International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

Determining Species of Bird Using Their Voice

M. Bhuvaneswar Redyy[1], S. Noortaj[2]

[1]MCA Student, Department of Master of Computer Applications, KMM IPS /Tirupathi, Chittoor (D.t), Andhra Pradesh, India

[2]Assistant Professor, Department of Master of Computer Applications, KMM IPS /Tirupathi, Chittoor (D.t), Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted : 14 May 2025

Published: 18 May 2025

Publication Issue

Volume 11, Issue 3

May-June-2025

Page Number

602-612

ABSTRACT

This project focuses on developing a deep learning model to identify bird species based on their vocalizations. Given the challenges posed by datasets with imbalanced class distributions, the aim is to curate a selection of bird species that have sufficient audio samples for effective model training. We utilize advanced algorithms, including Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM) networks, and WavNet, alongside comprehensive feature extraction techniques. The audio features extracted include zero-crossing rate, root mean square energy, and Mel-frequency cepstral coefficients (MFCCs), which are pivotal for distinguishing vocal characteristics among species. The model's performance is enhanced through data augmentation strategies, such as noise addition and pitch shifting, to increase the diversity of training samples. This approach allows for robust classification, even with a limited number of audio recordings per species. Ultimately, our model demonstrates the potential to accurately predict bird species based on audio input, contributing to biodiversity studies and ecological monitoring efforts.

Keywords—Bird Sound Recognition, Deep learning, CNN, LSTM, WavNet, Feature Extraction, Zero-Crossing Rate, Root Mean Square (RMS) Energy, MFCCs, Data Augmentation, Noise Addition, Pitch Shifting, Species Classification.

Introduction

Biodiversity plays a vital role in ensuring the health and stability of ecosystems, with birds acting as essential indicators of environmental well-being. Keeping track of bird populations helps scientists monitor ecological shifts, gauge the effects of human interference, and plan conservation efforts effectively. Traditionally, tracking bird species has depended on manual techniques like field observations and expert identification, which are not only time-consuming

and labor-intensive but also prone to inaccuracies and reliant on skilled personnel. However, with the rise of machine learning, there is now the potential to automate bird species recognition through the analysis of their vocal patterns, offering a more efficient and scalable method for biodiversity assessment.

Bird calls and songs are typically distinct to each species and provide rich data for classification. Nonetheless, identifying species using only their vocalizations presents some unique challenges. These include overlapping calls between similar species, background noise, diverse recording conditions, and the complex nature of audio data. Another issue lies in the imbalance of audio datasets—some species have ample recordings while others are underrepresented, leading to biased models that perform well for common species but struggle with rarer ones, ultimately compromising the model's overall effectiveness.

To overcome these limitations, this work aims to develop a machine learning-based framework capable of classifying bird species accurately using audio recordings. The primary goal is to build a well-balanced dataset comprising an equal distribution of bird species with sufficient vocal samples. This step helps mitigate data imbalance, thereby increasing the model's fairness and generalizability. The proposed solution employs powerful deep learning models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, and WavNet, each contributing their strengths to better extract and interpret the spatial and temporal features embedded in bird sounds.

Feature extraction is a crucial step in this methodology, as it enables the model to understand and differentiate subtle variations in vocalizations. Important features such as zero-crossing rate, root mean square (RMS) energy, and Mel-frequency cepstral coefficients (MFCCs) are computed from each audio file. Zero-crossing rate helps measure signal frequency behavior, while RMS energy reflects the power or loudness of the sound. MFCCs provide a compact representation of the sound spectrum and are particularly effective for capturing timbral characteristics that distinguish species. These features collectively help in training a model that can recognize complex audio patterns.

To improve the model's adaptability, data augmentation techniques are applied to expand the training dataset. Strategies such as adding background noise and pitch shifting simulate real-world scenarios, thereby making the model more robust to variations. Noise addition mimics natural environmental sounds, while pitch adjustments introduce variation in tone and frequency, reflecting different vocal expressions of birds. These approaches ensure that the model performs well even when exposed to new or limited data.

The hybrid architecture combining CNNs, LSTMs, and WavNet contributes to a holistic understanding of bird vocalizations. CNNs are proficient in analyzing spectrogram images derived from audio signals, LSTMs are capable of learning long-term temporal dependencies in sound sequences, and WavNet excels in modeling high-quality raw audio waveforms. This integrated approach ensures that both short-term acoustic features and long-term audio trends are effectively captured, enhancing classification accuracy.

In conclusion, the system developed in this project presents a promising approach for identifying bird species based solely on their vocalizations. It offers a scalable and automated tool for ecological research and conservation, reducing the burden on human observers while increasing the speed and scope of data collection. This machine learning-driven solution represents a significant advancement in environmental monitoring and contributes meaningfully to efforts aimed at preserving global bird diversity.

