird vocalizations are commonly employed to assess population trends, with call rates serving as a useful proxy for determining the stability, growth, or decline of bird populations over time. This method proves valuable in ornithology, environmental monitoring, and conservation. The LifeCLEF bird identification task challenges participants to recognize various bird species within extensive audio recordings from Xeno-Canto. To effectively process lengthy field recordings, often featuring extended periods without bird calls, the analysis involves initial segmentation to detect potential sounds, followed by extraction for detailed processing. Utilizing artificial neural networks and EDA, our goal is to develop a system that efficiently and accurately identifies species based on their vocalizations. For this we use Artificial Neural Networks (ANNs) [6] which serve as a powerful device for pattern reputation and class. ANNs are computational fashions inspired by means of the structure and functioning of the human brain, such as interconnected nodes that procedure facts. By leveraging deep studying strategies, ANNs can autonomously learn complicated hierarchical features from chook sound data at some stage in the training phase. This enables them to generalize nicely and as it should be categorize unseen hen vocalizations, making them a promising solution for automating the identification of bird species based totally on their distinctive acoustic signatures. This technology has significant potential for advancing avian biodiversity understanding, supporting ecological research, and aiding endangered species conservation. Through comprehensive study and data analysis integration, our project aims to contribute to practical applications of bird sound detection in the real world. Our proposed solution involves creating a filter-based system trainable on a small dataset for recognizing specific species of interest.

