IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

AI-Powered Identification of Bird Species

Anusha S[1] and Prof. Parimal Kumar K R[2]
Student, Department of MCA[1]
Assistant Professor, Department of MCA[2]
Vidya Vikas Institute of Engineering and Technology, Mysore

Abstract: *Birds are essential to ecosystems, and recognising them is essential for conservation and biodiversity monitoring. It has over 9,000 bird species worldwide, many of which are rare and challenging to identify; expert manual classification is labour-intensive and prone to mistakes. This study aims at solving these difficulties. This study presents an AI- powered bird species recognition system that combines visual and auditory characteristics. The system recognises bird vocalisations using spectrogram-based audio processing and classifies pictures using CNN. Converting bird calls into spectrograms enables the model to analyse them as visual information, and CNNs have shown promise in image recognition. The project uses audio datasets like Kaggle for sound recognition and the Caltech-UCSD Birds 200 dataset for picture training. By combining visual and auditory cues, the suggested method achieves robust classification while lowering misidentification from background noise or species-to-species visual similarity.*

Keywords: Bird species, Machine Learning, Convolutional Neural Networks, Audio Recognition, Spectrogram

I. INTRODUCTION

Birds are relevant ecological indicators as they are quick to adapt to the environment. Bird observation gives helpful data to the ecological research, habitat management, and biodiversity conservation. Nevertheless, the traditional method of observing birds in the field and identifying them manually with the help of experts is the most common type of traditional bird identification, which is expensive, time- consuming and, most often, inaccurate.

Various species of birds can be identified with the help of photographs and sounds. Image recognition is inhibited by large inter-class similarity and also changes in light and angle and position although it has the benefit of strong visual features such as plumage, beak and body shape. Conversely, auditory recognition may be at times impaired by external noise, concurring calls, and variations in the quality of voice amongst users. The use of bird cries is important in field studies where visual observations are rare yet human beings also love sight recognition over hearing.

This study introduces a dual-modality system that takes into account the drawbacks of single-modality approaches. sound and picture identification. CNNs are used to interpret photographs of birds and CNN-based models are used to analyse bird vocalization that have been converted into spectrograms. By utilizing the complementary qualities of both modalities, this integrated approach increase the reliability of species identification. The project builtds a comprehensive recognition framework using Xeno-Canto audio recordings and CUB-200-2011 picture dataset.

II. LITERATURE REVIEW

In recent years, various academics have investigated AI-based bird species identification with image and audio datasets. [1] applied SVM to classify bird vocalisations based on spectral and auditory characteristics. While the model reached excellent accuracy, it had memory management concerns and needed high-quality audio recordings. [2] and [4] used CNN- based image recognition models, which offered good classification performance but were influenced by intraclass variance caused by changes in lighting, angle, and backdrop.

[3] used supervised machine learning with feature selection and noise reduction approaches to improve audio categorisation, but the system's performance decreased dramatically as the amount of species increases.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

[5] and [6] addressed image-based identification with CNN, AlexNet, and SVM techniques. Their algorithms enabled real- time prediction via online or mobile applications, but accuracy was strongly reliant on camera quality and dataset quantity, making it challenging to distinguish birds with subtle visual similarities. [7] pioneered bird identification in Jordan by introducing a VGG-19 model PCA for size reduction, despite delayed training and noisy backgrounds.

In contrast, [8] devised an audio-based strategy employing HFCC, k-NN, and HMM models, which correlated well with avian auditory traits but resulted in significant confusion between similar-sounding species. [9] used deep CNNs with spectrogram-based input, whereas [10] used MobileNet with transfer learning for bird sound classification. Both methods handled complicated sounds well, but struggled in loud situations and with bigger datasets.

Overall, most extant systems specialise in either sound-based or image-based identification, with CNNs being the most popular method. However, issues such as background noise, limited dataset and intraclass heterogeneity continue to impede accuracy. Few studies have attempted to integrate both picture and audio modalities, indicating an untapped research opportunity for developing more powerful and comprehensive identification algorithms.

III. METHODOLOGY

Technology used in this project depends on a dual- modality framework that integrates picture and audio recognition for identifying different bird species. The initial step is data collection. The image data is find in the Caltech-UCSD Birds 200 (CUB-200-2011) dataset of over 11,000 annotated photos of 200 bird species. This dataset has rich annotations such as part locations, properties, and bounding boxes and is, therefore, a useful standard to image- based classification. There is also recording of bird vocalisations in audio format in repositories such as Xeno- Canco which has great field recordings of the bird sounds and cries. The system can more easily differentiate similar species due to the presence of picture and sound data as it has access to both visual and sound signals.

