

Wild Bird Species Identification Using ResNeXt50 Feature Extraction and Random Forest Classification

Konda Rupa Narendra Babu, Gade Yasaswi, Kadiyam Vasanthi, Kancharlapalli Tej Deep, Garika Sambasiva Rao,

Assistant professor, Dept of AI&DS, Seshadri Rao Gudlavalleru Engineering College,
Gudlavalleru, Andhra Pradesh

*UG student, Dept of AI&DS Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh

*UG student, Dept of AI&DS Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh

*UG student, Dept of AI&DS Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh

*UG student, Dept of AI&DS Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, Andhra Pradesh

Abstract— Accurate recognition of wild bird species from audio signals is essential for ecological monitoring and biodiversity conservation. This paper presents a hybrid ensemble-based system for bird species identification, aiming to improve recognition performance and robustness in real-world environments. The proposed approach utilizes a pre-trained ResNeXt50 deep learning model for feature extraction, followed by a Random Forest classifier for final classification. By combining deep feature learning with ensemble learning, the system achieves improved accuracy, robustness, and generalization across diverse field conditions. Furthermore, a Flask-based web interface with secure authentication is developed to facilitate system testing and performance analysis. Experimental results demonstrate that the proposed hybrid ensemble model outperforms standalone lightweight CNN models and is effective for real-time bird species recognition.

Keywords: Wild Bird Species Identification, ResNeXt50, Random Forest, Hybrid Ensemble Model, Deep Feature Extraction, Bioacoustic Signal Analysis, Flask Web Interface, Lightweight Classification System

INTRODUCTION

The functions of birds species in ensuring the ecological balance and maintaining biological diversity include seed dispersal, regenerating habitats, and the stability of the ecosystem. These are the ecological processes that directly lead to the recreational of the forests, the dynamics of the plant populations, and the natural ecosystems in general. The behavioral ecology of birds and their distribution is thus a crucial area in ecological studies and conservation policies because it allows scientists and policy makers to develop effective policy interventions that can protect the habitat and manage biodiversity. Experiments on the useful nature of bird-dispersed habitats have been previously noted to be of functional significance, particularly in fragmented forests where the activity of the birds plays a significant role in the recruitment of plants and in the sustainability of forests. Birds in such environments have been found to be important biological agents, which close the gaps in habitats and provide ecological connectivity. In the same way, species characteristics and movement patterns of birds have been indicated to have a strong influence on seed dispersal network and distance, which highlights the importance of precise monitoring of bird species in the natural habitat. The results show that morphological, flight, and foraging diversifications have a direct impact on ecological processes on the local and landscape levels.

New ecological studies have also been concerned with the large-scale data gathering and biodiversity evaluation using bird occurrence datasets in wetland, heritage ecosystems to help in conservation and policy-making. These data sets are very useful in understanding population dynamics, habitat utilization, movement and species susceptibility, therefore enhancing evidence-based conservation practices. Together with ecological research,

computer vision and machine learning have allowed automatic detection and re-identification of wild animals, making the monitoring process more efficient and consuming less human resources. Continuous monitoring as with images, videos, and audio records in large quantities using these technological advances is not limited in the same way that the traditional field surveys were. Moreover, the use of optimized planning and automation in agricultural and environmental applications through technology proves the increasing significance of intelligent systems in the real-world ecological management. These systems make the operations more efficient, minimize the wastage of resources, and help to achieve the sustainable development objectives, by incorporating data-driven decision-making into the environmental processes.

All these evolutions reveal that precise and automated and efficient systems to identify species of birds are needed, which has driven the need to adopt intelligent and data-driven systems to monitor bioacoustic and ecological environments..

I. LITERATURE SURVEY

Recent research has shown that deep learning has transformed radically the way bird sound recognition is done as models can learn meaningful acoustic patterns directly out of raw audio data. Lostonlen et al. [1] have demonstrated that the convolutional neural networks can automatically learn species-specific sound-related features, which results to better bird recognition compared to the traditional signal processing techniques.