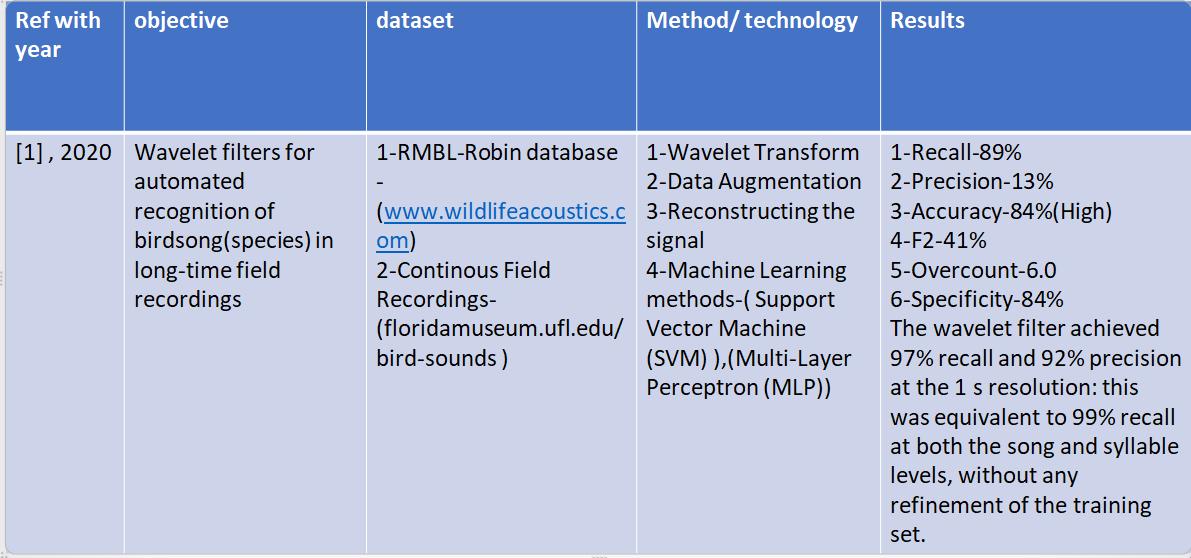


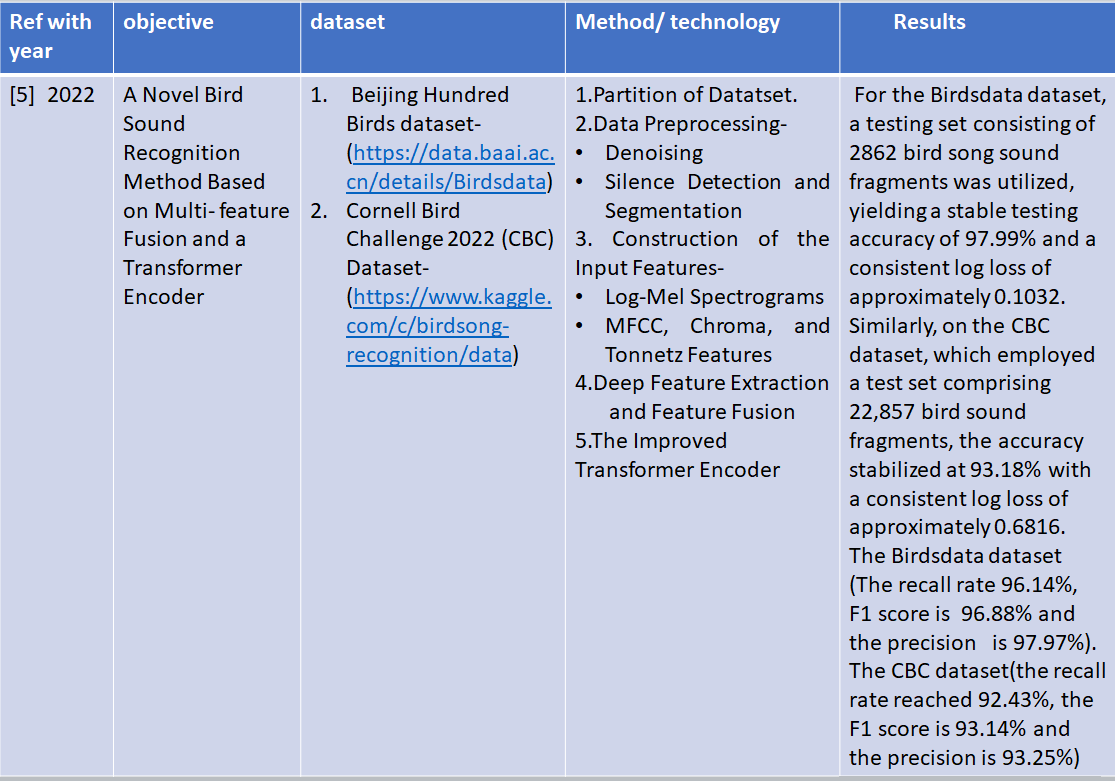
# (Fig 2) Song preferences predict the quality of voice