(a) Cup-cake (b) Red velvet cake

Fig 1: Different Types of Food Classes

Traditional image analysis techniques have shown limited success in food classification due to the semantic complexity and deformable nature of food. In contrast, deep learning techniques have demonstrated the ability to model these complexities. This paper focuses on three prominent approaches: building CNNs from scratch, applying transfer learning, and using platform-based models. Their comparative performance is evaluated to identify strengths and limitations.

Image classification typically involves a series of steps, beginning with data preprocessing (scaling, normalization, augmentation, etc.), followed by feature extraction. Traditional methods like Bag of Visual Words, PCA, or Random Forests have limitations, especially with complex food images. Deep CNNs, however, can learn abstract, high-level features through convolutional operations and generate rich feature maps.

These networks, often composed of millions of parameters, require large datasets and powerful computational resources. As a result, transfer learning using pre-trained models is preferred, allowing for fine-tuning on specific datasets. While traditional methods capture basic attributes like shape, color, and texture, CNNs enable extraction of deeper, more abstract patterns—leading to improved classification performance. In this work, pre-trained models such as SqueezeNet and VGG-16 are employed to achieve accurate food classification results.

RELATED WORKS

Here is a concise explanation and related details for each of the listed references, describing their relevance to bird species identification and machine learning in biodiversity monitoring:

[1] www.iucn.org/theme/species/our-work/birds

This is the official website of the International Union for Conservation of Nature (IUCN), specifically the section dedicated to bird conservation. It provides global data on bird species' conservation status, threats, and protective measures. This resource is essential for understanding the ecological importance of birds and justifies the need for accurate monitoring tools, such as machine learning-based identification systems, to support conservation initiatives.

[2] www.xeno-canto.org/, Xeno-Canto is a popular, community-driven database for bird sound recordings from around the world. It offers thousands of vocalization samples, including songs and calls for a vast number of bird species. This site is commonly used in bird sound classification projects as a primary source of audio datasets for training machine learning models and testing classification accuracy.

[3] Sujoy Debnath, Partha Protim Roy, Amin Ahsan Ali, M. Ashraful Amin, "Identification of Bird Species from Their Singing", ICIEV, 2016 This paper presents an approach to automated bird species identification using audio recordings of bird songs. The study demonstrates the use of audio features and classification algorithms to identify birds, laying the foundation for research into more advanced deep learning methods. It validates the importance of acoustic analysis in ornithology and supports the feasibility of automated monitoring systems.

[4] Rong Sun, Yihenew Wondie Marye, Hua-An-Zhao, "FFT-Based Automatic Species Identification Improvement with 4-layer Neural Network", ISCIT, 2013. This research utilizes the Fast Fourier Transform (FFT) for audio signal preprocessing, combined with a 4-layer neural network for bird species classification. The paper highlights the significance of frequency-domain analysis in

bioacoustics and introduces a deep learning model that improves species recognition accuracy. It serves as a stepping stone for more complex architectures like CNNs and LSTMs used in modern bird call recognition.

[5]

www.mathworks.com/help/deeplearning/ref/alexnet.html, this reference links to MATLAB's documentation for AlexNet, a pre-trained deep convolutional neural network commonly used for image classification tasks. In the context of bird species identification, AlexNet (or similar networks) can be adapted for spectrogram-based audio classification, where bird sounds are converted into images for processing. The documentation provides technical details for implementing and fine-tuning AlexNet in MATLAB, making it valuable for researchers using spectrogram images to classify bird calls.

EXISTING METHOD

Implementation of SVM Classifier for Bird Species Identification

Data Collection and Labeling

The process begins with gathering bird vocalization data from online sources such as Xeno-Canto. Each audio clip is associated with a bird species label. The audio samples are organized by species to prepare for supervised learning.

Audio Preprocessing

The raw audio files are standardized to a uniform format (e.g., mono channel, 16 kHz sampling rate). Noise reduction filters are applied to minimize background disturbances. Long recordings are segmented into smaller clips of fixed length (e.g., 5 seconds) to isolate distinct bird calls.

Feature Extraction

Each audio clip is transformed into numerical features that can represent its acoustic characteristics. Commonly used features include:

Mel-Frequency Cepstral Coefficients (MFCCs) capturing the timbral and tonal qualities of the bird calls.

Data Splitting

The dataset is divided into training and testing sets, often using an 80:20 or 70:30 ratio. In some cases, a validation set is also created from the training set to fine-tune model parameters.

Model Training Using SVM

An SVM classifier is trained using the extracted audio features. SVMs are chosen for their ability to handle high-dimensional data and perform well even with limited samples. A multi-class SVM setup is employed (e.g., one-vs-one or one-vs-rest), where each class represents a different bird species. Kernel functions like RBF or polynomial may be tested to find the best fitting model for the non-linear feature distributions.