Deep learning has also been shown to be helpful in large-scale experiments, in real-life conditions. Stowell et al. [2] organised the first bird audio detection challenge and was reported to have improved performance with noisy outdoor recordings by the modern neural models, which makes them applicable in the practical biodiversity monitoring systems.

Kahl et al. [3] developed BirdNET, which is a deep learning-based system that can be deployed to track birds on a continuous basis to aid in the use of large-scale ecological studies. They proved that automated systems can be effective in detecting and classifying the species of birds in various habitats, and it is no longer possible to use manual field surveys.

Hybrid modeling strategies have also been addressed in order to increase the strength of classification. Salamon et al. [4] combined both traditional machine learning techniques and deep neural networks and found out that the mixture of various features representations enhances recognition accuracy, particularly in highly acoustic settings.

Knight et al. [5] examined how automated acoustic recognition systems should be tested so as to attain valid and valuable performance results. The study has highlighted that the different quantifications of evaluation can lead to the different meanings of model accuracy and reliability.

The field has been developed with the help of benchmarking,

e.g. the LifeCLEF bird identification task. As Goeau et al. [6] state, deep learning models are much better than the existing methods when trained on large and varied audio data, and have become a new standard in bioacoustic classification studies.

According to Hershey et al., different CNN structures were compared in the context of large-scale audio classification and it was found that more profound models can always perform better in recognizing a list of diverse categories of sounds [7].

Tokozume and Harada [8] provided an end-to-end CNN to identify environmental sounds and demonstrated that raw audio data can be directly input into the network without the manually-crafted features. In their study, Sanchez et al. [9] developed a deep learning-based system of bird species recognition and achieved a high classification score in the real-world recording situation.

METHODOLOGY

The offered extended system is based on a hybrid ensemble-based approach to enhance the accuracy and robustness of the identification of wild bird species in the real world. First, the recording of bird sounds is

transformed to spectrogram representation and cleaned up to eliminate noise and normalize the dimensions of the input. This is followed by the use of a pre-trained ResNeXt50 model as a deep feature extractor that uses the grouped convolution design to extract rich and discriminative frequency-temporal features of bird vocalizations. The extracted deep features are then inputted in the Random Forest classifier that is able to make decisions based on the ensemble in order to increase the accuracy of the classification and in order to minimize overfitting. This is a combination of deep learning and ensemble learning, which enhances generalization in a wide range of datasets and field settings. Moreover, a small-weight Flask based Web interface with sensitive user authentication is constructed to allow test of the models and performance analysis to allow an efficient and user-friendly interaction with the system in testing during the time of validation of the experimental results.

456 bird audio recordings from websites like Xeno-canto and Kaggle are included in the collection. By capturing audio samples in real environments with background noise, real-world applicability was guaranteed. The dataset includes four species of birds that belong to the balanced class. To avoid data leakage, recordings from the same source or session were split 80:20 across training and testing sets. For consistent feature representation, audio recordings were converted into spectrograms using the Short-Time Fourier Transform (STFT).

A) Proposed System

The proposed system introduces a hybrid ensemble-based wild bird species identification framework based on deep learning and ensemble learning dual-way, high-accuracy, and high-robustness welfare in a real world setting. The first step is to record bird sounds under variety of field conditions and convert them into representation of time-frequency spectrogram with normal audio preprocessing methods. A deep feature extractor used as a pre-trained ResNeXt50 is used to learn rich and discriminative acoustic features on the spectrograms. Grouped convolutions in ResNeXt50 facilitate effective feature acquisition in the form of the architecture, coupled with good representational ability, which makes it an appropriate architecture to analyze complex bioacoustic signals.

To further improve the performance of classification, the extracted deep features are fed to the Random Forest classifier which is engaged in ensemble based decision making to achieve better generalization and minimise overfitting. The proposed hybrid method is a successful combination of the advantages of deep feature learning and conventional machine learning, which leads to higher accuracy on a variety of species of birds and in various environments. Also, a web interface based on Flask and secure authentication is created to provide testing and evaluation of the system. The interface permits the user to upload the bird sound data, see the classification output, and analyze the model performance, which guarantees a friendly and secure interface of usage in terms of experimental verification and application the model

in the real environment.

