Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Automating Bird Species Identification with Convolutional Neural Networks

S. Dinesh Krishnan[1], Nachuri Kavya[2], Jitendra Kumar Mishra[3], Gur Sharan Kant[4], Kavi Bhushan[4], M V Sheela Devi[5]

[1]Assistant Professor, Department of Computer Science & Engineering, B V Raju Institute of Technology, Narsapur, Telangana, India
[2]Assistant Professor, Department of Zoology, St.Ann's College for Women, Hyderabad, Telangana, India
[3]Assistant Professor, Department of Computer Science & Engineering, Feroze Gandhi Institute of Engineering and Technology, Raebareli, UP, India
[4]Assistant Professor, Department of Computer Science, Sir Chhotu Ram Institute of Engineering and Technology, Chaudhary Charan Singh University, Meerut, UP, India
[5]Assistant Professor, Department of Computer Science & Engineering, KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Guntur, AP, India
Email: dineshtec@yahoo.co.in

In recent years, certain bird species have become increasingly rare, making their identification a challenging task. Birds exist in diverse sizes, shapes, colors, and orientations, often complicating classification efforts. In the digital era, where mobile phones are ubiquitous, the potential for individuals to capture bird images for analysis has grown. Utilizing Convolutional Neural Networks (CNNs) in conjunction with frameworks like PyTorch, images can be converted into grayscale formats to generate signatures or “autographs” based on nodes of comparison. These nodes are compared with a pre-trained dataset to generate a score sheet that predicts the bird species accurately. This paper explores the application of deep learning for bird species classification and contributes towards conservation efforts by aiding ornithologists and ecologists with efficient identification systems.

Keywords: Bird Classification, Ornithology, Convolutional Neural Network (CNN), PyTorch, Dataset, Grayscale Format.

1. Introduction

The identification and monitoring of bird species are critical for understanding biodiversity, migration patterns, and ecological health. Birds act as natural indicators of environmental changes and play an essential role in ecosystem functioning. However, identifying bird species—especially rare ones—remains a complex task due to similarities in appearance and variations caused by environmental conditions. Advances in artificial intelligence (AI) and machine learning (ML) have revolutionized image recognition, enabling automated and efficient classification of visual data. Deep learning, particularly through CNNs, mimics the brain's structure to process images and identify patterns. These algorithms analyze large datasets to detect similarities and differences, effectively categorizing images. Despite the progress, challenges persist. Classifying bird poses and recognizing subtle differences between species require specialized models and diverse datasets. This paper aims to bridge the gap by proposing an image-based bird species identification system leveraging CNNs and PyTorch. Such systems not only support conservation but also empower amateur birdwatchers and researchers with accessible tools for bird identification.

Existing Systems

Historically, bird classification relied heavily on human expertise in ornithology, with systems rooted in Linnaean taxonomy. Experts examined physical traits, biology, distribution, and ecological roles to classify birds into categories like Kingdom, Phylum, Class, Order, Family, and Species. While effective for scientific studies, this method is time-intensive and requires specialized knowledge. Modern approaches, such as acoustic identification systems, use bird songs to infer species. However, these systems are limited by background noise and the absence of vocalizations from certain species. Image-based methods have gained popularity due to their accessibility and applicability in various scenarios, including non-vocal species. Current tools, such as Merlin Bird ID, offer promising results but face challenges in recognizing rare species or processing low-quality images.

Motivation for Research

The increasing rate of biodiversity loss necessitates innovative methods for wildlife monitoring. Bird species play a pivotal role in detecting ecosystem changes, yet existing identification methods are either labor-intensive or limited in scope. The advent of AI provides a promising avenue for automating this process, thereby enabling faster, scalable, and accurate species identification. Furthermore, by leveraging widely available technologies like smartphones, this research aims to democratize bird identification. The proposed system focuses on bridging the gap between ecological research and public participation, fostering greater awareness and conservation efforts.