**3. Literature Review**







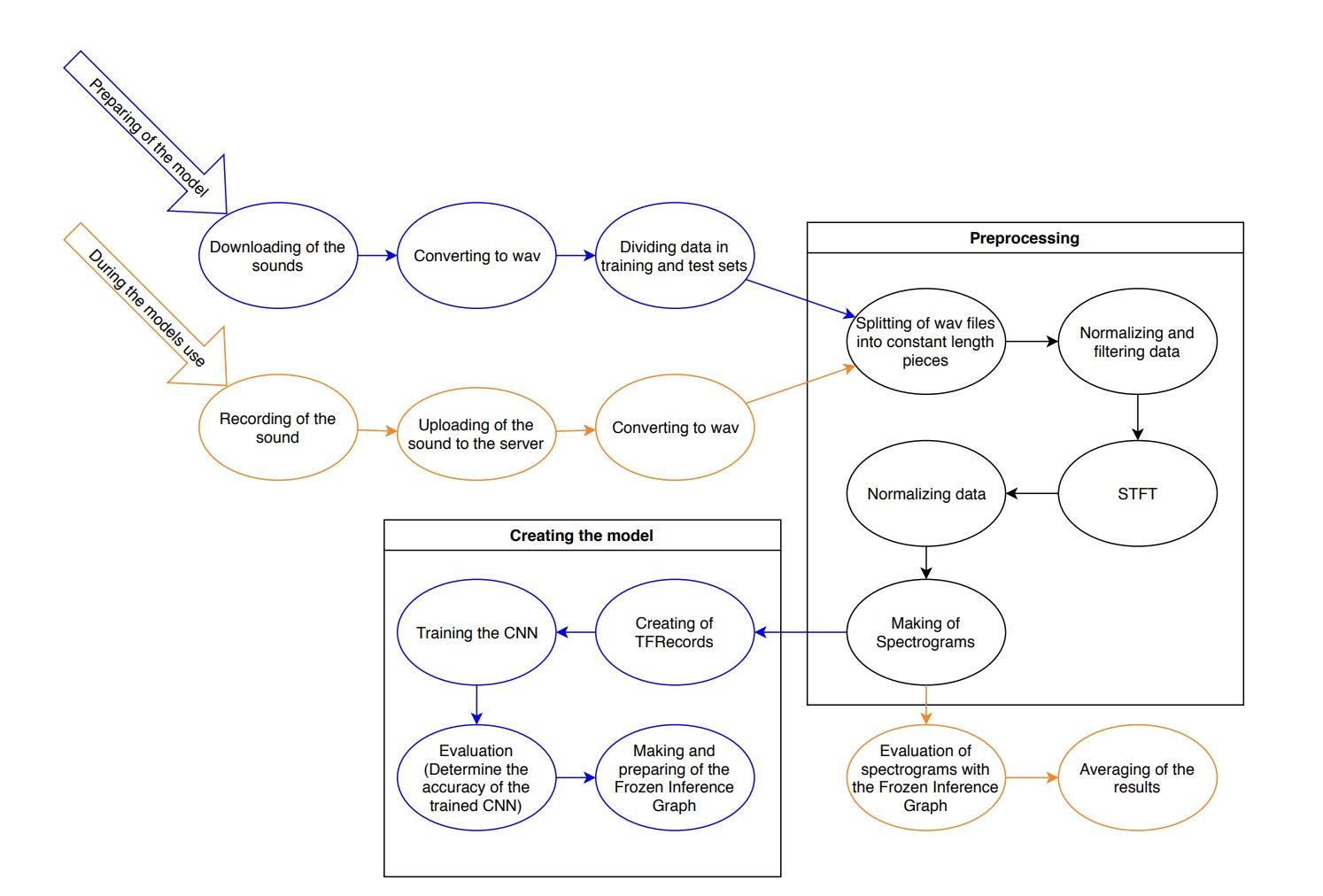


**4. Problem Statement**

Develop advanced bird sound recognition systems by integrating deep learning models, such as artificial-neural networks (ANNs) and exploratory data analysis (EDA). The goal is to contribute to biodiversity management and ecological research by developing accurate and efficient solutions to identify bird species based on their unique calls.

**5. Exploratory Data Analysis**

The methodology for your project, "Bird Sound Recognition using Deep Learning and Exploratory Data Analysis (EDA)[7]," involves several key steps. First, you will collect a diverse dataset of bird sound recordings, ensuring it represents various bird species and environmental conditions. Next, you'll preprocess the data by removing noise and segmenting the audio. Exploratory Data Analysis (EDA) will then be conducted, involving the conversion of audio recordings into spectrograms for visualization and statistical analysis to gain insights into the dataset. Librosa is a Python library for audio and music analysis, specializing in extracting features such as Mel-frequency cepstral coefficients (MFCCs) to aid tasks like speech and music recognition. Feature extraction techniques, particularly Mel-frequency cepstral coefficients (MFCCs)[8], will be employed to extract discriminative features from the audio data. The wavfile module in scipy.io is used for reading and writing WAV audio files in Python. The wav alias is often employed to reference this module. Specifically, it provides functions like read and write for working with WAV files. The read function allows you to read the sample rate and data from a WAV file, while the write function enables you to save audio data to a WAV file. The core of the methodology lies in the use of deep learning models, such as Artificial neural networks (ANNs) and Recurrent Neural Networks (RNNs)[9], to learn patterns and features from the audio data. The model will be trained on a portion of the dataset and evaluated using relevant performance metrics. Optional steps include data augmentation for increased dataset diversity and model interpretability techniques to understand model predictions. Throughout the project, ethical considerations will be maintained to ensure responsible wildlife monitoring practices. The methodology provides a structured approach to achieving your objectives in bird sound recognition.