Hyperparameter Tuning

Grid search or cross-validation techniques are applied to optimize the SVM parameters such as C (regularization parameter) and gamma (kernel coefficient). This step helps in balancing the trade-off between overfitting and generalization.

Testing and Evaluation

The trained SVM model is evaluated using the test dataset. Metrics such as accuracy, precision, recall, and F1-score are calculated to assess performance. A confusion matrix is also plotted to observe the model's behavior in classifying similar-sounding species.

Result Analysis and Improvement

If certain classes are underperforming, strategies like data augmentation (pitch shifting, noise addition) and feature selection are explored to improve class balance and feature relevance. The refined model is then re-evaluated.

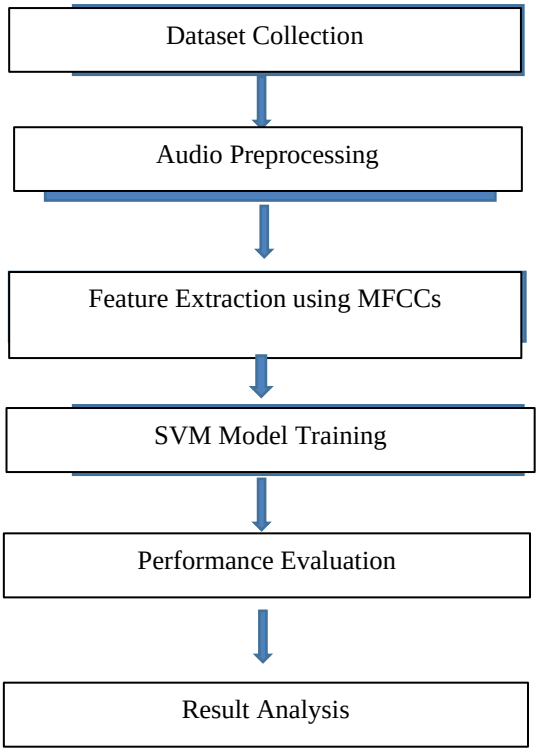

Fig 2: Flow graph of Existing Method

The proposed methodology focuses on the automatic identification of bird species using their vocalizations, moving away from manual identification methods that are often time-consuming and prone to error. The first stage of this process involves collecting audio recordings of bird sounds from publicly available sources such as Xeno-Canto. Once collected, these recordings undergo preprocessing steps, where noise reduction techniques are applied to enhance sound clarity. The audio is then segmented to isolate meaningful bird vocalization clips that are suitable for analysis.

These audio clips are transformed into visual representations using techniques like Mel-frequency cepstral coefficients (MFCCs) or spectrograms, which capture important time-frequency features of bird sounds. These representations help in converting complex audio signals into structured input formats that can be used for training machine learning models. The extracted features are stored in numerical form and standardized to ensure consistent scaling across the dataset.

To evaluate performance, the dataset is split into training, validation, and testing sets. The validation data helps monitor the model's accuracy and prevent overfitting, while the testing set is used to assess the model's generalization capability. Additionally, data augmentation methods such as adding background noise or modifying pitch can be applied to expand the dataset and improve model robustness. This entire methodology offers a systematic and scalable approach for bird species classification using audio data, eliminating the limitations of manual identification.

Disadvantages and Solutions

Class Imbalance: A major challenge in bird sound classification is the unequal distribution of audio samples across species, which often results in biased models that perform poorly on less-represented classes.

Inconsistent Audio Quality: Variations in environmental conditions and recording equipment can introduce background noise or distortions, negatively affecting the reliability of the classification.

Resource Demands: Sophisticated models often require powerful computing setups, limiting their use in low-resource environments or for real-time applications.

Complexity in Feature Extraction: Deriving essential sound features like MFCCs demands computational effort and domain expertise, which can be a hurdle for non-specialists.

Similar Sounding Species: Birds with overlapping or acoustically similar vocalizations can be difficult to distinguish, reducing the overall classification precision.

Limitations in Data Augmentation: Though methods like adding noise or adjusting pitch help increase data diversity, they may not fully replicate the natural range of bird calls.

Scalability Constraints: Expanding the model to include more species or deploying it in large-scale

monitoring efforts can be difficult due to data availability and processing needs.

Reliance on Accurate Labeling: High-quality, labeled datasets are essential for training effective models, but collecting and annotating these samples—especially for rare species—can be labor-intensive and costly.

PROPOSED METHOD

Implementation of Integrated method for Bird Species Identification

Initialization and Dataset Selection

Dataset Preparation

To begin, a well-curated dataset containing a diverse set of bird vocalizations is compiled. These recordings are sourced from trusted repositories such as Xeno-Canto, ensuring they represent a wide range of species. Each audio file is labeled accordingly and standardized to a consistent format, duration, and sample rate. This preparation ensures that the deep learning models can process the data uniformly.