The datasets are acquired and then processed in order to be made ready to be analyzed. Image data is downsampled to fixed-size dimensions to improve the diversity and generalisation of dataset and augmented with image manipulation approaches like rotation, cropping and flipping. Reduction of noise approaches like spectral gating and high- pass filtering are used to lower the effect of outside disturbances like wind or overlapping animal sounds in audio recordings.The processed recordings are then converted into spectrograms or Mel-frequency cepstral coefficients (MFCCs), which effectively convert the audio signals into image-like representations that can be analysed using the same CNN- based architecture created for visual data.

The core system is the model architecture, which is built on CNN. CNN layers are used for image recognition stream to extract hierarchical characteristics like simple edges and textures as well as sophisticated properties like plumage and body form. In the audio recognition stream, spectrograms run through the same convolutional procedure, allowing the model to collect critical time-frequency properties of bird vocalisations. Each stream consists of convolutional and pooling layers for feature extraction, activation layers such as ReLU for nonlinearity, and fully connected layers for classification. The outputs of the two streams are joined at a fusion step, where the decision-making layer incorporates both image- based and audio-based predictions, increasing the overall robustness of the system.

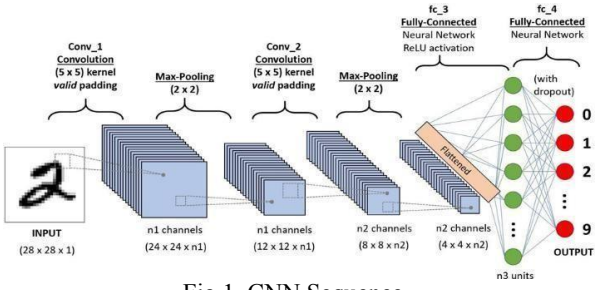

Fig 1. CNN Sequence

Finally, to ensure the output of the system, it is trained and tested. Both image and audio models are trained by backpropagation and gradient descent, and trained and tested on subsets of data. The combination of the two modalities ensures that the shortcomings of one modality can be compensated by the other, and dropout layers and data

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

augmentation methods are employed to reduce overfitting. The systems tested using common performance metrics like accuracy, precision, recall, and F1-score, which allow us to have a broad analysis of the benefits and drawbacks of the models. This methodology confirms that the system has the capability of recognizing birds by their visual as well as auditory features which would result to a more effective and correct way of species recognition.

IV. IMPLEMENTATION

The suggested system turns the dual-modality system into a prototype system that is able to process audio and visual data. The CNN models were prepared with python and the basic deep learning capabilities were given by TensorFlow and Keras platforms. Audio preprocessing libraries like Librosa were used to process audio to make it easier to obtain characteristics and convert audio into spectrograms. The hardware environment includes the use of GPU-enabled system which made the training much shorter and the computing process much more efficient. To enhance the resistance of the models to the real- life conditions, the images contained in the CUB-200 dataset were scaled to homogenous size and enhanced by adding rotation, brightness and flipping variables. Each audio recording was filtered prior to being converted into spectrograms which were used as image-like data to CNN pipeline to remove background noise.

The fusion stage involved the combination of the outcome of the image and audio trained recognition models independently to come up with a final classification. Its user interface was made to be very user-friendly and allowed a user to record the song of a bird or post an image. The data would then be automatically processed in the system and the predicted species and confidence levels on the same would be displayed. The technology was also still useful to researchers and enthusiasts in its ability to combine the backend processing with a simple user facing app. The implementation phase's use of dual-modality inputs demonstrated the system's potential for practical deployment in a range of field conditions in addition to its technical capability.

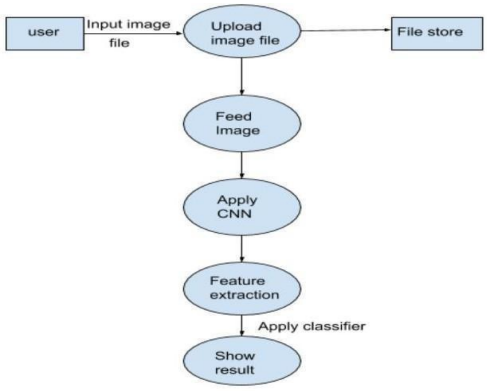

Fig 2. Flow System

V. RESULTS AND DISCUSSION

The project's results highlight the benefits of combining image and audio data for bird species identification. When tested individually, the image-based CNN produced excellent results, properly distinguishing visually dissimilar species like Painted Bunting and Brewer's Blackbird. However, its performance was low consistent in the case of species that comparable plumage patterns or when photos were cluttered with background elements. The audio- based model was effective at recognising bird sounds, especially when the recordings were clear and isolated, but it struggled in loud environments or when many birds vocalised at the same time. These limitations raised the issues of relying on a single medium only.