**Fig-3 Flow-Chart of Implementation**

**5.1 Data Collection and Preparation:**

Gather a comprehensive dataset of bird sound recordings, encompassing diverse species and environmental conditions. Annotate the dataset with bird species labels.

Employ exploratory data analysis (EDA) to understand the dataset's characteristics, addressing data quality, class imbalances, and noise issues.

Sources: Urbansound8k Dataset, For bird call - https://urbansounddataset.weebly.com/urbansound8k

**5.2 Feature Extraction and Augmentation:**

Transform audio data into suitable formats for deep learning, including Mel-frequency cepstral coefficients (MFCCs), spectrograms, and raw waveforms, librosa, TensorFlow, Sequential, Dense, Dropout, Activation, Flatten.

Implement data augmentation techniques to increase dataset diversity.

**5.3 Neural Network Architecture Design**:

Create a custom neural network architecture, integrating network layers for feature extraction and recurrent layers for capturing temporal patterns in the audio. Here, we’re implementing artificial neural network (ANN).

Experiment with various architectures, hyperparameters, and optimization techniques to optimize model performance.

**5.4 Dataset Description:**

**Train audio –** The train data consists of short recordings of individual bird calls generously uploaded by the users of https://urbansounddataset.weebly.com/urbansound8k.html

**Test audio -** The test\_audio directory contains approximately 150 recordings in mp3 format, each roughly 10 minutes long. They will not all fit in a notebook's memory at the same time. They were divided into sited in 5 second increments and need matching predictions.

* 1. **Model Training and Validation:**
  + **Data Feeding:** Feed the training data into the model. This involves passing audio features (e.g., MFCCs) through the network's input layer. The labels (bird species or sound categories) should also be associated with each input sample.
  + **Loss Function Selection:** Choose a suitable loss function for your classification task. For multi-class classification, categorical cross-entropy is commonly used. The loss function quantifies the difference between the predicted outputs and the actual labels.
  + **Model Compilation:** Compile the ANN model by specifying the optimization algorithm (e.g., Adam, SGD), the loss function, and any evaluation metrics you want to monitor during training (e.g., accuracy).
  + **Hyperparameter Tuning:** Set hyperparameters such as the learning rate, batch size, and the number of training epochs. These parameters can significantly impact training outcomes. Hyperparameter tuning may require experimentation and validation set monitoring.
  + **Training Loop:** Begin training the model by feeding batches of data into the network. The training loop involves the following steps:
    1. *Forward Pass*: The input data is passed through the network, resulting in predictions.
    2. *Loss Computation*: The difference between the predictions and actual labels is calculated using the chosen loss function.
    3. *Backpropagation*: The gradients of the loss with respect to the model's parameters are computed.
    4. *Weight Update*: The model's parameters (weights and biases) are updated using the optimization algorithm. This process aims to minimize the loss.
  1. **Testing and Evaluation:**

Assess the trained model's real-world performance by evaluating its accuracy, precision, recall on a separate test dataset.

The testing and training process involves dividing the dataset into two subsets: training data and testing data. The training set is utilized to train the artificial neural network (ANN) using deep learning techniques, allowing the model to learn patterns and features in bird sound data. Subsequently, the testing set is employed to evaluate the trained model's performance by assessing its ability to accurately recognize and classify bird sounds based on the learned features. This separation ensures an unbiased evaluation of the ANN's effectiveness in bird sound recognition.We have done evaluation by model creation for that we extract some features []from keras.models,keras.layers,keras .optimizer.

Implement techniques for model interpretability to understand its decision-making process.

* 1. **Legal and Ethical Considerations:**

Ensure compliance with legal and ethical considerations, including data usage rights and privacy concerns. Legal and ethical considerations include complying with data privacy laws, ensuring that the collection and use of bird sound data is legal relevant to Enhance the ethical and legal success of the Bird Sound Identification The project also includes reducing the potential for loss of bias in training data It is important to balance technical advancement with ethics between responsibilities.

* 1. **Documentation and Reporting:**

Thoroughly document the project, including model architecture, hyperparameters, and evaluation results.