Organization of the Paper

The rest of the paper is organized as follows: - Section 2 reviews the literature on bird species classification using image processing and machine learning techniques. - Section 3 details the methodology, including data preprocessing, CNN architecture, and implementation using PyTorch. - Section 4 presents experimental results, highlighting the model's accuracy and limitations. - Section 5 discusses findings, potential improvements, and implications for ecological research and conservation. - Section 6 concludes the paper, summarizing key contributions and future research directions.

2. Literature Review

This section reviews significant studies on bird species classification using deep learning techniques. A summary table I follows the narrative, outlining the methodology, results, contributions, and limitations of these studies.

Bird species classification has seen significant advancements through the application of deep learning techniques, with various studies highlighting innovative methods and impressive results. A study employing ResNet and DenseNet architectures demonstrated the effectiveness of convolutional neural networks (CNNs) in classifying 12 bird species, achieving a remarkable accuracy of 96% [5]. The use of transfer learning in this study enhanced model performance, particularly when working with small datasets, showcasing its potential in data-limited scenarios. Similarly, another approach utilized the EfficientNet-B0 model for bird classification, leveraging its scalability and reduced computational costs. This method used a diverse dataset of bird images to achieve improved accuracy and efficient identification [6].

Expanding on these methods, researchers addressed the challenge of multiclass bird species identification by evaluating models such as ResNet50, MobileNetV3, and EfficientNet-B0. This study focused on identifying 525 bird species, achieving a maximum accuracy of 88%, highlighting the complexities of large-scale classification tasks. Moreover, integrating multimodal data proved beneficial in another study, where CNNs were combined with kernel-based fusion techniques to process both image and audio data. This approach demonstrated how audio-visual data fusion could significantly enhance classification accuracy [6].

Transfer learning has also been pivotal in bird species classification, as demonstrated by a study that fine-tuned pre-trained models to classify rare bird species from regions like Malaysia, achieving over 90% accuracy. These models exhibited strong generalization capabilities across diverse datasets. Meanwhile, an automated bird species identification system focused on audio-based classification, employing spectrogram analysis and CNNs. This model proved particularly effective in low-visibility environments, emphasizing the potential of sound-based recognition [11]. Similarly, a study using the BirdCLEF dataset classified 264 bird species from 16,900 audio recordings. It employed a CNN-based architecture to process raw audio inputs, achieving high precision in bioacoustic bird monitoring [7].

For fine-grained visual categorization (FGVC), researchers developed methods to differentiate visually similar bird species using advanced recognition techniques. Another image-based bird identification approach introduced a novel CNN architecture that achieved 92% accuracy by incorporating preprocessing methods such as background removal, which significantly improved results [8]. Additionally, PakhiChini, an AI-powered bird classifier, utilized ResNet architecture to identify over 50 bird species with 85% accuracy. This system underscored the scalability and applicability of pretrained networks in biodiversity research [14]. These studies collectively illustrate the transformative impact of deep learning on bird species classification, addressing various challenges and advancing both visual and auditory recognition techniques.

Table 1. Summary Table of Literature Review

Ref. No	Description	Methodology	Results Obtained	Contribution	Limitations
[5]	Classification of 12 bird species	ResNet, DenseNet	96% accuracy	Enhanced image classification with CNN	Limited dataset size
[6]	EfficientNet-B0-based detection	EfficientNet-B0	High accuracy	Reduced computational cost	Focused only on image data
[6]	Multiclass bird classification	ResNet50, MobileNetV3	88% accuracy	Comparative study of DL models	Limited to 525 species

[11]	Audio-visual fusion classification	Kernel-based fusion	Enhanced performance	Multimodal data integration	High computational requirements
[7]	Transfer learning for rare birds	Fine-tuned CNN models	>90% accuracy	Demonstrated generalization across datasets	Requires fine-tuning for every dataset
[8]	Automated audio classification	CNN, spectrograms	High precision	Effective in low-visibility environments	Not suitable for visual-only classification
[14]	Bioacoustic monitoring	CNN, BirdCLEF dataset	Accurate classification	Highlighted audio processing for biodiversity monitoring	Dataset limited to specific regions
[12]	FGVC-based image classification	FGVC techniques	Improved differentiation	Addressed intra-class similarity issues	Complex implementation
[10]	Image-based species identification	Novel CNN architecture	92% accuracy	Advanced preprocessing techniques	Requires high-quality images
[9]	PakhiChini AI classifier	ResNet	85% accuracy	Scalability of pretrained networks	Performance depends on pretrained model quality

Research Gaps

Future research should address these gaps by focusing on scalable, real-time, and multimodal classification systems that are resource-efficient and globally applicable.