B) System Architecture

The system architecture of the proposed hybrid ensemble- based method of recognizing species of birds is produced in a gradual and modular manner to ensure performance and scalability. The raw recordings of the birds are initially captured in the field and transmitted via an audio preprocessing module where noise removal, normalization and segmentation takes place. The processed audio signals are then converted to the time frequency spectrogram representations that are fed into the deep learning part. The stage assists in ensuring that critical frequency and temporal attributes of bird vocalization are preserved to ensure that they are effective in learning features.

The spectrograms are then inputted to a ResNeXt50 built feature extraction module that has already been trained with recovered high-level and discriminative acoustic features via the application of grouped convolution processes. The resulting deep feature vectors are then issued to a Random Forest classifier where the ensemble based classification enhances the robustness and generalization of the classification to the various bird species. Finally, a secure user-authenticated flask-powered web interface is added to display the prediction results. With the use of this interface, a user is able to post samples of the sound of the birds, visualize a result of the

classification, and evaluate the work of the system that is the reason why the entire architecture is suitable to apply it to the real-time testing process and apply it to the work in the field.

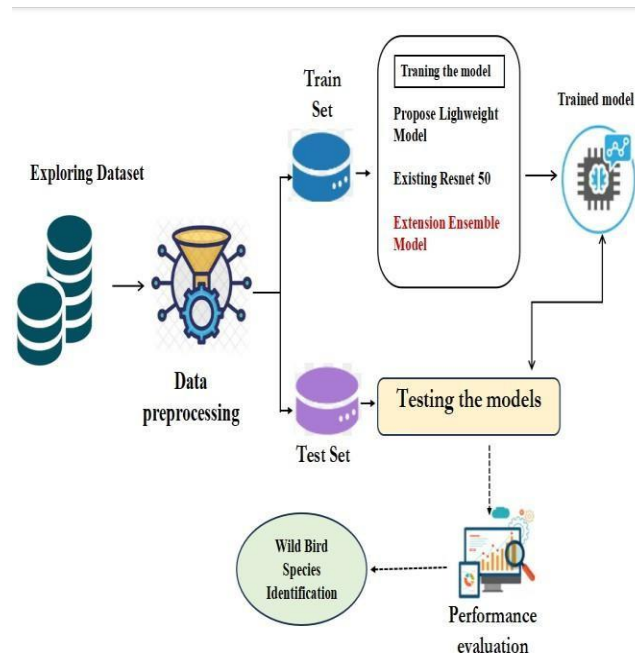


Fig. 1 illustrates the overall system architectureMODULES

a) Audio Data Utilization Module.

The module plays part in the recording of the uncooked bird sounds in other field settings. It is going to target recording the audio samples in the different ecological and environmental conditions to improve the model robustness and generalisation.

b) Preprocessing and Spectrogram Generating Module.

It reduces noise in the audio signals received in this module, normalizes and divides. The audio is then subjected to processing to get representations of time-frequency spectrograms without the loss of acoustically significant characteristics of bird vocalizations.

c) ResNeXt50 Deep Feature Extraction Module.

This module is based on a pre-trained ResNeXt50 model that is used to retrieve high-level and discriminative features of spectrogram inputs. Grouped convolutions aid in improved learning of complex frequency temporal patterns of the sounds of birds.

d) Random Forest (Ensemble Classification Module).

The deep features received are presented to a Random Forest classifier which is an ensemble based decision making. This module enhances accuracy of classification, reduces overfitting and enhances strength within different datasets.

e) Web Interface and Authentication Live Chat in flask. The module is user friendly and has the capability of uploading the bird samples of sounds and viewing the results of the classification. Test and assessment: This provides system integrity and authorized system access.