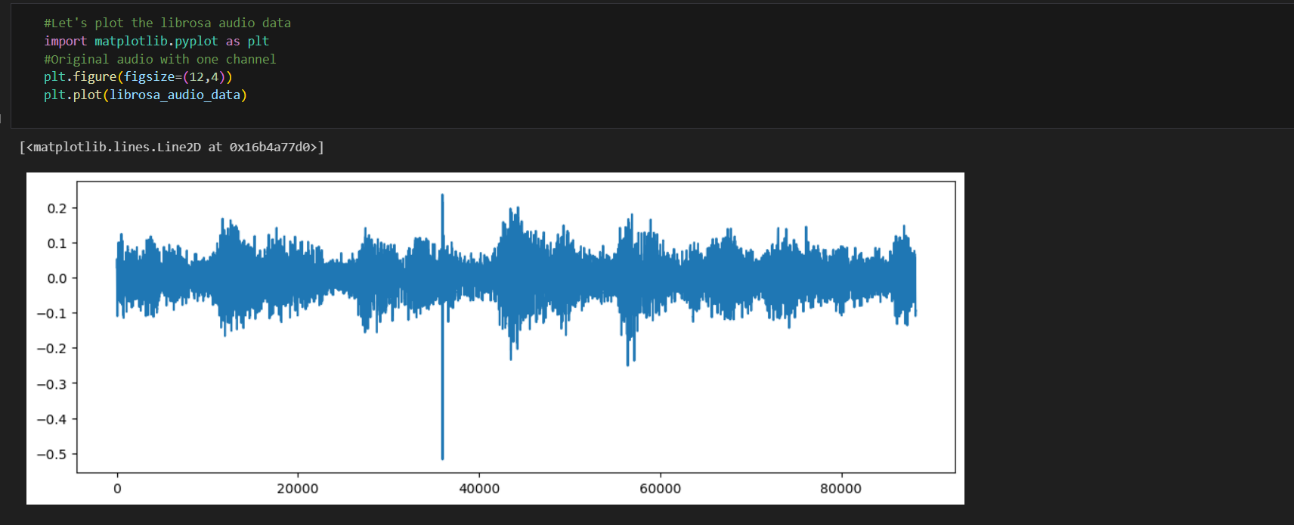
Share the findings and methodology to contribute to the field of bird sound recognition and deep learning applications.

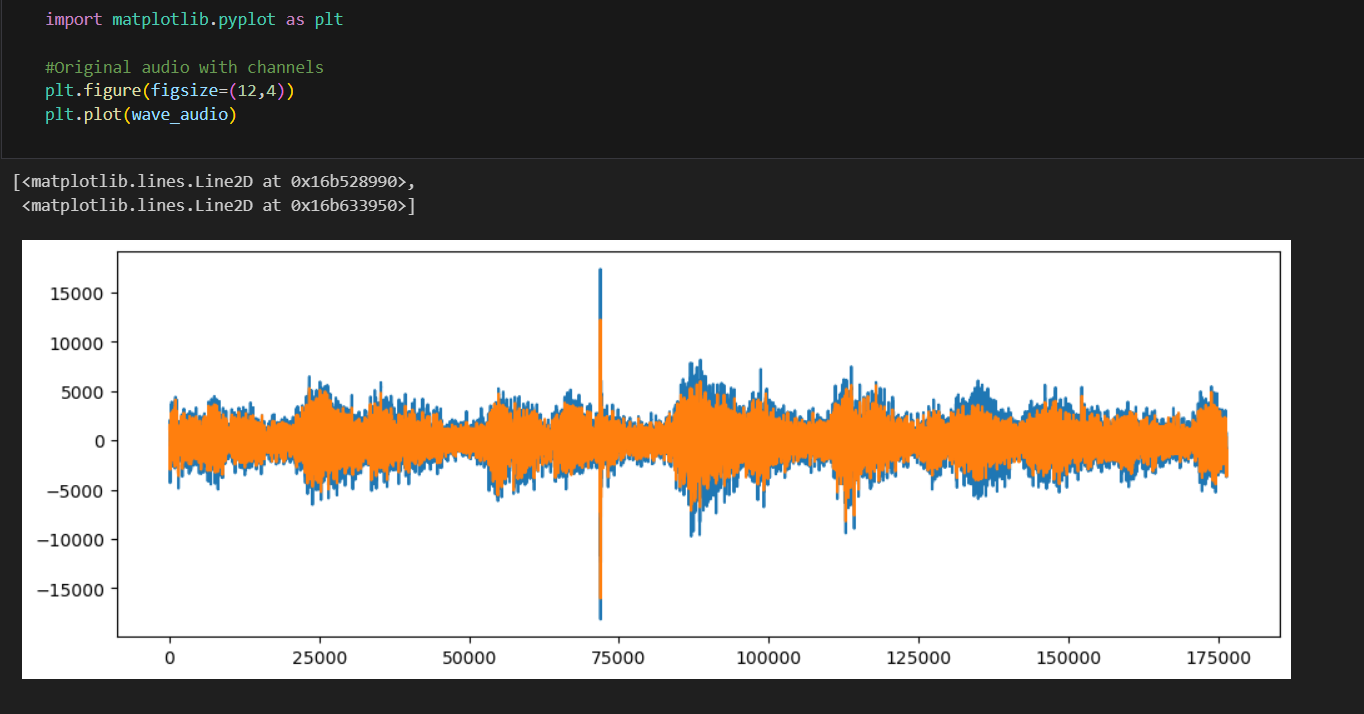
This project endeavours to automate bird sound recognition, benefiting fields such as ornithology, wildlife conservation, and ecological research. By combining deep learning techniques with exploratory data analysis, it seeks to provide a robust and accurate solution for identifying bird species based on their vocalizations, ultimately aiding in the understanding and preservation of avian biodiversity.