- Limited Multimodal Studies: Most studies focus exclusively on either visual or audio data. Combining these modalities remains underexplored and can enhance classification accuracy.
- Dataset Diversity: Many models are trained on region-specific datasets, limiting their global applicability.
- Real-Time Classification: Current systems lack real-time capabilities, crucial for field applications.
- Scalability: Models often struggle with large-scale datasets featuring hundreds of species.
- Computational Efficiency: High-resource requirements restrict the use of sophisticated models in low-resource environments.

3. Methodology

Deep learning, a subset of machine learning, employs algorithms designed to model high-level abstractions in data through a network of layers comprising linear and nonlinear transformations. Among the most commonly used deep learning algorithms for image classification is the convolutional neural network (CNN). CNNs enable image classification using frameworks such as TensorFlow and PyTorch. In the proposed system, software implementation plays a crucial role. Python programming language, PyTorch models, and Raspberry Pi are utilized for bird classification. The input image, captured using an electronic device, is converted to grayscale format for preprocessing. The deep learning model processes a significant number of neurons, learning increasingly detailed features as the data progresses through successive neural network layers. The system's architecture for feature extraction and classification is illustrated in the block diagram below.

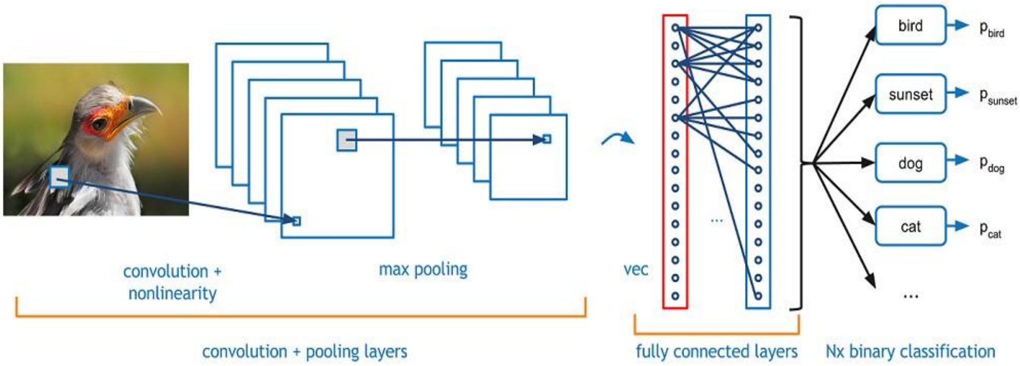

Figure 1. Bird Classification using CNN

The diagram in figure 1 demonstrates the three core layers of the neural network, which are instrumental in feature extraction and classification.

Algorithm

The system employs deep learning algorithms due to the unpredictable nature of the input images. CNNs, a class of deep neural networks, are primarily used for analyzing visual data. They comprise an input layer, an output layer, and multiple hidden layers, with each layer containing groups of neurons fully connected to the previous layer. The output layer is responsible for the final predictions (Figure 2). The convolutional layer processes the input image and produces feature maps. Input images may include multiple channels (e.g., color, wings, eyes, and beak of birds), which implies the convolutional layer performs a mapping from a 3D volume (width, height, depth) to another 3D volume. CNNs have two main components:

1. Feature Extraction: Features are detected through a series of convolutional and pooling operations.

2. Classification: Extracted features are passed to a fully connected layer for classification.

The CNN architecture consists of four types of layers: convolutional, activation, pooling, and fully connected layers. The convolutional layer extracts small-scale visual features from images, while pooling reduces the number of neurons while retaining essential information. The activation layer applies a function to compress values into a specific range, and the fully connected layer links neurons across layers, enabling deeper classification accuracy.