C) ALGORITHMS

- a. ResNeXt50 (Deep Feature Extraction Algorithm)

ResNeXt50 is a deep convolutional neural network that uses grouped convolutions to efficiently learn rich and discriminative features from bird sound spectrograms. In the proposed system, it is used as a pre-trained feature extractor to capture complex frequency–temporal patterns while maintaining computational efficiency.

b. Random Forest (Ensemble Classification Algorithm)

Random Forest is an ensemble learning algorithm that builds multiple decision trees and combines their outputs for final prediction. It is used to classify bird species based on deep features extracted by ResNeXt50, improving accuracy, robustness, and generalization while reducing overfitting.

c. Hybrid Ensemble Learning Strategy

This strategy integrates deep learning (ResNeXt50) with traditional machine learning (Random Forest). The combination leverages the strong representation capability of deep networks and the stability of ensemble classifiers, resulting in enhanced performance under diverse field conditions.

D) Confusion Matrix and Error Analysis

The Extension Ensemble Model delivers good classification accuracy across all bird species with few misclassifications, according to this investigation. Due to overlapping vocalization frequency patterns, the majority of mistakes occur between acoustically related species, such as Sparrow and Bulbul. Using deep ResNeXt50 features, the Random Forest classifier enhances decision bounds.

II. DATA SET DESCRIPTION

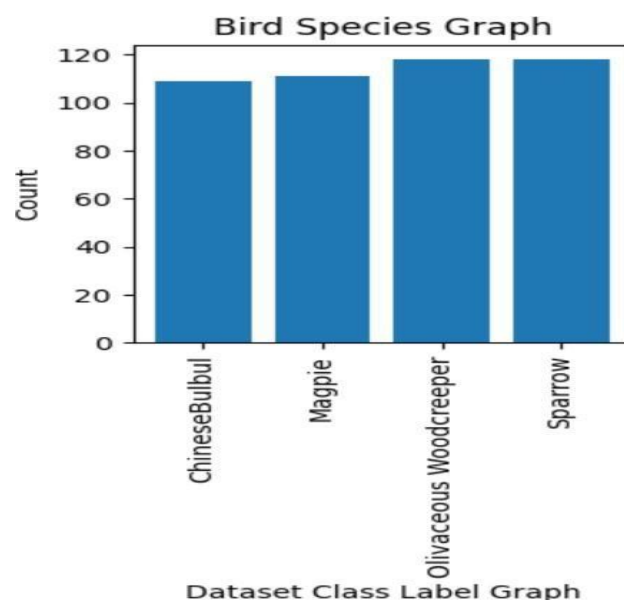


Fig. 2. Data set

The data employed in the research is a compilation of 456 records of bird sounds that were obtained through various Internet sites that were publicly available. These audio files are of four bird species, namely, Chinese Bulbul, Magpie, Olivaceous Woodcreeper, and Sparrow, and each class contains about 100 or more audio samples, which is a rather balanced distribution of classes, as can be seen in Fig. 2.

All audio samples were preprocessed and standardized and then trained. To evaluate the model reliably, the data was divided into 80 percent (364 recordings) and 20 percent (92 recordings) training and testing respectively.

The entire collection comprises 456 audio tapes which were recorded in various online sources. This even distribution in the bird species enhances the generalization of models and aids in the effective functioning of the model in the actual acoustic environment.

The dataset sources are publicly available and used strictly for academic and research purposes. The recordings follow open-access licenses such as Creative Commons (CC BY or CC BY-NC). No personally identifiable information is included in the dataset, ensuring compliance with ethical data usage standards. EXPERIMENTAL RESULTS

a) *Web Interface Security*

For secure authentication, the Flask web interface employs username-password validation. By verifying and cleaning user input, SQL injection and cross-site scripting are avoided. Momentary processing of uploaded audio files protects data privacy. The basic threat model uses controlled access and authentication to stop unauthorized access and data tampering. This guarantees the security of testing and deployment.

