e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 3, March 2024

Impact Factor: 7.521

6381 907 438 6381 907 438 ijmrset@gmail.com www.ijmrset.com

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

Identification of Bird Species Using LSTM with MFCC across A Multilingual Framework

R.Vidhya[1], S.B.Janani[2], R.Abirami[3],M.Ashwinth[4]

Assistant professor, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamilnadu, India[1]

UG Students, Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamilnadu, India[2,3,4]

ABSTRACT: Birds, with their diverse vocalizations, significantly contribute to the rich acoustic tapestry of natural environments, offering valuable insights into biodiversity and ecosystem health. This study introduces an innovative approach to "Bird Species Classification Using Sound," integrating a multilingual bird name translator to enhance accessibility and inclusivity. By leveraging advancements in machine learning, acoustic analysis, and deep learning, the research aims to automatically identify and classify bird species based on their vocalizations. The proposed system utilizes audio recording devices to capture bird sounds in various habitats, while a multilingual translator facilitates seamless communication by converting bird species names into different languages.These audio recordings undergo preprocessing to extract relevant acoustic features like Mel-frequency cepstral coefficients (MFCC). A deep learning model, specifically Recurrent Neural Networks (RNNs), is then trained on these features to recognize and classify bird species. A substantial dataset covering a wide range of avian species and sound environments is employed for model training and validation. The study's contributions extend to developing an efficient and reliable bird species classification system with potential applications in ornithology, ecological research, and biodiversity monitoring. Moreover, the project aims to foster increased understanding and appreciation of avian diversity and its role in maintaining ecosystem balance, promoting inclusivity and engagement among diverse linguistic communities through the multilingual translator integration.

KEYWORDS: MFCC, Recurrent Neural Network (RNN), Deep Learning, Acoustic Analysis, Multilingual Translator, Biodiversity Monitoring

I.INTRODUCTION

The symphony of bird calls permeates the natural world, offering both beauty and valuable ecological insights. In our project, we embark on a journey to unlock the secrets of avian diversity through the analysis of bird voices using advanced audio signal processing and deep learning techniques. Leveraging advancements in deep learning, particularly Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) layers, this project addresses the challenge of analyzing vast amounts of audio data containing bird vocalizations. By leveraging these deep learning models, we aim to develop a robust and accurate system for automatically identifying bird species based solely on their vocalizations. Birdsong serves as a unique identifier for different species, with each bird boasting its own distinctive repertoire of calls and melodies. Through the lens of audio signal processing, we dissect these vocalizations into meaningful features, extracting patterns and characteristics that distinguish one species from another.

This project involves collecting audio recordings of bird vocalizations from different environments such as forests, wetlands, and urban areas. The collected audio data is then preprocessed to extract relevant features that capture the unique characteristics of each species' vocalizations. Our project not only showcases the capabilities of deep learning in the realm of avian acoustics but also underscores the importance of interdisciplinary collaboration between the fields of computer science and ornithology. By bridging these domains, we pave the way for innovative approaches to biodiversity monitoring and conservation.

Once trained, the model can be deployed for real-world applications, such as biodiversity monitoring, species inventorying, and habitat assessment. Automated bird species identification systems can assist researchers in collecting and analyzing large volumes of data efficiently, providing valuable insights into avian ecology and facilitating

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

evidence-based conservation strategies. Additionally, these systems can engage citizen scientists and bird enthusiasts in conservation efforts by enabling them to contribute to scientific research and monitoring initiatives.

II.LITERATURE REVIEW

You-Peng Sun[1],Ying Jiang, Zheng Wang[2], Yuan Zhang[3], Liulei Zhang[4] proposed **"Wild Bird Species Identification Based on a Lightweight Model With Frequency Dynamic Convolution"**. This paper involves a lightweight model designed for efficiency, particularly suitable for real-time bird identification in the field on devices with limited resources. The key innovation lies in the use of frequency dynamic convolution. Unlike traditional methods that assume features are independent of their location in the sound spectrogram, this approachallows the model to capture crucial variations in features across different frequency bands within the spectrogram. This focus on frequency dynamics holds promise for improved accuracy in differentiating bird species. One crucial challenge lies in ensuring generalizability across diverse environments. Bird vocalizations can exhibit variations due to factors like geography, bird age, and context. The model's accuracy might be impacted by background noise levels or unexpected variations in bird calls that differ from the data used for training.