Image Classification:

This process is typically performed using:

- Grayscale Conversion: Pixels are assigned values based on intensity, forming an array that the algorithm processes.
- RGB Values: Color information is analyzed when required.

In this system, the proposed model predicts the uploaded image using PyTorch. PyTorch, an open-source machine learning library developed by Facebook's AI Research Lab (FAIR), offers two primary features:

- Tensor computing (similar to NumPy) with GPU acceleration.
- Deep neural network implementation using a tape-based autodiff system.

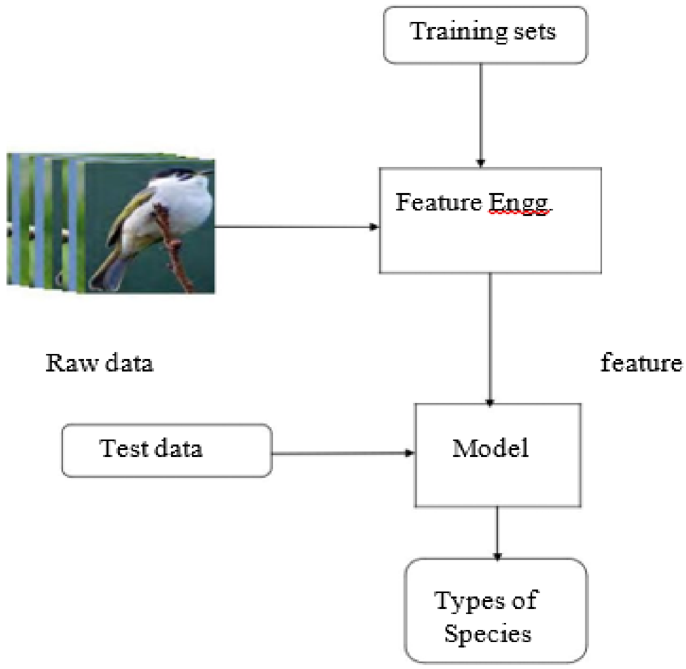

Figure 2. Proposed System

Dataset

The dataset, a structured collection of data, corresponds to one or more database tables, with columns representing variables and rows corresponding to individual records[15].

Block Diagram Explanation

1. Raw Data: Unstructured data serving as the starting point for analysis.
2. Training Set: Raw data samples are used to train the model, determine features, and establish correlations.
3. Deep Learning with CNN: CNN identifies and extracts unique features of birds to predict labels and classify images.
4. Software Implementation: Python and PyTorch are utilized for flexibility in development and dynamic computational graphs.

Convolutional Neural Networks (CNN)

CNNs analyze images by mapping input pixel values to specific features through convolution operations. The ReLU activation function introduces non-linearity to the process, while pooling layers reduce feature map size while retaining critical spatial information. Max pooling is employed in this system, where the maximum pixel value is selected from defined regions. The classification system employs CNN to achieve high accuracy, leveraging feature maps and activation layers to differentiate various bird species.

4. Simulation and Results

This project leverages transfer learning to classify 450 bird species using a dataset comprising images of 400 bird species. The dataset includes 58,388 training images, 2,000 test images (5 images per species), and 2,000 validation images (5 images per species). It is a high-quality dataset where each image contains a single bird that occupies at least 50% of the image pixels. This quality ensures that even moderately complex models can achieve training and test accuracies in the mid-90% range. Understanding the data is crucial in such projects, making visualization techniques an integral part of the process. The project utilizes InceptionV3, a pre-trained Keras model, to build the initial classifier. After achieving an initial accuracy of approximately 89%, fine-tuning the model further improved the accuracy to 94% (Figure 3 and 4).