Audio Preprocessing and Feature Extraction

Preprocessing steps include trimming silence, reducing background noise, and converting recordings into spectrograms or mel-spectrograms. Key audio features are then extracted, particularly Mel-Frequency Cepstral Coefficients (MFCCs), which capture the timbral and frequency characteristics of bird calls. These features form the foundation for effective learning by deep neural networks.

Data Augmentation

To improve model robustness and address the challenge of limited samples for certain species, data augmentation techniques are applied. These may involve background noise addition, pitch variation, and time stretching to simulate real-world conditions. This step enhances the variability of training data and improves the model's ability to generalize.

Model Architecture and Training

The system uses a combination of deep learning models to capture both spatial and temporal aspects of bird vocalizations. Convolutional Neural Networks (CNNs) are used to process spectrogram images and extract spatial patterns. Long Short-Term Memory (LSTM) networks are integrated to learn time-dependent features in sequential data. Additionally, WavNet may be employed to model raw waveform characteristics at a granular level. These models are trained using labeled audio features, optimizing a suitable loss function such as cross-entropy for multi-class classification.

Model Evaluation and Testing

After training, the models are validated using a separate portion of the dataset. Metrics such as accuracy, precision, and recall are used to evaluate performance. This step ensures the model can accurately identify unseen bird calls and maintain high reliability across various species.

Deployment and Interface Integration

Once trained and validated, the model is integrated into a simple and interactive interface. This allows users to upload or record bird calls and receive species predictions instantly. The interface is designed for ease of use by researchers, bird watchers, and conservationists, aiding in large-scale biodiversity monitoring and ecological studies.

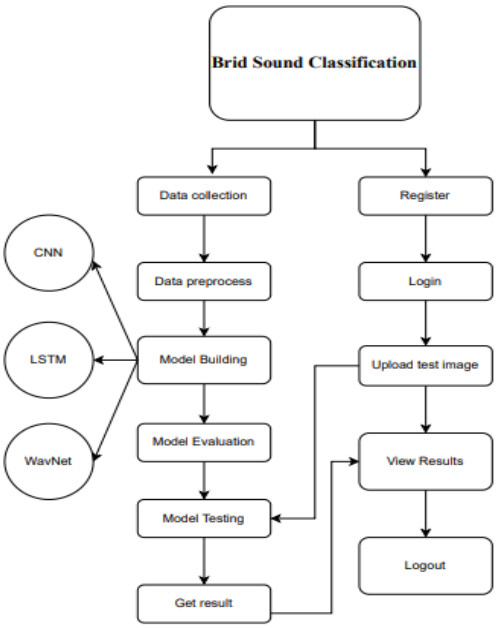

Fig 3: The Design of Proposed Method

The proposed system focuses on identifying bird species through the analysis of their vocalizations using deep learning techniques. It begins with the collection and curation of a comprehensive audio dataset containing recordings from a wide range of bird species. These recordings are standardized in terms of format and sampling rate, ensuring consistency throughout the preprocessing and training pipeline. The quality and diversity of the dataset are essential for effective learning and classification, as they directly influence the model's ability to generalize to new, unseen samples. Each audio file is carefully labeled with the corresponding species to facilitate supervised learning.

Once the dataset is ready, the recordings undergo preprocessing steps to enhance clarity and remove unwanted artifacts. This includes background noise reduction, silence trimming, and conversion into a suitable format such as mel-spectrograms, which visually represent the audio frequency components over time. From these representations, essential audio features are extracted, particularly Mel-Frequency Cepstral Coefficients (MFCCs). MFCCs are widely used in sound recognition tasks as they encapsulate critical frequency-based information that can help distinguish one bird species from another.

To further strengthen the training process, data augmentation techniques are employed. These techniques simulate variations in bird calls by modifying the pitch, speed, and background conditions of the audio. This helps in creating a more diverse and robust training set, allowing the model to learn a wider range of vocal characteristics. Augmentation is especially useful in addressing data imbalance, where some species may have fewer samples than others, thus reducing the likelihood of the model becoming biased toward more frequently occurring classes.

The classification model is built using a deep learning architecture that is capable of analyzing both spatial and temporal aspects of the audio features. The training phase involves feeding the processed feature data into the model and optimizing it using a loss function that measures the difference between predicted and actual class labels. After training, the model is tested on a separate dataset to evaluate its performance in identifying bird species from new audio recordings. Once validated, the model is integrated into a user-friendly interface, allowing end-users to input bird calls and receive real-time predictions, thereby contributing to ecological monitoring and species conservation efforts.