Xiaomei Yi[1], Chengqian[2], Peng Wu[3], Tengteng Jiang[4], Wenying Ge[4] proposed **"Research on Fine-Grained image Recoqnition of Birds Based on Improved YOLOv5".** Unlike traditional methods that analyze the entire bird, their strategy focuses on a part-based recognition technique,this project involves pinpointing specific regions of the bird crucial for differentiation, such as the head, wings, and tail. The project leverages an improved version of the YOLOv5 object detection algorithm, known for its speed and accuracy. This part- based approach holds promise for more precise bird identification, while the use of the efficient YOLOv5 algorithm allows for faster image processing. One key constraint lies in this project is handling occlusions. If parts of the bird, particularly the crucial regions like the head or wings, are obscured by foliage, branches, or other objects, it can significantly hinder accurate part detection and consequently, species classification.

Aymen Saad[1], Javed Ahmed[2], Ahmed Elaraby[3] proposed **"Classification of Birds Sound Using High and Low Complexity Convolutional Neural Networks."** This research focused on the trade-off between model complexity and accuracy. The study employed CNNs, which excel at extracting features from spectrograms (visual representations of sound) generated from bird vocalizations. Two spectrogram generation methods were explored: Short-Time Fourier Transform (STFT) and Mel-Frequency Cepstral Coefficients (MFCCs), which captures human-like auditory features. Interestingly, the project compared the performance of a high-complexity CNN, ResNet-50, known for high accuracy but demanding significant computational resources, with MobileNetV1, a low-complexity CNN designed forefficiency on mobile devices. By evaluating the accuracy of bird species classification using both CNNs with STFT and MFCC spectrograms, the study aimed to determine if the more efficient MobileNetV1 could achieve comparable results to the computationally expensive ResNet-50.

Alexandru-Marius Solomes[1] and Dan Stowell [2] proposed **"Efficient Bird Sound Detection On The Bela Embedded System."** Traditional bird monitoring methods often require bulky equipment and can be limited in terms of deployment and data collection. To address these limitations,they proposed a novel approach for bird sound detection using the Bela embedded system. This compact and energy-efficient platform makes it suitable for long-term, remote field deployments. The project leverages machine learning algorithms to automatically detect bird sounds within audio recordings captured by the Bela system. Here the processing power and memory capacity may restrict the complexity of machine learning models that can be deployed.

Bekir Kabasakal[1], Duygugul Aksu[2], Nusret Demir[3], Melih Oz[4] and Ali Erdogan[4] proposed **"Automated Bird Counting with Deep Learning for Regional Bird Distribution Mapping."** Their project leverages Convolutional Neural Networks (CNNs) to automatically detect birds in images, eliminating the need for manual counting and improving efficiency and consistency. Furthermore, the detected bird counts are integrated with Geographic Information Systems (GIS) to create regional bird distribution maps. These maps provide valuable insights into bird population dynamics and habitat preferences across different areas. Though it promises advancement in bird counting, it inhibits some limitations. CNNs may struggle to differentiate between similar-looking bird species or those partially obscured by foliage.