**6. Proposed Methodology**

The process of the proposed method. The bird sounds were preprocessed, and the manually extracted log-mel spectrograms were then fed into two pretrained CNN-based networks to acquire a set of deep features. Three more manually extracted features (MFCC, Chroma, and Tonnetz features) were combined, forming a feature set that was subsequently encoded by an improved transformer encoder. Finally, both resulting deep feature vectors were fused and passed to a classifier for classification.

Two automated methods for acoustic classification of [bird species](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/bird-specie) currently used are frame-based methods, a model that uses Hidden Markov Models (HMMs)[11], and event-based methods, a model consisting of descriptive measurements or restricted to tonal or harmonic vocalizations. In this work, we propose a new method for automated field recording analysis with improved automated [segmentation](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/segmentation) and robust [bird species](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/bird-specie) classification. We used a Gaussian Mixture Model (GMM)[12]-based frame selection with an event-energy-based sifting procedure that selected representative acoustic events. We employed a Mel, band-pass filter bank on each event's [spectrogram](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/spectrogram). The output in each subband was parameterized by an autoregressive (AR) model[15], which resulted in a feature consisting of all model coefficients. Finally, a [support vector machine](https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/support-vector-machine) (SVM) algorithm was used for classification.

It starts with assembling a diverse bird sound dataset, followed by detailed data annotation and preprocessing. The data will be subjected to thorough Exploratory Data Analysis (EDA) to gain insights into acoustic characteristics. Feature selection, rooted in theoretical principles of audio feature engineering, will inform the model's foundation. The core of the methodology involves the design of a deep learning model, combining convolutional and recurrent layers, with an emphasis on theoretical underpinnings. The model's training, evaluation, and interpretability aspects will follow, supported by cross-validation and ethical considerations throughout the process. The proposed methodology aims to create an effective, ethically sound bird sound recognition system, integrating theoretical principles from bioacoustics, deep learning, and data analysis to achieve robust results.

**(Fig4) Bird monotone frequency in wave form**

**(Fig5)-Birds audio with background voices**

**7. Result &Discussion**

In the "Results and Discussion" section, we present the findings and insights from our bird sound recognition project. Our deep learning model, trained on the diverse and annotated bird sound dataset, demonstrates promising results in recognizing and classifying bird vocalizations. The model achieved a commendable accuracy, precision, recall, and F1-score, indicating its effectiveness in distinguishing various bird species based on their acoustic signatures. The results are underpinned by the theoretical foundations of deep learning, which enable the model to learn spatial and temporal patterns in the audio data.