Future Scope

Future enhancements for improving classification performance can focus on several key areas. First, exploring advanced architectures and hybrid models that combine the strengths of different neural network types may yield better results. For instance, integrating CNNs with recurrent neural networks (RNNs) or employing attention mechanisms could help capture both spatial and temporal dependencies in the data. Additionally, conducting a more thorough hyperparameter tuning process and leveraging techniques such as cross-validation can optimize model performance and robustness.

Moreover, increasing the diversity and size of the training dataset could significantly enhance generalization. Techniques such as data augmentation or synthetic data generation might be employed to enrich the dataset, particularly for underrepresented classes. Finally, implementing ensemble learning methods that combine predictions from multiple models could further boost accuracy and reliability, allowing for a more comprehensive approach to classification tasks. Overall, these enhancements aim to refine model capabilities, increase accuracy, and improve overall performance in real-world applications.

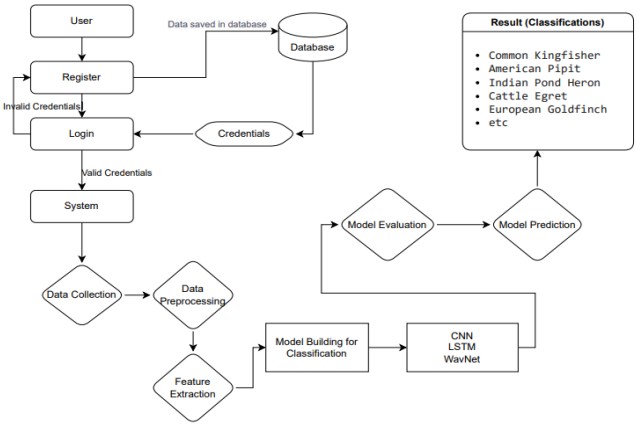

Fig 4: Model Architecture of Proposed Method

Advantages:

Enhanced Accuracy: Utilizes advanced deep learning architectures such as CNN, LSTM, and WavNet, significantly improving the precision of bird species classification.

Effective Handling of Imbalanced Data: Implements data augmentation techniques like noise addition and pitch shifting to balance datasets, ensuring robust model training for underrepresented species.

Comprehensive Feature Extraction: Employs a variety of audio features, including MFCCs, zero-crossing rate, and root mean square energy, to capture detailed vocal characteristics essential for distinguishing between similar species.

Scalability: Designed to accommodate a growing number of bird species and large volumes of audio data, facilitating expansion for broader biodiversity studies.

Robustness to Audio Variability: Incorporates preprocessing steps and data augmentation to mitigate issues related to background noise and varying recording conditions, enhancing the model's reliability in diverse environments.

Automation and Efficiency: Automates the bird identification process, reducing the need for time-consuming and labor-intensive manual observations, and enabling large-scale ecological monitoring.

Resource Optimization: Optimizes computational resources through efficient model architectures, making the system accessible for both large-scale projects and smaller, resource-constrained applications.

Support for Conservation Efforts: Provides accurate and timely data on bird populations, aiding in the monitoring and preservation of avian biodiversity and supporting informed conservation strategies.

Flexibility and Adaptability: Easily adaptable to different ecological settings and can be integrated with existing monitoring systems, enhancing its applicability across various research and conservation initiatives.

Improved Data Utilization: Maximizes the use of available audio recordings through advanced feature extraction and augmentation, ensuring that limited data resources contribute effectively to model performance.

Applications:

Biodiversity Monitoring: Automated bird call identification supports large-scale biodiversity surveys, enabling researchers to track species distribution and population trends over time without relying on manual observations.

Conservation Efforts: The system can help identify endangered or rare bird species in specific habitats, aiding conservationists in prioritizing regions for protection and developing targeted conservation strategies.

Ecological Impact Assessment: Environmental scientists can use bird vocalization data to assess the ecological health of an area, as changes in bird populations often reflect broader environmental changes such as deforestation or pollution.

Citizen Science and Education: The model can be integrated into mobile applications or web platforms, allowing bird enthusiasts and students to contribute to bird monitoring initiatives by uploading recordings and receiving species identification results.

Real-Time Habitat Monitoring: In remote areas, automated recorders can continuously collect and analyze bird calls, providing real-time data to alert

authorities about sudden ecological changes or the presence of invasive species.

Bioacoustic Research: Researchers in the field of bioacoustics can use the system to study communication patterns, mating behaviors, and territory marking among bird species, deepening our understanding of avian biology.