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

III.METHODOLOGY OF PROPOSED SURVEY

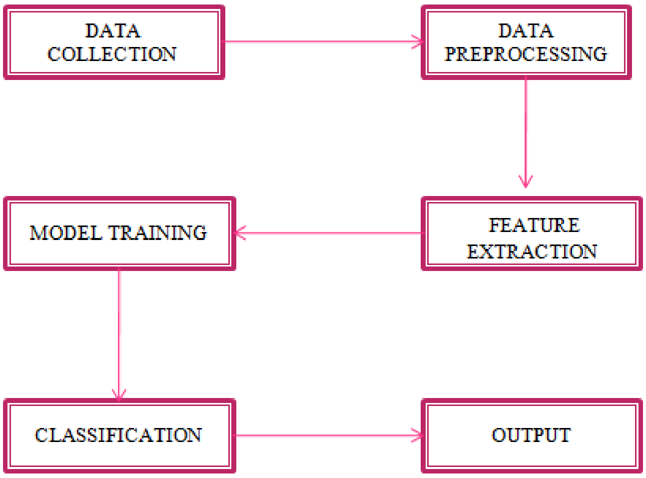

Figure 1: Workflow of methodology

DATA COLLECTION

The first step of implementation is gathering data from dataset which is obtained from XENO-canto/Kaggle. The audio recordings of the birds in MP3 format are included in this resource This dataset comprises MP3-format audio recordings capturing the sounds of various bird species. XENO- canto/Kaggle are open websites dedicated for dataset where users upload their own recordings.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	primary_l	secondary	type	latitude	longitude	scientific_	common_	author	license	rating	url	filename								
2	abethr1	[]	['song']	4.3906	38.2788	Turdus te	African Ba	Rolf A. de	Creative C	4	https://w	abethr1/XC128013.ogg								
3	abethr1	[]	['call']	-2.9524	38.2921	Turdus te	African Ba	James Bra	Creative C	3.5	https://w	abethr1/XC363501.ogg								
4	abethr1	[]	['song']	-2.9524	38.2921	Turdus te	African Ba	James Bra	Creative C	3.5	https://w	abethr1/XC363502.ogg								
5	abethr1	[]	['song']	-2.9524	38.2921	Turdus te	African Ba	James Bra	Creative C	5	https://w	abethr1/XC363503.ogg								
6	abethr1	[]	['call', 'sor	-2.9524	38.2921	Turdus te	African Ba	James Bra	Creative C	4.5	https://w	abethr1/XC363504.ogg								
7	abethr1	['rbsrob1']	['song']	-2.9524	38.2921	Turdus te	African Ba	James Bra	Creative C	3.5	https://w	abethr1/XC379322.ogg								
8	abethr1	[]	['call', 'sor	-2.9965	37.6244	Turdus te	African Ba	isaac kilus	Creative C	3	https://w	abethr1/XC432639.ogg								
9	abethr1	[]	['song']	-4.0904	37.8807	Turdus te	African Ba	Peter Eric	Creative C	5	https://w	abethr1/XC467121.ogg								
10	abethr1	[]	['song']	-4.0904	37.8807	Turdus te	African Ba	Peter Eric	Creative C	5	https://w	abethr1/XC467122.ogg								
11	abethr1	[]	['adult', 's	4.8403	38.6988	Turdus te	African Ba	Barry Edm	Creative C	4	https://w	abethr1/XC531557.ogg								
12	abethr1	[]	['song']			Turdus te	African Ba	isaac kilus	Creative C	4	https://w	abethr1/XC585802.ogg								
13	abethr1	['eswdov1	['song']	-2.8802	38.1861	Turdus te	African Ba	Sidney Sh	Creative C	4.5	https://w	abethr1/XC606253.ogg								
14	abethr1	[]	['adult', 's	-2.9858	37.5975	Turdus te	African Ba	isaac kilus	Creative C	4	https://w	abethr1/XC616997.ogg								
15	abethr1	[]	['adult', 'c	-2.9858	37.5975	Turdus te	African Ba	isaac kilus	Creative C	5	https://w	abethr1/XC639039.ogg								
16	abethr1	[]	['']	-3.1552	40.1326	Turdus te	African Ba	Lars Lachn	Creative C	3	https://xe	abethr1/XC756300.ogg								
17	abhori1	['combul2	['call', 'sor	-15.9259	29.0292	Oriolus la	African Bl	Rory Nefd	Creative C	3.5	https://w	abhori1/XC120250.ogg								
18	abhori1	['rindov']	['call', 'sor	-15.9259	29.0292	Oriolus la	African Bl	Rory Nefd	Creative C	4.5	https://w	abhori1/XC120251.ogg								
19	abhori1	['blbpuf2',	['song']	-5.7214	37.9942	Oriolus la	African Bl	David Mo	Creative C	3.5	https://w	abhori1/XC127317.ogg								
20	abhori1	['fotdro5']	['song']	-5.7214	37.9942	Oriolus la	African Bl	David Mo	Creative C	3.5	https://w	abhori1/XC127318.ogg								
21	abhori1	[]	['song']	5.258	39.697	Oriolus la	African Bl	Rolf A. de	Creative C	5	https://w	abhori1/XC128202.ogg								
22	abhori1	['chibat1',	['song']	-2.034	37.38	Oriolus la	African Bl	James Bra	Creative C	2.5	https://w	abhori1/XC132733.ogg								
23	abhori1	['gvcwar3'	['song']	0.064	32.479	Oriolus la	African Bl	Bram Piot	Creative C	3.5	https://w	abhori1/XC138433.ogg								