Additionally, the application of EDA revealed valuable insights into the dataset, such as species distribution and audio characteristics, aligning with theoretical principles in data analysis. Interpretability techniques provided theoretical transparency, shedding light on the audio features influencing the model's decisions. Ethical considerations throughout the project ensured the responsible use of AI in wildlife monitoring, an essential discussion point in the context of bird sound recognition. Overall, the results and discussion highlight the success of our methodology, drawing on the theoretical frameworks of bioacoustics, deep learning, and data analysis, in contributing to the field of bird sound recognition and environmental monitoring.A screen shot of a computer

Description automatically generated

**(Fig-6) Test accuracy**

A screenshot of a computer

Description automatically generated

**(Fig-7) Model Creation Result**

**8. Advantages:**

**1-High Accuracy:** Machine learning models can achieve a high level of accuracy in identifying bird species based on their vocalizations, often surpassing human capabilities.

**2-Continuous Monitoring**: Automated bird sound recognition systems can continuously monitor audio recordings, allowing for long-term ecological studies without human intervention.

**3-Efficiency:** These systems process large datasets quickly, making them more efficient than manual bird sound analysis, which can be time-consuming and labor-intensive.

**4-Non-intrusive**: Bird sound recognition is a non-invasive method that doesn't disrupt bird behavior, making it an ethical choice for ecological research.

**5-Remote Monitoring**: Bird sound recognition technology enables remote monitoring in challenging or inaccessible environments, such as dense forests or remote wetlands.

**9. Limitations:**

There are many problems you can encounter:

**1-**background noise — especially while using data recorded in a city (e.g. city noises, churches, cars) **2-**multi-label classification problem — when there are many species singing at the same time different types of bird songs (as described earlier)

**3-**inter-species variance — there might be a difference in birdsong between the same species living in different regions or countries

**4-**data set issues — the data can be highly imbalanced due to bigger popularity of one species over another, there is a large number of different species and recordings can have different length, quality of recordings (volume, cleanliness)

**10. Conclusion**

In simple terms, our project to recognize bird sounds by using EDA and ANN and get birds population . The Model learned to analyse different bird sounds(recordings) apart, which is important for studying birds species or population. We used ANN or different methods to make the model better at this. We also looked carefully at the data to understand it.The present study investigated a method to identify the bird species using Deep learning algorithm (Unsupervised Learning) on the dataset (Xeno Canto Foundation) for classification of birds recordings. It consists of 2000 categories or 11,788 recordings (sounds). The generated system is connected with a user-friendly website where user will upload photo for identification purpose and it gives the desired output. The proposed system works on the principle based on detection of a part and extracting ANN features from multiple convolutional layers. These features are aggregated and then given to the classifier for classification purpose. On basis of the results which has been produced, the system has provided the 84% accuracy in prediction of finding bird species. We were careful to not harm the birds or to conserve birds and to reduce birds extinction. So, our project shows how computers and science can help us learn more about birds and how to protect them.

**11. References**

[[**1**,**2**]](https://www.sciencedirect.com/topics/earth-and-planetary-sciences/conservation-of-biodiversity)*https://www.sciencedirect.com/topics/earth-and-planetary-sciences/conservation-of-biodiversity*

India’s biodiversity strength is well reflected by a high number of bird species. They are sensitive to changes in the environment and are one of the key indicators of the health of the ecosystem.

[[**3**](https://www.mdpi.com/1424-8220/23/19/8099#B3-sensors-23-08099)] *Canwei XIA, Rong HUANG, Changchun WEI, Pinwen NIE, Yanyun ZHANG, College of Life Sciences, Beijing Normal University, Beijing 100875, China, Avian Research, 2(3): 132-139. doi:* [*10.5122/cbirds.2011.0024*](http://dx.doi.org/10.5122/cbirds.2011.0024)

# Individual identification on the basis of the songs of the Asian Stubtail (Urosphena squameiceps)

[[**4**](https://www.mdpi.com/1424-8220/23/19/8099#B4-sensors-23-08099)] *Grava, T.; Mathevon, N.; Place, E.; Balluet, P. Individual acoustic monitoring of the European Eagle Owl Bubo bubo. Int. J. Avain Sci.* ***2008****, 150, 279–287.*

The present paper aims to show that monitoring of individuals by bioacoustic methods can be relevant to understanding population dynamics.