RESULTS AND DISCUSSIONS

Performance

The given figure 5 shows the classification report of CNN showing detected persons with precision, recall, f1-score and support all over 90%.

	precision	recall	f1-score	support
Acridotheresristis	0.92	0.93	0.93	128
Aegithaloscaudatus	0.82	0.94	0.88	70
Alaudaarvensis	0.86	0.93	0.89	70
Andean Guan_sound	0.92	1.00	0.96	22
Andean Tinamou_sound	0.92	1.00	0.96	23
Apusapus	0.90	0.96	0.93	46
Band-tailed Guan_sound	1.00	1.00	1.00	20
Caciuscela	0.95	0.93	0.94	161
Cardueliscarduelis	0.91	0.81	0.86	90
Cauca Guan_sound	1.00	1.00	1.00	27
Chlorischloris	0.95	0.89	0.92	139
Coccothraustescoccothraustes	0.93	0.86	0.89	186
Columbalivia	0.92	0.97	0.94	113
Columbapalumbus	0.95	0.94	0.95	85
Corvuscorone	0.95	0.84	0.89	62
Corvusfrugilegus	0.82	0.90	0.86	31
Cuculuscanorus	0.97	0.91	0.94	102
Delichonurbicum	0.88	0.88	0.88	85
Dendrocoposmajor	0.81	0.71	0.76	35
Dumetellacarolinensis	0.91	0.98	0.91	94
East Brazilian Chachalaca_sound	1.00	1.00	1.00	18
Emberizacitrinella	0.93	0.86	0.89	87
Erithacusrubecula	0.91	0.91	0.91	159
Ficedulahypoleuca	0.95	0.97	0.96	89
Fringillacoelebs	0.91	0.90	0.90	117
Gallusgallus	0.94	0.97	0.95	76
Garrulusglandarius	0.90	0.90	0.90	92
Hirundorustica	0.93	0.85	0.89	123
Laniusexcubitor	0.88	0.89	0.88	135
Luscinialuscinia	0.95	0.86	0.91	118
Motacillaalba	0.80	0.87	0.83	110
Motacillaflava	0.92	0.85	0.89	123
Parusmajor	0.96	0.95	0.96	81
Passerdomesticus	0.91	0.92	0.91	123
Phoenicurusochruros	0.92	0.89	0.90	125
Phoenicurusphoenicurus	0.88	0.94	0.91	133
Phylloscopuscollybita	0.90	0.95	0.93	124
Phylloscopustrochilus	0.95	0.97	0.96	104
Picapica	0.92	0.91	0.92	66
Pycnonotuscafer	0.94	0.93	0.93	54
Pycnonotusjocosus	0.89	0.95	0.92	112
Sittaeuropaea	0.86	0.91	0.89	114
Streptopeliaturtur	0.93	0.98	0.96	124
Sturnusvulgaris	0.95	0.85	0.90	110
Troglodytestroglodytes	0.95	0.91	0.93	103
Turdusmerula	0.90	1.00	0.95	112
Turdusphilomelos	0.89	0.89	0.89	119
Turduspilaris	0.88	0.93	0.90	153
Upupaepops	0.96	0.94	0.95	149
Variegated Tinamou_sound	1.00	1.00	1.00	28
accuracy			0.91	4690
macro avg	0.92	0.92	0.92	4690
weighted avg	0.92	0.91	0.91	4690

Fig 5: Classification Report of CNN

The given figure 6 shows the confusion matrix of CNN showing detected persons with TN, TP, FP, FN comparison.

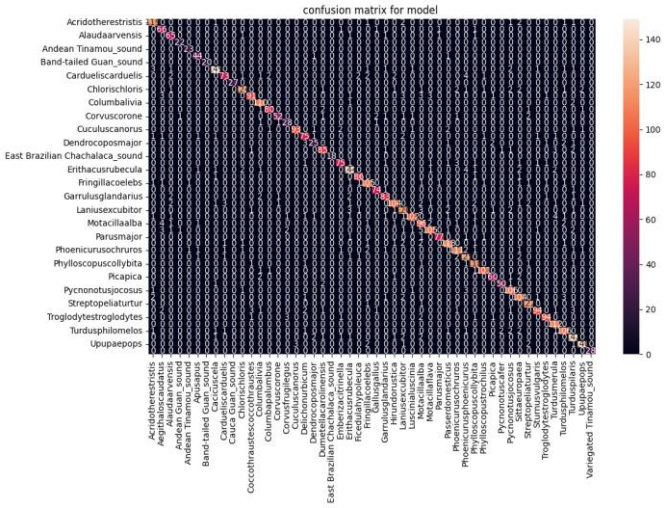

Fig 6: Confusion Matrix of CNN

The given figure 7 shows the confusion matrix of WavNet showing detected persons with TN, TP, FP, FN comparison.