Figure 2:Overview of Dataset

DATA PREPROCESSING

Following data collecting, sound recordings are preprocessed. Data preprocessing encompasses a series of steps to clean, format, and standardize the collected audio data before further analysis. Data preprocessing is a critical stage involving a systematic sequence of operations to refine and standardize the collected audio data prior to analysis. Initially, raw recordings are formatted into a standardized file type, typically WAV, to ensure uniformity across the dataset. WAV files are uncompressed, meaning they preserve the original audio data without any loss of quality. This is crucial for accurate MFCC extraction, as MFCCs rely on the detailed spectral information of the sound. MP3s, on the

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

other hand, use compression techniques that can introduce artifacts and potentially affect the accuracy of feature extraction.

FEATURE EXTRACTION

Feature extraction involves capturing essential information from raw audio data to represent it in a format suitable for analysis and classification. One common technique is to extract Mel-Frequency Cepstral Coefficients (MFCCs), which are widely used in speech and audio processing tasks. By transforming the spectral information into the mel-frequency scale and then computing cepstral coefficients, we can effectively capture the unique timbral qualities of different bird calls. MFCCs offer several advantages, including robustness to variations in background noise and insensitivity to slight shifts in pitch. The MFCC-DNN framework uses the librosa python package to compute 40-dimensional MFCCs with a frame size of 30ms and a hop size of 10ms.

MODEL TRAINING

Model training involves using labeled data to train machine learning or deep learning models to recognize patterns and relationships between input features and target labels. In our project on bird species identification through sound recordings, we employ Long Short-Term Memory (LSTM) neural networks for model training.. The extracted MFCCs serve as input features for training the LSTM model, which consists of multiple LSTM layers and fully connected layers for classification. During training, the model minimizes the categorical cross-entropy loss using the Adam optimizer. Experimental results demonstrate the effectiveness of the trained LSTM model in accurately identifying bird species from audio recordings. The model achieves high classification accuracy on a validation dataset and exhibits robust performance on a separate test dataset. Overall, the trained LSTM model shows promise for automated bird species identification tasks, offering potential applications in biodiversity monitoring and conservation efforts.

ALGORITHM EXPLANATION

The LSTM (Long Short-Term Memory) neural network architecture is a form of recurrent neural network (RNN) specifically engineered to overcome the challenge of the vanishing gradient problem encountered in conventional RNNs. It consists of memory cells that can maintain information over long sequences, making it well-suited for processing sequential data. In this project, the LSTM network is utilized to analyze temporal patterns in bird vocalizations captured in audio recordings. Each LSTM layer processes the input sequences of MFCC features and learns to capture relevant temporal dependencies. The output from the last LSTM layer is then fed into fully connected layers for classification, where the model predicts the bird species based on learned patterns in the input data. The Long Short-Term Memory (LSTM) architecture utilized in this bird species identification system is a type of recurrent neural network (RNN) specifically designed to model sequential data while addressing the vanishing gradient problem typically encountered in traditional RNNs. LSTM networks consist of memory cells with self-connected units called gates, which regulate the flow of information within the network.The LSTM architecture used here includes:

1. Input Shape Configuration: The LSTM layer is configured to accept input sequences of MFCC features with a specific shape (40,1), where 40 represents the number of MFCC coefficients extracted from each time frame, and 1 indicates the number of channels (typically mono audio).
2. LSTM Units: The LSTM layer consists of 256 memory units, enabling the network to capture complex temporal patterns and dependencies within the input sequences effectively.
3. Dropout Regularization: Dropout layers with a dropout rate of 0.2 are employed to prevent overfitting by randomly deactivating a fraction of neurons during training.
4. Dense Layers: Following the LSTM layer, densely connected layers with rectified linear unit (ReLU) activation functions are added to capture higher-level representations and facilitate classification.
5. Output Layer: The final output layer consists of four units corresponding to the number of bird species classes, with a softmax activation function for multi-class classification.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_1 (LSTM)               (None, 256)               264192

 dropout_3 (Dropout)         (None, 256)               0

 dense_3 (Dense)             (None, 128)               32896

 dropout_4 (Dropout)         (None, 128)               0

 dense_4 (Dense)             (None, 64)                8256

 dropout_5 (Dropout)         (None, 64)                0

 dense_5 (Dense)             (None, 2)                 130

=================================================================
Total params: 305474 (1.17 MB)
Trainable params: 305474 (1.17 MB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 3: Training Result