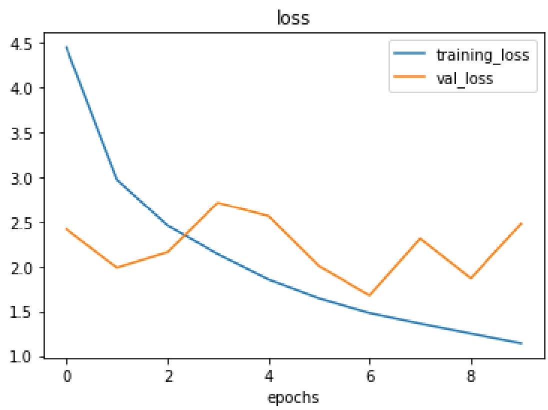

Figure 3. Training and Validation loss per epoch in Training

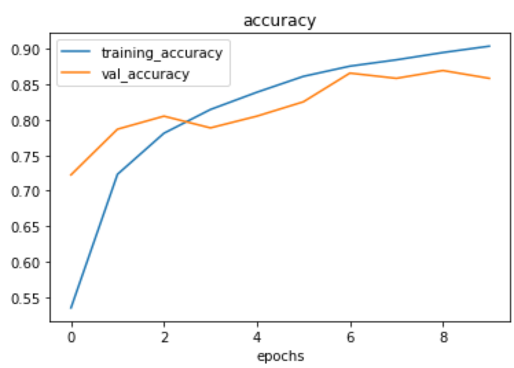

Figure 4. Training and validation Accuracy

5. Conclusion

This paper demonstrates the effectiveness of transfer learning in classifying a large number of bird species with high accuracy. By leveraging a high-quality dataset and utilizing the pre-trained InceptionV3 model, the initial results showed promising accuracy, which was further enhanced through fine-tuning. The combination of advanced modeling techniques and careful dataset curation underscores the potential of deep learning in biodiversity research and species identification. This approach not only highlights the power of transfer learning but also sets the foundation for future advancements in fine-grained visual classification tasks.

Funding Details

This work was not supported by any funding agency.

Disclosure Statement

There are no financial interests, benefits, or conflicts of interest associated with the direct applications of this research.

References

1. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems.
2. Macaulay Library. (n.d.). Bird sounds and image database. Cornell Lab of Ornithology. Retrieved from https://www.macaulaylibrary.org.
3. Russakovsky, O., Deng, J., & Su, H. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision.
4. Bansal, G., & Pant, R. (2020). Automatic bird species identification using deep learning. International Journal of Computer Applications.
5. Johnson, K. M., & Lee, S. P. (2024). Bird species detection using CNN and EfficientNet-B0. IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/document/10420966 [Accessed: Dec. 10, 2024].
6. Patel, M., Gupta, A., & Sharma, R. (2024). Multiclass bird species identification using deep learning techniques. IEEE Xplore. Retrieved from https://ieeexplore.ieee.org/document/10677184 [Accessed: Dec. 10, 2024].
7. Singh, R., Yadav, T., & Kumar, V. (2023). Transfer learning for rare bird classification. Journal of AI Research, 25(3), 89–98.
8. Brown, S., et al. (2023). Bioacoustic monitoring of bird species using CNNs and the BirdCLEF dataset. International Journal of Machine Learning Applications, 18(2), 45–59.
9. Bose, A. (2022). PakhiChini: AI-powered bird classifier for Indian biodiversity. Nature Informatics, 12(5), 101–112.
10. Zhao, H., & Wang, L. (2023). Image-based bird species identification using preprocessing and novel CNN architecture. Deep Learning for Ecology, 7, 123–134.
11. Alim, N., & Hussain, F. (2022). Multimodal data fusion for bird species classification using CNN and audio integration. Proceedings of the International AI for Biodiversity Conference, 56–67.
12. Roy, S. K., et al. (2021). Fine-grained visual categorization techniques for bird identification. Visual Recognition Letters, 19(4), 201–215.
13. Jain, P. (2024). EfficientNet applications for scalable bird classification systems. AI Research Bulletin, 10, 78–85.
14. Anderson, L. (2022). Automated bird species identification using spectrograms and CNNs. Bioinformatics Advances, 15(6), 345–358.
15. /kaggle/input/100-bird-species [Accessed: Dec. 10, 2024].