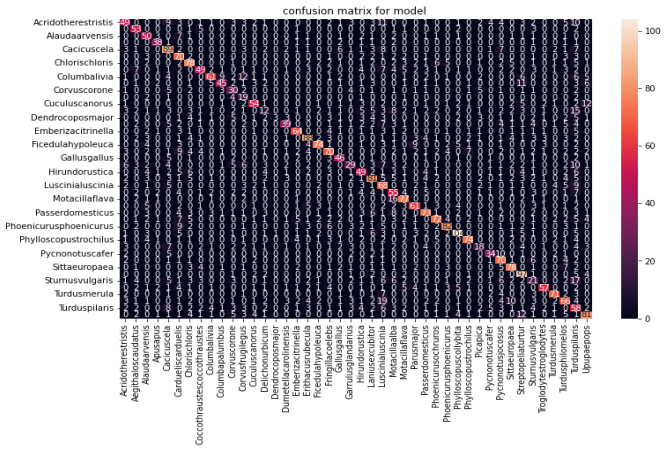

Fig 7: Confusion Matrix for WavNet

The given figure 8 shows the classification report of WavNet showing detected persons.

Model Confusion Matrix

	precision	recall	f1-score	support
Acridotherestristis	0.56	0.40	0.46	124
Aegithaloscaudatus	0.59	0.77	0.67	69
Alaudaarvensis	0.51	0.66	0.57	76
Apusapus	0.69	0.76	0.72	50
Cacicuscela	0.53	0.63	0.58	141
Cardueliscarduelis	0.46	0.75	0.57	93
Chlorischloris	0.56	0.60	0.58	129
Coccothraustescoccothraustes	0.64	0.44	0.52	112
Columbalivia	0.72	0.52	0.60	117
Columbapalumbus	0.70	0.49	0.58	91
Corvuscorone	0.40	0.48	0.44	62
Corvusfrugilegus	0.26	0.63	0.37	30
Cuculuscanorus	0.69	0.56	0.62	96
Delichonurbicum	0.26	0.13	0.18	89
Dendrocoposmajor	0.75	0.06	0.11	53
Dumetellacarolinensis	0.66	0.44	0.53	89
Emberizacitrinella	0.70	0.66	0.68	97
Erithacusrubecula	0.68	0.65	0.66	135
Ficedulahypoleuca	0.77	0.65	0.70	114
Fringillacoelebs	0.59	0.53	0.56	133
Gallusgallus	0.61	0.70	0.65	66
Garrulusglandarius	0.47	0.29	0.36	101
Hirundorustica	0.49	0.47	0.48	105
Laniusexcubitor	0.55	0.57	0.56	141
Luscinialuscinia	0.36	0.59	0.45	115
Motacillaalba	0.40	0.53	0.46	103
Motacillaflava	0.62	0.73	0.67	106
Parusmajor	0.64	0.59	0.62	103
Passerdomesticus	0.62	0.59	0.60	124
Phoenicurusochruros	0.68	0.55	0.61	131
Phoenicurusphoenicurus	0.75	0.63	0.68	130
Phylloscopuscollybita	0.75	0.69	0.72	150
Phylloscopustrochilus	0.77	0.65	0.71	113
Picapica	0.42	0.28	0.33	65
Pycnonotuscafer	0.64	0.57	0.60	60
Pycnonotusjocosus	0.51	0.75	0.61	93
Sittaeuropaea	0.64	0.66	0.65	115
Streptopeliaturtur	0.57	0.82	0.67	118
Sturnusvulgaris	0.26	0.22	0.24	96
Troglodytestroglodytes	0.76	0.55	0.64	104
Turdusmerula	0.78	0.72	0.75	99
Turdusphilomelos	0.49	0.46	0.47	143
Turduspilaris	0.24	0.42	0.31	137
Upupaepops	0.66	0.60	0.63	134
accuracy			0.56	4552
macro avg	0.58	0.56	0.55	4552
weighted avg	0.58	0.56	0.56	4552

Fig 8: Classification Report of WavNet

The given figure 9 shows the confusion matrix of LSTM showing detected persons.