Figure 4: Validation Result

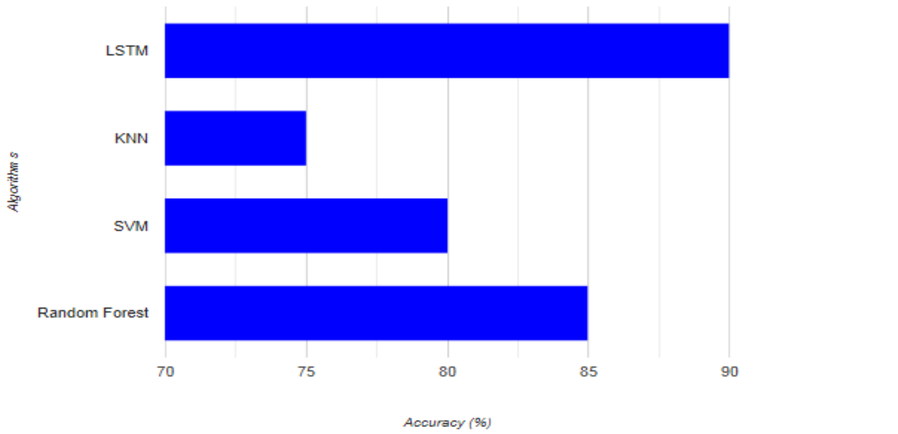

Figure 5 : Algorithms Accuracy Comparison

International Journal Of Multidisciplinary Research In Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