However, recognition accuracy was improved by the dual- modality technique. Misclassification, which would otherwise happen happen in single-modality models, could be eliminated by merging both streams of prediction. As an example, the auditory model frequently gave the decisive characteristic required to have a successful identification in cases where two species were identical on a visual basis. Conversely, the image model made up by giving constant

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

visual cues when audio signal were distorted by noise in the background. This complementary behaviour helped to increase the overall robustness, and made the integrated system more appropriate to real-world application with frequently imperfect conditions. Nevertheless, the research did not leave without challenges and noted the unequal representation of species in the data set, difficulties with getting high quality audio in environment, and misdiagnosis of similar sounding and visual phenomena. These results shows that, while dual- modality models are a major improvement, there is need for additional development using larger datasets and advanced deep learning approaches.

VI. CONCLUSION

This survey reveals that a dual-modality method combining visuals and sounds is a more dependable and accurate solution for identifying bird species than single- mode systems. By using CNN for visual features and spectrogram-based analysis for audio information, the system overcomes the constraints that exist when depending simply on images or sounds. The results show that combining the two modalities improves classification accuracy, decreases errors, and strengthens the system against ambient noise and visual similarity.

The ramifications of this work matters for biodiversity monitoring, ecological research, and conservation efforts, as it provides a practical tool that can be utilised by both professionals and laypeople. Future work could concentrate on increasing the dataset to include more diverse global species, using sophisticated architectures like attention-based models or transformers to better sequence learning in audio, and deploying Mobile-friendly models o enable real-time field applications. By incorporating these enhancements, the system has the potential to become a significant platform for large-scale bird monitoring, facilitating both scientific study and public participation in bird conservation efforts.

REFERENCES

[1] Bhavya Hegde, Bhagyashree V Bhat, Bhavyashree v Bhat, “Detection of Bird Species from Voice Feature,” International Research journal of Engineering and Technology(IRJET), 2023.

[2] “Bird Species Identification Using CNN,” International Journal of Scientific Research in engineering and Management (IJSREM), 2023, Dr. D. Siri, J. Snehith Reddy, V, Ajay Kumar Reddy, G. Jagadeshwara Sai, G. Maneesh and P. Vijay Kumar.

[3] Chingh Seh Wu, Sasanka Kosuru and Samaikya Tippareddy, “Identifying Bird Species from Audio Data,” 2nd International Conference on smart Electronics and Communication(ICOSEC) IEEE,2023.

[4] “Bird Species Identification Using Convolutional Neural Network,” International Conference on Emerging Trends in Engineering and Technology (ICETET), IRJMETS, 2022; Dharaniya R, Preetha M, Yashmi S.C.

[5] “Image-based Bird Species Identification,” 5th International Conference on Communication and Electronics Systems(ICCES), IEEE,2020, Anisha Singh, Akarshika Jain and Bipin Kumar Rai

[6] “Bird Identification by Image Recognition,” international Conference on Advances in Communication and Computing Technology (ICACCT), BEISP, 2022, Madhuri A, Tayal, Atharva Mangrulkar, Purvashree Waldev and Chithra Dangra.

[7] Suleyman AI-Showarah and Sohyb Obailtat, “Deep Learning-based Bird Identification Sysyem,” International Journal of Advanced Computer Science and Application (IJACSA), 2021.

[8] “Automatic Bird Species Recognition Based on Birds Vocalisation,” EURASIP Journal on Audio, Speech and Music Processing, Springer, 2018, by Jiri Stastny, Michal Munk and Lubos Juranek

[9] Mario Lasseck, Deep Convolutional Neural Networks for Audio-based Bird Species Identification,” Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, 2018

[10] “Bird Sound Recognition Using a convolutional Neural Network,” 17th IEEE International Conference on Machine Learning and Application (ICMLA). IEEE, 2018, by Agnes Incze, Henrietta-Bernadett Jancso, Zoltan Szilagyi, Attila Farkas and Csaba Sulyok.