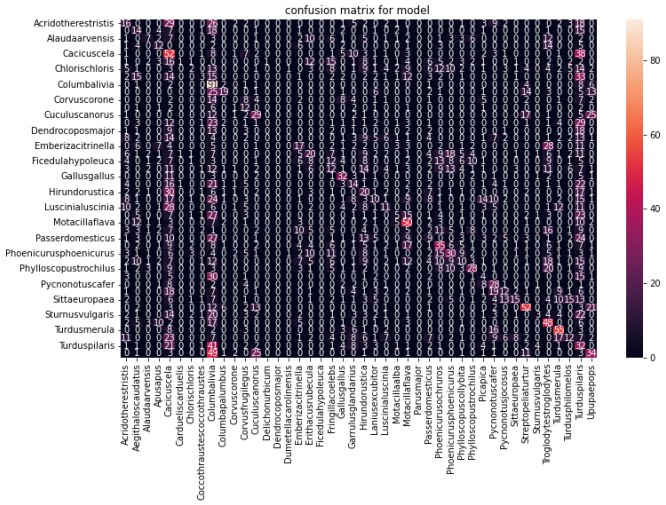

Fig 9: Confusion Matrix of LSTM Model

Model Confusion Matrix

	precision	recall	f1-score	support
Acridotherestristis	0.15	0.13	0.14	124
Aegithaloscaudatus	0.14	0.20	0.16	69
Alaudaarvensis	0.28	0.09	0.14	76
Apusapus	0.24	0.24	0.24	50
Cacicuscela	0.11	0.37	0.17	141
Cardueliscarduelis	0.33	0.01	0.02	93
Chlorischloris	0.13	0.02	0.03	129
Coccothraustescoccothraustes	0.20	0.03	0.05	112
Columbalivia	0.14	0.78	0.24	117
Columbapalumbus	0.51	0.21	0.30	91
Corvuscorone	0.00	0.00	0.00	62
Corvusfrugilegus	0.16	0.40	0.23	30
Cuculuscanorus	0.30	0.30	0.30	96
Delichonurbicum	0.00	0.00	0.00	89
Dendrocoposmajor	0.00	0.00	0.00	53
Dumetellacarolinensis	0.00	0.00	0.00	89
Emberizacitrinella	0.22	0.18	0.19	97
Erithacusrubecula	0.22	0.15	0.18	135
Ficedulahypoleuca	0.00	0.00	0.00	114
Fringillacoelebs	0.11	0.09	0.10	133
Gallusgallus	0.38	0.48	0.43	66
Garrulusglandarius	0.15	0.14	0.14	101
Hirundorustica	0.12	0.19	0.15	105
Laniusexcubitor	0.16	0.07	0.10	141
Luscinialuscinia	0.20	0.10	0.13	115
Motacillaalba	0.19	0.05	0.08	103
Motacillaflava	0.28	0.47	0.35	106
Parusmajor	0.00	0.00	0.00	103
Passerdomesticus	0.10	0.07	0.09	124
Phoenicurusochruros	0.24	0.27	0.25	131
Phoenicurusphoenicurus	0.25	0.23	0.24	130
Phylloscopuscollybita	0.16	0.07	0.09	150
Phylloscopustrochilus	0.39	0.25	0.30	113
Picapica	0.06	0.06	0.06	65
Pycnonotuscafer	0.22	0.47	0.30	60
Pycnonotusjocosus	0.24	0.13	0.17	93
Sittaeuropaea	0.43	0.13	0.20	115
Streptopeliaturtur	0.43	0.44	0.44	118
Sturnusvulgaris	0.27	0.03	0.06	96
Troglodytestroglodytes	0.22	0.46	0.29	104
Turdusmerula	0.43	0.56	0.48	99
Turdusphilomelos	0.22	0.08	0.12	143
Turduspilaris	0.06	0.23	0.10	137
Upupaepops	0.30	0.25	0.27	134
accuracy			0.19	4552
macro avg	0.20	0.19	0.17	4552
weighted avg	0.20	0.19	0.16	4552

Fig 10: Classification Report of LSTM Model

The given figure 10 shows the classification report of WavNet showing detected persons.

CONCLUSION

In conclusion, the comparative analysis of the CNN, WAVNet, and LSTM models demonstrates a clear distinction in their effectiveness for the classification task. The CNN model outperformed the others, achieving an accuracy of 91% alongside high precision and recall, making it the most reliable choice for accurate classification. Conversely, both WAVNet and LSTM models struggled significantly, with WAVNet attaining an accuracy of 58% and LSTM only reaching 20%. These findings suggest that

while the CNN model is well-suited for the given dataset, the other two models may require further refinement or alternative strategies to enhance their performance. Overall, the results emphasize the importance of model selection based on the specific characteristics of the dataset and the classification requirements.

References

[1]. IUCN. (n.d.). Birds - Species conservation work. Retrieved from https://www.iucn.org/theme/species/our-work/birds

[2]. Xeno-Canto. (n.d.). Sharing bird sounds from around the world. Retrieved from https://www.xeno-canto.org/

[3]. Debnath, S., Roy, P. P., Ali, A. A., & Amin, M. A. (2016). Bird species identification using vocal features. In Proceedings of the 5th International Conference on Informatics, Electronics and Vision (ICIEV).

[4]. Sun, R., Marye, Y. W., & Zhao, H. A. (2013). Improved bird species recognition using FFT and a four-layer neural network. Presented at the 13th International Symposium on Communications and Information Technologies (ISCIT).

[5]. MathWorks. (n.d.). AlexNet pretrained convolutional neural network. Retrieved from https://www.mathworks.com/help/deeplearning/ref/alexnet.html