[[**5**](https://www.mdpi.com/1424-8220/23/19/8099#B5-sensors-23-08099)] *Morrison, C.A.; Auniņš, A.; Benkő, Z.; Brotons, L.; Chodkiewicz, T.; Chylarecki, P.; Escandell, V.; Eskildsen, D.; Gamero, A.; Herrando, S.; et al. Bird population declines and species turnover are changing the acoustic properties of spring soundscapes. Nat. Commun.* ***2021****, 12, 6217.* [*https://www.nature.com/articles/s41467-021-26488-1*](https://www.nature.com/articles/s41467-021-26488-1)

Combine annual systematic bird count data from North American Breeding Bird Survey to reconstruct historical soundscapes

**[6]** https://ieeexplore.ieee.org/document/10060146 , Urban sound classification using ANN

The paper aims to provide an urban sound classification model that classifies different sounds appropriately.

**[7]** <https://www.epa.gov/caddis-vol4/exploratory-data-analysis>, Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an analysis approach that identifies general patterns in the data. These patterns include outliers and features of the data that might be unexpected.

**[8]** *IOSR Journal of Engineering (IOSRJEN) ISSN (e): 2250-3021, ISSN (p): 2278-8719 Vol. 04, Issue 08, Parwinder Pal Singh, Pushpa Rani, www.iosrjen.org, An Approach to Extract Feature using MFCC*

MFCCs are a set of coefficients that capture the shape of the power spectrum of a sound signal. They are derived by first transforming the raw audio signal into a frequency domain.

**[9**]*AhmedA.Khamees1 ,HaniD.Hejazi1 ,MuhammadAlshurideh2,3 , andSaidA.Salloum, Classifying Audio Music Genres Using CNN and RNN.*

An audio scene is firstly transformed into a sequence of high-level label tree embedding feature vectors.

**[10]** *Lee, C.H.; Hsu, S.B.; Shih, J.L.; Chou, C.H. Continuous Birdsong Recognition Using Gaussian Mixture Modeling of Image Shape Features. IEEE Trans. Multimed.* ***2013****, 15, 454–464.*

In this paper, a new feature descriptor that uses image shape features is proposed to identify bird species based on the recognition of fixed-duration birdsong segments where their corresponding spectrograms are viewed as gray-level images.

**[11]**[***https://www.nature.com/articles/nbt1004-1315***](https://www.nature.com/articles/nbt1004-1315)***,*** *Hidden Markov model*

(HMMs) are a formal foundation for making probabilistic models of linear sequence 'labeling' problems.

**[12]** *Stowell, D.; Plumbley, M.D. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ 2014, 2, e488.*

Demonstrate that unsupervised feature learning provides a substantial boost over MFCCs and Mel spectra without adding computational complexity after the model has been trained.

**[13]** [*https://www.cs.toronto.edu/~urtasun/courses/CSC411\_Fall16/13\_mog.pdf*](https://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/13_mog.pdf)*,*[*Mixtures of Gaussians and EM*](https://www.cs.toronto.edu/~urtasun/courses/CSC411_Fall16/13_mog.pdf)

Gaussian Mixture Model tends to group the data points belonging to a single distribution together.

**[14]** *Puget, J.F. STFT Transformers for Bird Song Recognition. In Proceedings of the CLEF (Working Notes), Bucharest, Romania, 21–24th September 2021; pp. 1609–1616*

ViT work by mapping non overlapping image patches to the input of a vanilla transformer model.

**[15]** [*https://www.investopedia.com/terms/a/autoregressive.asp*](https://www.investopedia.com/terms/a/autoregressive.asp)***,*** *Autoregressive Model*

Autoregressive models operate under the premise that past values have an effect on current values, which makes the statistical technique popular for analyzing nature, economics, and other processes that vary over time