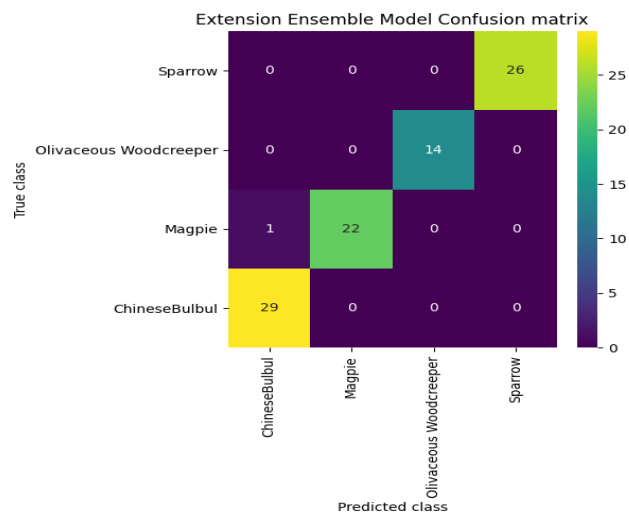


Fig 3. Confusion Matrix of Extension Ensemble Model

The confusion matrix of the proposed extended ensemble model for the classification of Chinese Bulbul, Magpie, Olivaceous Woodcreeper, and Sparrow is displayed in Figure 3. The matrix demonstrates that the majority of diagonal samples are correctly classified, demonstrating the accuracy of the model. One Magpie is mistakenly identified as a Chinese Bulbul. The hybrid ensemble model identifies bird species with minimal errors, according to the results.

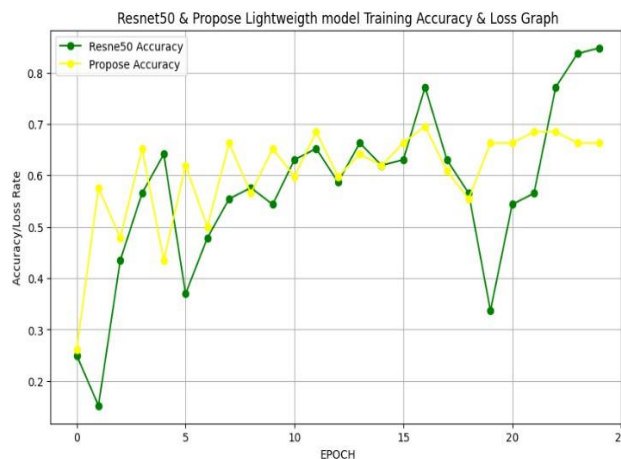


Fig 4 Training Accuracy Graph

Multiple period training accuracy trends for both the suggested lightweight model and the current ResNeXt50 model are displayed in Figure 4. The graph shows that although the lightweight model has lower and inconsistent accuracy, the ResNeXt50 model slowly improves with stable convergence. While the deeper model

learns more effectively and is more accurate, the lightweight model is computationally efficient but performs badly.

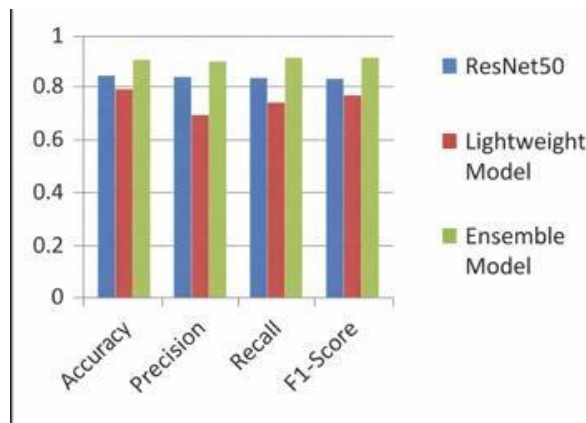


Fig. 5. Performance Comparison Graph

According to figure 5, the performance comparison graph is the graph that introduces a graphic overview of the tested models in terms of accuracy, precision, recall, and F1- score. Ensemble Model has a higher value on all metrics, meaning that it is more predictive. The ResNet50 model exhibits steady and competitive performance, as the scores of the evaluation measures are balanced. The Lightweight Model on the other hand has comparatively low performance pointing to a trade-off between efficiency and accuracy.

ML Model	Accuracy	Precision	Recall	F1-Score
ResNeXt 50	0.847	0.842	0.839	0.828
Lightweight Model	0.782	0.645	0.709	0.746
Extension Ensemble Model	0.929	0.921	0.929	0.930

Table1 performance comparison table

Table 1 presents the comparative performance of the evaluated models in terms of accuracy, precision, recall, and F1-score. The Ensemble Model achieves the highest accuracy of 0.929, outperforming ResNet50 (0.847) and the light weight Model (0.782), indicating its superior generalization capability.

In terms of precision, the Ensemble approach records a value of 0.921, demonstrating its effectiveness in reducing false positive predictions. ResNet50 maintains a strong precision of 0.842, where as the light weightModel shows a lower value of 0.645, reflecting reduced prediction reliability.

For recall, the Ensemble Model again leads with a score of 0.929, highlighting its ability to correctly identify the majority of relevant instances. ResNet50 achieves a recall of 0.839, while the light weight Model records 0.709, indicating comparatively lower sensitivity.

The F1-score further supports the overall trend, where the Ensemble Model attains the highest value of 0.930, indicating a well-balanced performance. ResNet50 achieves an F1-score of 0.828, while the lightweight Model

records 0.746. Overall, the results demonstrate that the ensemble-based approach consistently outperforms the individual models across all evaluation metrics.

VI . CONCLUSION

Experimental results demonstrate that the proposed model outperforms both the baseline ResNeXt50 and the lightweight model across all evaluation metrics, indicating improved generalization and reliability under varying conditions.

In addition, the integration of a secure and authenticated Flask-based web interface enhances system usability and supports practical deployment. The overall findings confirm that the proposed approach is robust, scalable, and suitable for real-time applications.

In conclusion, the proposed hybrid ensemble model provides an effective solution for accurate wild bird species recognition and can contribute to ecological monitoring and biodiversity conservation. Future work will focus on improving model efficiency and extending the system to larger and more diverse datasets

REFERENCES

- [1] V. Lostanlen, J. Salamon, B. McFee, M. Cartwright, A. Farnsworth, S. Kelling, and J. P. Bello, "Bird species identification from sound using convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 256–260.
- [2] D. Stowell, M. Wood, Y. Stylianou, and H. Glotin, "Automatic acoustic detection of birds through deep learning: The first bird audio detection challenge," *Methods Ecol. Evol.*, vol. 10, no. 3, pp. 368–380, Mar. 2019.
- [3] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecol. Informat.*, vol. 61, Art. no. 101236, 2021.
- [4] J. Salamon, J. P. Bello, A. Farnsworth, and S. Kelling, "Fusing shallow and deep learning for bioacoustic bird species classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 141–145.
- [5] E. C. Knight, K. C. Hannah, G. J. Foley, C. D. Scott, R.M. Brigham, and E. M. Bayne, "Recommendations for acoustic recognizer performance assessment with application to five common automated signal recognition programs," *Avian Conservation and Ecology*, vol. 12, no. 2, Art. no. 14, 2017.
- [6] H. Goëau, H. Glotin, W. P. Vellinga, R. Planqué, and A. Joly, "LifeCLEF bird identification task 2016: The arrival of deep learning," in *Proc. CLEF Working Notes*, Évora, Portugal, Sep. 2016.
- [7] S. Hershey *et al.*, "CNN architectures for large-scale audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 131–135.
- [8] Y. Tokozume and T. Harada, "Learning environmental sounds with end-to-end convolutional neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 2721–2725.
- [9] J. C. Sánchez, M. E. Pérez, and R. L. Vidal, "Automatic bird species recognition using deep learning," *Appl. Soft Comput.*, vol. 82, Art. no. 105572, 2019.