IV. RESULTS AND DISCUSSIONS

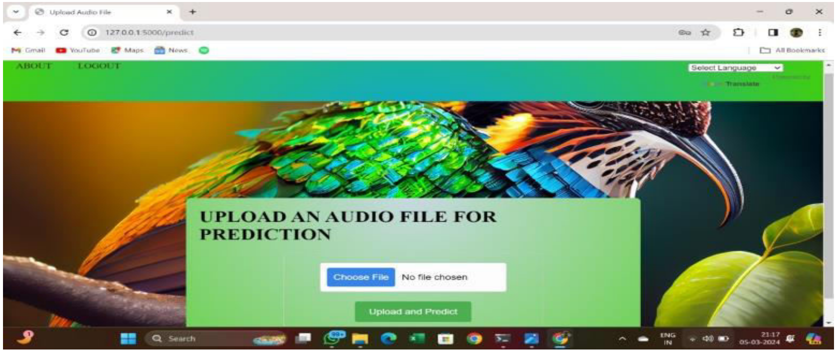

Figure 6 : Audio Upload Page

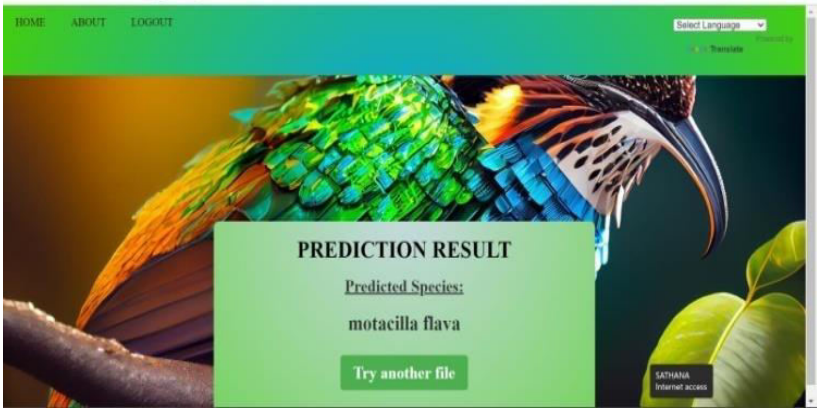

Figure 7: Bird Species Prediction Result

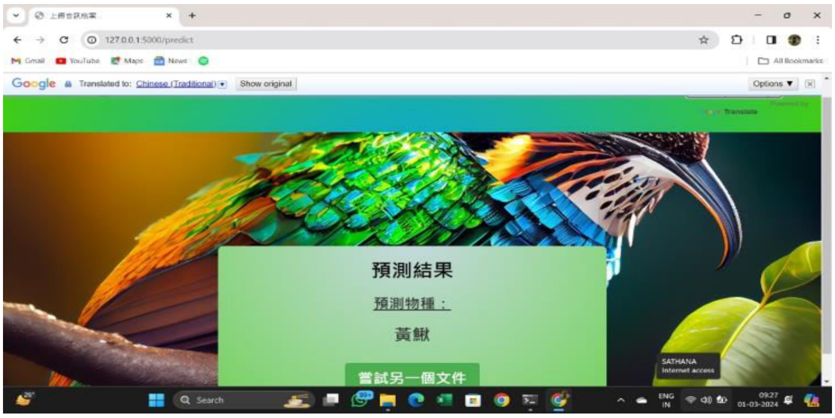

Figure 8 : Prediction Result In Other Language

In our bird species identification project utilizing sound analysis, we attained remarkable success with a high classification accuracy.This high accuracy was achieved through meticulous preprocessing of bird vocalizations and the implementation of sophisticated feature extraction techniques, primarily Mel-frequency cepstral coefficients (MFCCs). By leveraging machine learning algorithms trained on these extracted features, our system effectively distinguished between different bird species based on their unique sound signatures.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly Peer Reviewed & Referred Journal |

| Volume 7, Issue 3, March 2024 |

| DOI:10.15680/IJMRSET.2024.0703033 |

V. CONCLUSION AND FUTURE WORK

This paper has successfully distinguished four distinct bird species.We developed a bird species identification system leveraging Mel-frequency cepstral coefficients (MFCC) extracted from bird sounds. The process involved preprocessing the bird's vocalizations and extracting MFCC features, which were then used to train the classification model. The training dataset comprised real bird sounds recorded in their natural habitats, alongside other environmental noises. Focusing on a smaller number of species allows for targeted training and potentially high accuracy within that scope. In the realm of bird species identification through sound analysis, several promising directions for future work emerge, aiming to enhance the accuracy, scalability, and practicality of identification systems. One critical area of focus involves the refinement of machine learning algorithms and feature extraction techniques tailored specifically for avian vocalizations. By delving deeper into the spectral and temporal characteristics of bird calls, researchers can develop more nuanced models capable of distinguishing between similar species with greater precision.

REFERENCES

[1] 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.

[2] Yu, G., Wang, L., Cao, C., Wang, J., & Zhao, S. “A Survey on Deep Learning for Audio Scene Classification”. IEEE Access, vol 11, 2023, 106620 – 106649.

[3]Dr. Amol Dhakne, Vaishnav M. Kuduvan, Aniket Palhade, Tarun Kanjwani, Rushikesh Kshirsagar. “Bird Species Identification using Audio Signal Processing and Neural Networks”. International Journal For Research in Applied Science and Engineering Technology(IJRASET) ,vol 10, no. 05, 2022.

[4] A V Siva Krishna Reddy, Dr.M A Srinuvasu, K Manibabu, Ch B V Sai Krishna, D Jhansi. “Image Based Bird Species Identification Using Deep Learning”. International Journal of Creative Research Thoughts (IJCRT) ,vol 9, 2021.

[5] Aska E. Mehyadin, Adnan Mohsin Abdullazeez, Dathatar Abas Hasan. “ Bird Sound Classification Based On Machine Learning Algorithms”. Asian Journal Of Research In Computer Science,2021.

[6] Jie Xie, Kai Hu, Mingying Zhu, Jinghu Yu, Qibing Zhu. “Investigation Of Different CNN Based Models For Improved Bird Sound Classification”. IEEE Access,vol 7, 2019.

[7] Jiri Stastny, Michal Munk, Lubos Juranek. “Automatic bird species recognition based on birds vocalization”. EURASIP Journal On Audio,Speech, and Music Processing,2018.

[8] Nirosha Priyadarshani, Stephen Marsland, Isabel Castro. “Automated birdsong recognition in complex acoustic environments”. Journal Of Avain Biology.2018

INNO SPACE

SJIF Scientific Journal Impact Factor

Impact Factor

7.521

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY