International Journal of Engineering Trends and Technology
ISSN: 2231-5381/ https://doi.org/10.14445/22315381/IJETT-V74I11P118

Volume 74 Issue 11, 253-266, November 2026
© 2026 Seventh Sense Research Group®

Original Article

Real-Time Bird Species Detection using YOLOv8 and Edge Computing for Intelligent Biodiversity Monitoring

Rohit Samkaria[1], Rajesh Singh[2], Anita Gehlot[3], Rahul Mahala[4], Purnendu Shekhar Pandey[5]

[1,2,3]*Uttaranchal Institute of Technology, Uttaranchal University, Dehradun, Uttarakhand, India.*
[4]*Law College Dehradun, Uttaranchal University, Dehradun, Uttarakhand, India.*
[5]*GL Bajaj Institute of Technology and Management, Greater Noida, UP, India.*

[4]*Corresponding Author: rahulmahala98@gmail.com*

Received: 21 July 2026 Revised: 12 September 2026 Accepted: 18 September 2026 Published: 25 November 2026

Abstract *- Bird species monitoring is the most crucial component of biodiversity conservation, ecological research and wildlife management, and it is essential to be accurate and automated. Intelligent computer vision-based methods are essential for traditional methods that are tedious, time-consuming and observer dependent. In this study, the edge-enabled bird identification system is proposed based on the YOLOv8 framework, which is an object detection system and aims to detect and identify 17 bird species in real time. To boost the model's ability to generalise across various environmental scenarios, a publicly available Roboflow dataset of 1,131 annotated images was extended to 2,376 images using extensive data augmentation, such as rotation, flipping, brightness adjustment, scaling, blurring, and noise injection. The YOLOv8 model was trained on the augmented dataset in Google Colab and tested for its performance with precision, recall, F1-score, confusion matrix, and mean Average Precision (mAP). The trained model obtained a precision of about 0.85, recall of 0.80 and mAP@0.5 of 0.81, which reflects the ability to detect birds and classify them to their species. The optimized model was deployed with a Raspberry Pi 5 embedded with a Raspberry Pi Camera Module to demonstrate its practical utility, with the objective of detecting birds in real time without relying on cloud services, and on a low-cost embedded platform. Experimental results have verified that the developed system is an efficient, accurate and scalable solution for automatic bird monitoring, biodiversity assessment, ecological surveillance, wildlife conservation and intelligence at the edge for environmental monitoring.*

Keywords *- Bird species detection, YOLOv8, Edge Artificial Intelligence, Biodiversity monitoring, Raspberry Pi.*

1. Introduction

The Sustainable Development Goal 15 (Life on Land) and its targets focus too narrowly on biodiversity conservation while overlooking social justice, human livelihood, and nature, because it relies on outdated measures insufficiently addresses social injustices, which limits the ability to achieve the broader goal of transformation and sustainable development [1]. For keeping track of the species with more accuracy and place-specific assessments of land ecosystems, satellite imagery and spatial analysis is useful for policymakers to track environmental changes and implement targeted conservation strategies more effectively [2]. Essential Biodiversity Variables (EBVs) provide a consistent framework for supporting more effective conservation efforts, tracking biodiversity change, and enabling better assessment of progress toward biodiversity and sustainability goals, because it is important to improve global biodiversity monitoring. And integrating remote sensing with EBVs can improve national biodiversity monitoring because satellite data can help track ecosystem changes more efficiently and provide more reliable information to support conservation planning and biodiversity assessments [3, 4]. Locals help in collecting data can not be always accurate or evenly collected, but it can help in monitoring biodiversity on a large scale [5]. A traditional line transects survey compared to time-triggered camera traps for monitoring found that line transects are often better for recording a wider range of visible birds, whereas camera traps can detect shy or difficult-to-observe species with less observer bias, resulting in combining both methods that improves accuracy and completeness of bird biodiversity assessments [6].

Another efficient and reliable method to collect data for large-scale bird biodiversity surveys is autonomous sound recorders, as they can collect data continuously, provide a permanent record that can be reviewed later, and reduce observer bias [7]. Artificial Intelligence (AI) has transformed environmental monitoring just like any other sector as it can process large and complex environmental dataset, enhance monitoring efficiency and decision-making, and improve the detection of biodiversity trends, pollution, and ecological changes but its effectiveness depends on data quality,

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

transparency, and responsible implementation, and combing it with Deep Learning (DL) and Machine Learning (ML) for wildlife detection and monitoring from videos, sensor data, and image which reduces time and effort [8, 9]. Using Artificial Intelligence (AI) for acoustic wildlife monitoring can automatically analyze large volume of animal sound recordings, making detection and identification more accurate and faster, as it has the potential to improve biodiversity monitoring [10].

Combining ML and IoT techniques using sensors & AI systems created a smart bid deterrence and monitoring system that helps in tracking, monitoring, repelling, and detecting birds, which results in creating a smarter, automated solution that improves wildlife management and reducing human-bird conflicts, but also has to keep in mind the challenges that involves scalability, hardware limitations, and accuracy in a real-world environment.

For example, combining AI-powered acoustic sensors with Long Range Wide Area Network (LoRaWAN) and creating a smart farm protection system which uses wireless sensor networks and sound analysis for long-range monitoring and automated bird deterrence [11, 12]. The first Bird Audio Detection (BAD) Challenge using DL demonstrated high performance (up to ~88% AUC) as DL models and CNN trained on spectrogram-based audio features with data augmentation, achieved strong generalisation across diverse acoustic environments over traditional methods, despite the challenge that remains compared to human perception due to ambiguous short calls, faint signals, and noise [13].

India's Wild Life (Protection) Act, 1972 that supports conservation efforts, can be strengthen by integrating AI/ML technologies with environmental laws, because these ML technologies support decision-making, monitoring, and prediction in forest management, and legal frameworks ensure enforcement and protection [14].

When law and technology work together it strengthens the conservation like India's legal framework, especially the Wild Life (Protection) Act, 1972, and international agreements like Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) link these technologies such as IoT, drones, AI, and other digital models help in detecting poaching, monitoring animals, and improving enforcement through data-driven systems to combat illegal wildlife trade in India [15]. Objectives of the study are as follows:

- To develop a YOLOv8-based deep learning model for automatic bird detection and species identification.
- To improve bird classification accuracy using data augmentation and annotated image datasets.

2. Literature Review

Stowell et al. (2016) in their paper evaluate existing BAD approaches, dataset, and evaluation methods, highlighting the limitations of traditional signal-processing techniques in noisy and diverse acoustic environments and provided a standardized framework that encourages the development of more robust ML and DL models for ecological monitoring [16]. Fernández et al. (2018) in their paper proposed a camera-based system for real-time monitoring of broiler chicken welfare as it uses computer vision techniques that assess welfare indicators and flock behaviour, enabling continuous and automated monitoring as they support timely interventions in poultry farming [17].

Unlike the earlier models Weinstein et al. (2022) in their article developed a DL model as their approach was trained on diverse dataset to automatically detect birds in high resolution airbone images across multiple environments and bird species through a Deep Learning object-detection model based on Convolutional Neural Networks (CNNs), which can significantly reduce manual image annotation efforts [18]. BirdNET is a DL based system that uses CNNs to identify bird species from audio recordings was introduced by Kahl et al. (2021), as it is trained on a large-scale bird sound dataset, it can automatically recognize numerous species with high accuracy across different geographic regions while also reducing the need for manual analysis of acoustic data [19]. Using the VGG-16 deep Convolutional Neural Network (CNN) and transfer learning for classifying bird species, which was proposed by Islam et al. (2019) as this model is based on bird images that automatically identifies the species based on texture, shape, and color, and combining it with transfer learning helps in achieving high classification accuracy [20].

Another bird classification system that uses acoustic dataset from BirdCLEF-2024 dataset was proposed by Abhinav and Dhavale (2024) as they also compare these algorithms Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN) and found that in terms of accuracy, precision, F-1 score, and recall, Random Forest has achieved the best performance, and a Streamlit-based application was also developed to enable real-time bird species recognition for biodiversity monitoring [21]. Malarejes et al. (2025) in their paper conducted a comparative study using SVM, KNN, and CNN, and evaluated their performance using standard metrics on classification effectiveness and accuracy showed that CNN demonstrating the superiority of DL models for bird species recognition as compared to KNN and SVM [22]. By combining multiple classifiers and ensemble ML for bird species detection to improve overall performance showed that integrating different ML models to enhance prediction accuracy and reduce classification errors compared to individual models, making that effective as it provides more robust and reliable bird species identification [23].

Proposing a hybrid fusion-based DL CNN-based models and Transformer-style attention mechanism (self-attention–based feature weighting) called SA-BirdClassifierNet proposed by Tiruvikraman & Lokesh (2026), which performed better than You Only Look Once (YOLO) and ResNeSt in classification accuracy, as it does real-time classification and monitoring of migratory birds, as the combination of those techniques improve focus on robustness in dynamic environment, enhancing classification accuracy, and important acoustics/visual patterns [24]. Karanam et al. (2025), using YOLOv8 an object detection model proposed for bird species recognition system, can identify and classify bird species from images, that enables fast and accurate detection for habitat monitoring and conservation, shows that model performs effectively [25].

A multifeature fusion and a transformer encoder proposed by Zhang et al. (2023) to improve classification accuracy as it integrates multiple acoustic features, log-mel spectrograms, Chroma, MFCC, and Tonnetz, which is then combined with CNN feature extraction (ResNet50 and EfficientNetB3) with a transformer encoder that capture both global temporal dependencies and local spectral patterns. Experimental results showed achieviement up to 97.99% accuracy on the Birdsdata dataset and 93.18% on the CBC dataset, as Principal Component Analysis (PCA) was used te reduce feature redundancy before fusion demonstrated enhances robustness and recognition performance in complex acoustic environments through feature fusion combined with transformer-based modeling significantly [26].

Aiming to improve gradient flow and feature reuse through the use of Dense CNN for BAD was proposed by Pellegrini et al. (2017) as it operates on leverages dense connections and spectrogram-based audio representations to enhance learning of subtle acoustic patterns in noisy environment, as it improved detection performance [27]. Abebe Endalamaw et al. (2026) developed an effective model which is capable of processing live audio stream and classifying bird species with minimal latency, as it optimized neural network architectures and mobile friendly interface strategies for practical field deployment. This system is enabled on citizen science application and on-device recognition, which is suitable for environmental monitoring [28].

For real-time bird species monitoring and song classification, a 5G-enabled AI-IoT framework which integrates edge device or audio capture with AI models deployed across an IoT architecture was proposed by Segura-Garcia et al. (2024), which leveraging 5G connectivity to support scalable processing and low-latency data transmission, and this method proved to be efficient and responsive as compared to the traditional standalone system [29]. A flexible hybrid edge-computing IoT architecture that combines edge and cloud processing, where deploying lightweight deep learning models at the edge (e.g., Raspberry Pi) and offloading more complex processing to the cloud, for low-cost bird detection song was proposed by Delgado-Rajó and Travieso-Gonzalez (2025) as it supports cost-effective and scalable ecological monitoring, making it more practical in resource-constrained settings [30].

A system that enables long-term, non-evasive monitoring of breeding activity, conservation management, and providing high-resolution behavioural data useful for ecological research through Raspberry Pi–based nest camera systems as an affordable tool for remote monitoring of bird nesting behaviour and conservation studies was developed and evaluated by Hereward et al. (2021) showed that low cost, custom build camera units can reliably capture continuous behaviour data in nest environment while also reducing maintenance and equipment cost compared to commercial systems [31]. A Raspberry Pi-based, Radio-Frequency Identification (RFID)-equipped bird feeder system that integrates low-cost hardware with RFID tagging, which enables continuous non-invasive data collection on behavioural pattern, visitation frequency, and species presence without direct human observation, was developed by Youngblood (2019), and it provided a scalable and affordable approach for remote monitoring [32].

An edge-enabled visual sensing system integrated with Long Range Wide Area Network (LoRaWAN) telemetry proposed by Abubeker et al. (2026), that processes visual data locally for rapid bird detection by using edge computing by enabling timely response and reducing reliance on cloud connectivity, while LoRaWAN supports low-power, long-range communication for transmitting event data for real-time parakeet detection and deterrence in banana plantations [33]. For protecting the ecosystem on-edge artificial intelligence architecture proposed by Rossi Labianca et al. (2025) for real-time bird identification, which deploys AI models directly on edge devices to perform video/audio processing locally, enabling autonomous and fast species recognition without continuous cloud dependency, which results in bandwidth usage, and improved responsiveness, and reduces latency in field conditions [34].

Integrating sensing and edge AI framework in 6G network architecture will enhance low-latency, scalable sensing system by tightly integrating AI-driven inference with next-generation communication infrastructure was reviewed by Liu et al. (2025), because combining edge intelligence, sensing, and communication allows decision-making and real-time data processing at the edge network [35]. Félix-Jiménez et al., 2024 proposed an integrated bird monitoring system using YOLOv8 Small for detection and MobileNet V3 Large for species classification on mobile devices. YOLOv8 achieved 89.68% accuracy and 60.59% mAP@50-95, demonstrating the suitability of lightweight deep-learning models for efficient bird monitoring [36].

Ma et al., 2025 proposed YOLOv8-bird for bird detection and counting at Poyang Lake, incorporating receptive-field attention, DyASF-P2 feature fusion, a lightweight detection head, and Inner-ShapeIoU loss. On the PYL-5-2023 dataset, the model achieved 94.6% precision, 89.4% recall, 94.8% mAP@0.5, and 70.4% mAP@0.5:0.95, demonstrating its effectiveness for biodiversity monitoring in complex environment [37].

3. Methodology

The bird identification system that is developed is based on object detection using the YOLOv8 framework, which identifies and categorizes various species of birds in images. The primary objective of this study is to develop a learning model that can easily recognize the bird species even under various light and environment.

To accomplish this, the images are treated in the following way: pictures of birds are gathered, cleaned up pictures, added more pictures to the collection, trained the model, inspected, the model and compared its performance to that of the original model and finally, analyzed the results. Initially, 1131 images of 17 bird species, such as the Ashy Prinia, Asian Koel, Black Drongo, etc., were taken.

These pictures were annotated so that the model could learn where the birds are. The picture from various lighting and viewpoints were added. This enabled the model to learn the appearance of birds in their natural habitats. The model used 70% images for Training, 20% images for validation purpose and 10% images for testing purpose as shown in Table 1.

Table 1. Dataset distribution for training, validation and testing purpose

Dataset	Percentage	Approx. images
Training	70%	792
Validation	20%	226
Testing	10%	113
Total	**100%**	**1,131**

Augmentation process was applied for modifying the images, such as flipping, resizing, changing lighting and colors. This gave the model a bird's-eye perspective and different weather situations. The brightness was changed for the pictures, pictures were rotated sideways up and down, cut them out, made them larger and smaller, made the colors look brighter, blurred them a little, added noise and changed the colors. This meant that the model had more images to learn from and improve its ability to identify birds. After augmentation, the pictures became 2,376 in number, which is nearly twice as many as started with. The new pictures were still labeled and boxed, but looked slightly different around the birds, which enabled the model to learn more about what birds look like.

Representative examples of the augmented training samples are illustrated in Figure 1, which demonstrate that the augmentation process generated diverse image variations while maintaining accurate annotations around each bird instance. These variations in the diversity are further confirmed in the figures, which indicate that the increased data set is a solid basis for the training of a generalized object detection model to be applied to wildlife monitoring applications in which such variations are typical.



Fig. 1 Sample training batch visualization with labeled bird species

To provide a comprehensive overview of the dataset's structural composition, Table 2 delineates the quantitative distribution of annotated instances across all 17 bird species. It is pertinent to note that the cumulative number of annotated bounding boxes (2,558 instances) exceeds the total number of augmented images (2,376). This variation is a standard characteristic of multi-object detection datasets, attributable to the co-occurrence of multiple avian subjects within individual frames.

Furthermore, while the dataset exhibits an inherent class imbalance typical of ecological imaging—where certain species, such as the Rock Dove, are highly prevalent, and others, like the Fairywren (Malurus cyaneus), are comparatively underrepresented—the applied data augmentation pipeline effectively mitigated potential detection biases.

By generating significant intra-class visual variance, the preprocessing strategy ensured robust feature extraction and improved the generalization capabilities of the YOLOv8 model across all categories.

Table 2. Complete class-wise distribution of bird species instances in the dataset

No.	Bird Species	Object Instances
1	Ashy Prinia	213
2	Asian Koel	165
3	Black Crowned Tchagra	18
4	Black Drongo	195
5	Black Kite	57
6	Coppersmith Barbet	216
7	Fairywren Malurus Cyaneus	3
8	Himalayan Monal	201
9	Himalyan Bulbul	66
10	House Sparrow	75
11	Indian Grey Hornbill	213
12	Indian Robin	39
13	Jungle Myna	75
14	Red Vented Bulbul	245
15	Rock Dove	297
16	Rose-ringed Parakeet	246
17	Steppe Eagle	234
	Total Annotated Instances	**2,558**

The augmented dataset was used to train the YOLOv8 object detection model in the cloud computing environment of Google Colab. Google Colab is actually very convenient, as they have GPU support.

Thus, can train learning models which require a lot of computing power without purchasing expensive computers. YOLOv8 was selected as a suitable model for bird monitoring. YOLOv8 comes with numerous advantages over versions. Can detect objects with precision and speed. It's also more efficient in the aspect of training. YOLOv8 stands out from object detectors.

Typically, most object detectors perform their operations in two steps. First, they find the area where an object might be. They then attempt to determine the name of the object. In one step, YOLOv8 does all this. It identifies a given object. Figure out what it is at the same time. This leads to a much faster and efficient YOLOv8. It's quite good for wildlife watching, and for smart surveillance system. YOLOv8 consists of three parts.

The first one is the backbone network. Looks at pictures. Finds important features. The second is called the Neck Network. It overlays features from layers and merges them. The third part is the detection head. It relies on all of the information to determine where objects are and what they are. By looking at pictures in different ways, the YOLOv8 model can find birds of all sizes in complicated environments. This is really useful for YOLOv8 object detection model.

It is good at detecting birds. The YOLOv8n architecture was selected to balance detection accuracy and computational efficiency for edge deployment on the Raspberry Pi 5. The model was trained for 50 epochs using an input image resolution of 640 × 640 pixels.

The network parameters were optimized using the Stochastic Gradient Descent (SGD) optimizer with a batch size of 16, an initial learning rate of 0.01, and a momentum of 0.937.

The training objective was based on minimizing the bounding box loss, classification loss, and Distribution Focal Loss (DFL) to improve object localization and bird species classification performance.

The box test evaluates the performance of the network in locating the bird, and compares the bird's location with the location the network believes it has found. The classification loss is used to determine whether the network is choosing the type of bird.

The Distribution Focal Loss makes the network better, only at boxes around the birds, to find the birds very precisely. During training, the system always checks how well it is doing after each epoch and saves the version of the model as best.pt. This way, the last one it made isn't the final model because it didn't do as well.

The network desires to ensure the end model is the most accurate and, therefore, chooses the best one available. The convergence behaviour of the training process is illustrated in Figure 2, where the gradual reduction in training and validation losses demonstrates stable optimization throughout the learning process.

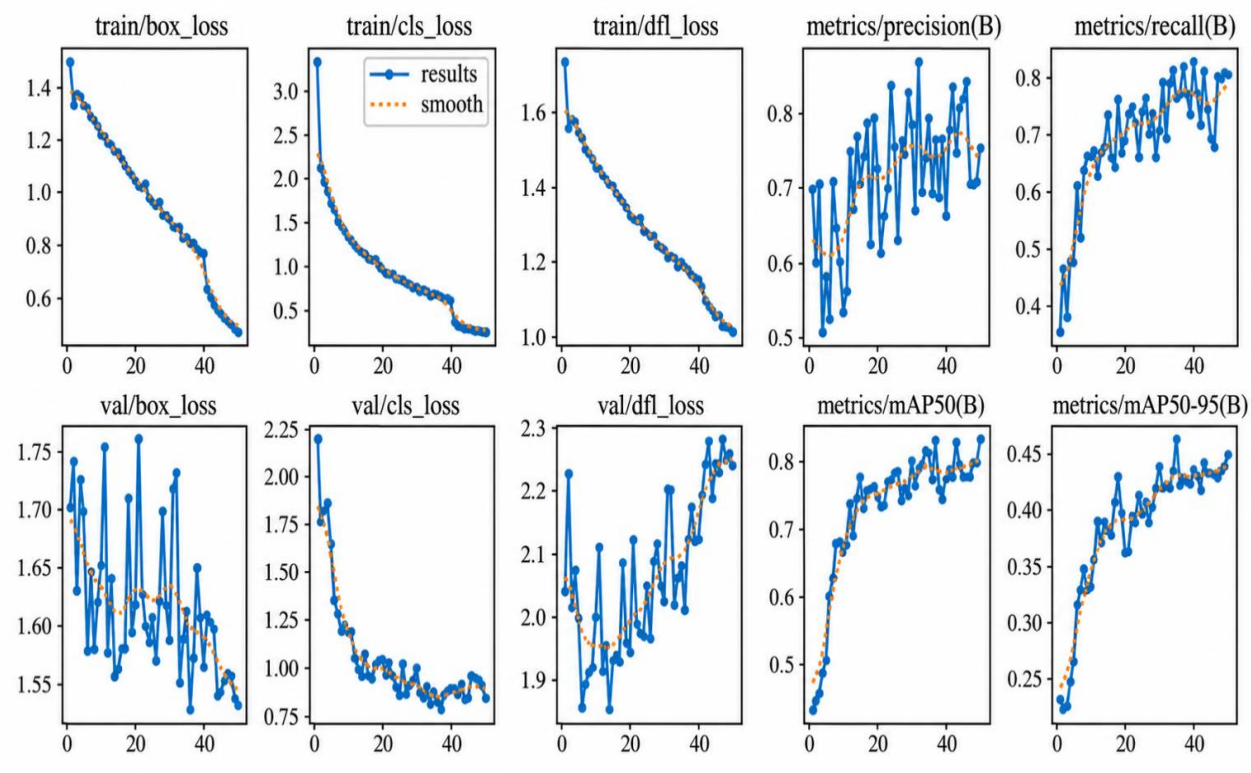

Fig. 2 Training and validation performance graphs including precision, recall, mAP, and losses

Moreover, the results demonstrate that the network can learn discriminative visual representations that are able to resolve multiple bird species with increasing confidence, as revealed by the simultaneous improvement of precision and recall, as well as mean Average Precision (mAP), in the same figure. After training, the model's performance was assessed using various commonly used object detection metrics to obtain a comprehensive evaluation of both localization and classification performance. The evaluation metrics were Precision, Recall, Mean Average Precision at an Intersection over Union (IoU) threshold of 0.5 (mAP@0.5), Mean Average Precision averaged over IoUs from 0.5 to 0.95 (mAP@0.5:0.95), F1-score, confusion matrix analysis, and confidence-based evaluation curves.

Precision refers to the percentage of birds correctly identified divided by the number of predicted birds and is an estimate of the number of false-positive predictions, while Recall refers to the percentage of birds that were correctly identified divided by the number of birds in the image.

The mAP@0.5 metric combines the detection accuracy at a typical IoU threshold, while mAP@0.5:0.95 provides a tougher test by averaging the detection performance at multiple localization thresholds. The F1-score is used to quantify the overall effectiveness of the detector by taking into account both Precision and Recall. The mathematical formulations for these metrics are defined as follows:

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-Score} = 2 * [(\text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})]$$

IoU = Area of Intersection between Predicted and Ground-Truth / Area of Union between Predicted and Ground-Truth

$$mAP@0.5 = (1 / N) * \Sigma\ AP_i@0.5$$

$$mAP@0.5:0.95 = (1 / 10N) * \Sigma\ \Sigma\ AP_i@t$$

Where:

- TP denotes True Positives, FP denotes False Positives, and FN denotes False Negatives.
- N = 17 represents the total number of bird species evaluated.
- AP_i represents the Average Precision for the i-th species.
- t represents the Intersection over Union (IoU) threshold.

Aside from the numerical performance metrics, graphical analysis is also carried out to be able to understand the behaviour of the trained model well. The dataset distribution shown in Figure 3 was analyzed to examine class balance, annotation density, and bounding-box characteristics.

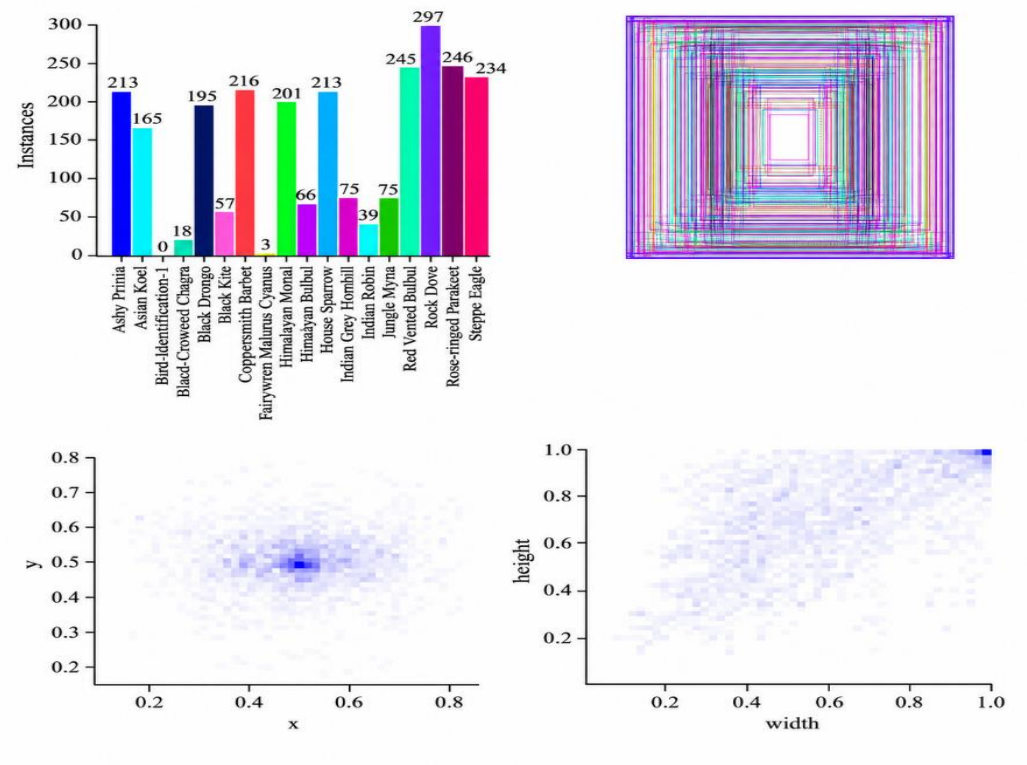

Fig. 3 Distribution of object instances and bounding box dimensions in the dataset

The normalized confusion matrix presented in Figure 4 was used to evaluate class-wise prediction accuracy and identify visually similar bird species responsible for classification errors.

The Precision–Confidence, Precision–Recall curve and F1–Confidence curve was further examined to determine the optimal confidence threshold for practical deployment and to analyze the trade-off between detection accuracy and recall. Finally, representative training samples and validation prediction results were visually inspected to verify annotation quality, localization accuracy, and prediction consistency under varying backgrounds, lighting conditions, and bird poses. The combination of quantitative performance metrics and qualitative visual analysis provides a comprehensive evaluation of the proposed YOLOv8-based bird identification system and demonstrates its suitability for automated wildlife monitoring, biodiversity assessment, and intelligent ecological observation applications.

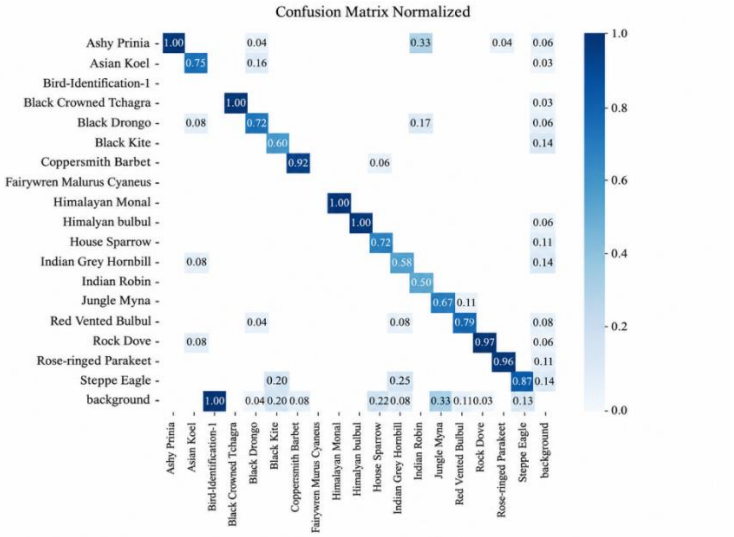

Fig. 4 Normalized confusion matrix representing prediction accuracy for each bird species

To demonstrate the practical applicability of the proposed bird identification system, a dedicated computer vision device was developed using a Raspberry Pi 5 integrated with a Raspberry Pi Camera Module, as shown in Figure 5. The edge computing platform used for acquiring images, preprocessing the images, and running the trained YOLOv8 model in real-time was the Raspberry Pi. The model was trained on Google Colab, and optimized model weights (best.pt) were transferred to the Raspberry Pi for inference. The camera module took continuous live images and video frames, which were then processed by the Raspberry Pi to classify and identify bird species in real time. No cloud computing was required for the predicted bird species, confidence score, and bounding box to be created on the device.

This choice of Raspberry Pi was made because of its small size, low power, cheap, and enough computing power to power edge artificial intelligence applications. The deployment of the trained model to embedded hardware is proof of the feasibility of implementing the proposed system in practical applications like wildlife sanctuaries, forests, agriculture fields and monitoring stations for biodiversity in places where internet connectivity is not always available. Moreover, the developed hardware prototype demonstrates that the proposed bird identification model based on YOLOv8 is not only computationally efficient but also is capable of being deployed on low-cost hardware in real-time, which facilitates portable, autonomous, and intelligent wildlife monitoring applications.



Fig. 5 Developed system

4. Results & Discussion

The test results of the proposed bird identification system using YOLOv8 model clearly shows the model learned to detect and classify the different bird species with a high level of accuracy and robustness. The model was trained for 50 epochs on the augmented dataset of 2376 images and continuously evaluated with standard metrics in object detection, such as precision, recall, mean Average Precision (mAP), F1-score, confusion matrix, and evaluation curve based on the confidence.

The model never diverged during the entire training process, signaling good convergence of the network architecture and training approach for multi-class bird identification. The training and validation performance curves clearly illustrate the progressive improvement in the learning behaviour of the model.

As can be seen, the training box loss, classification loss and Distribution Focal Loss (DFL) all reduced as the number of epochs increased, which means the network learned more and more representative visual features to accurately localize and classify birds. At the same time, the validation losses exhibited a downtrend without abrupt swings and divergence, indicating good generalization from the validation set of images, and relatively little overfitting. As both training and validation loss converge gradually, enough training epochs were selected to ensure learning meaningful feature representations without compromising model stability.

Moreover, the precision and recall curves exhibit continuous improvement throughout the training process, reaching approximately 0.85 and 0.80, respectively. These values show that most of the predictions were both correct and that there were also a high percentage of birds detected as well.

The model also achieved a score of mAP@0.5 around 0.81, which indicates that it has strong object detection ability in each bird species in the dataset. Even though the mAP@0.5:0.95 value is relatively low (around 0.45), it

reflects a more challenging evaluation criterion that includes several IoU thresholds, offering a more complete picture of the model's performance in maintaining localization accuracy even with more demanding evaluation requirements.

The training results validate that YOLOv8 managed to simultaneously learn object localization and bird species classification, thus enabling it to be used in real bird identification applications.

To evaluate class-specific performance, Table 3 details the class-wise Average Precision (AP@0.5) metrics. The model achieved high accuracy for distinct species (e.g., Himalayan Monal at 0.995), while lower scores (e.g., Black Kite at 0.461) reflect challenges such as background similarity or fewer training instances.

Table 3. Class-Wise Average Precision (AP@0.5) performance of the YOLOv8 model

Bird Species	AP@0.5
Ashy Prinia	0.988
Asian Koel	0.919
Black Crowned Tchagra	0.995
Black Drongo	0.921
Black Kite	0.461
Coppersmith Barbet	0.915
Himalayan Monal	0.995
Himalyan Bulbul	0.995
House Sparrow	0.775
Indian Grey Hornbill	0.703
Indian Robin	0.828
Jungle Myna	0.706
Red Vented Bulbul	0.843
Rock Dove	0.964
Rose-ringed Parakeet	0.967

The dataset characteristics also played a significant role in determining the overall detection performance of the proposed model. The dataset distribution analysis provides valuable insights into the composition of the training data and the spatial distribution of annotated bird instances.

The figure shows the frequency of each bird class as well as the distributions of the dimensions of the bounding boxes and the locations of the bird's center within the boxes. Clearly, there were some species with more instances than others and with more training samples than others, such as Rock Dove.

This class frequency variation is a common characteristic in wildlife datasets, as some species of birds are more easily observed or are more prevalent in image collections that are publicly available. While class imbalance may hinder the performance of minority classes in the classification task, this limitation was mitigated by making use of the extensive augmentation strategy in pre-processing, which resulted in generating variations of images for the minority classes. In addition, shows the distribution of the dimensions of the bounding boxes, meaning that the dataset contains both small and larger birds as well as birds in different locations within the image, which enables the network to learn multi-scale representations of objects.

The distribution of object centers within the various regions of the image is also indicative of the fact that the bird was not only placed near the image center, but throughout the image, as well. This diversity is crucial for enhancing the generalization of the models and achieving accurate detection in real-world scenarios with unpredictable bird positions.

Thus, the proposed model achieved good detection performance, which was contributed by the balanced spatial distribution and augmentation process. A more detailed evaluation of classification performance is provided by the normalized confusion matrix.

The confusion matrix provides a class-wise visualisation of the accuracy of predictions, matching the predictions with the ground truth labels. The confusion matrix shows several birds, such as Himalayan Monal, Ashy Prinia, Rock Dove and Rose-ringed Parakeet, are classified with very high accuracy, which suggests that the feature representations learned were highly discriminative for these birds.

The confusion matrix presented a higher performance for classification, which was presumably related to their unique colour patterns, shapes and morphological features. However, some minor classification errors were noticed among the visually similar bird species, such as Black Drongo and Jungle Myna. The species have similar body size, color, silhouette and viewing angles, requiring relatively more difficulty to separate species, particularly in poorly illuminated images or when taking long-range shots.

However, the amount of misclassifications was still relatively small compared to the number of correct predictions, indicating that the proposed YOLOv8 model successfully learned the visual features of the species. One of the other key findings is the relatively small rate of background misclassification.

Vegetation, tree branches, sky and buildings in the background were not commonly identified as bird objects, suggesting that the detector was able to distinguish foreground bird objects from background features.

This is especially crucial for wildlife monitoring systems, where complicated natural backgrounds can cause false positive detections. The confusion matrix thus verifies that the proposed model has a very good discriminative capability, with a good localization accuracy for various bird categories.

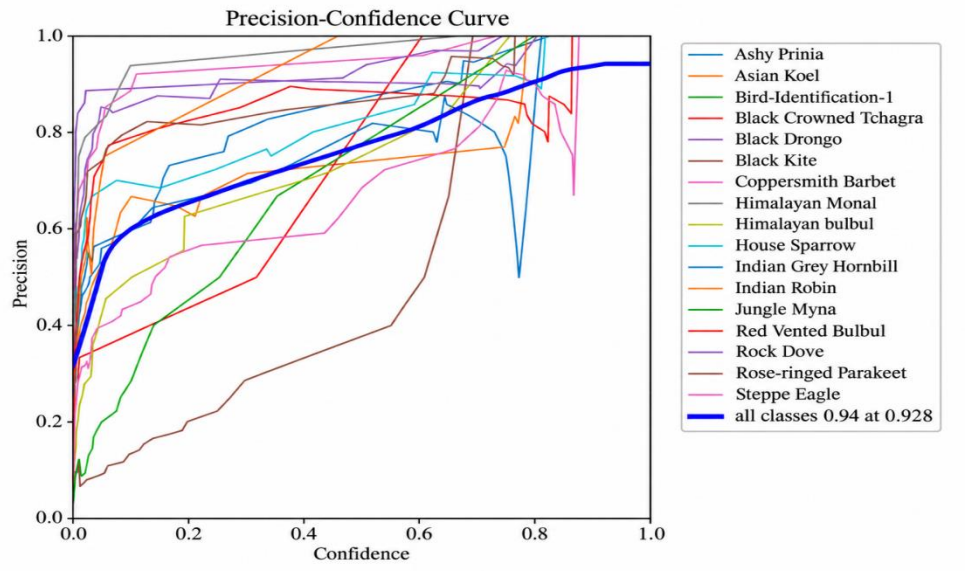

Fig. 6 Precision-Confidence curve for all bird classes

The Precision-Confidence relationship illustrated in Figure 6, and nearly all of the bird detections were predicted correctly, with an overall precision near 94% at the highest confidence values (close to 0.93). The behaviour indicates that the higher the confidence threshold, the lower the number of false-positive detections, and the better the classification accuracy. In applications for ecological monitoring, where misidentification of birds can cause problems in the assessment of biodiversity, it is especially beneficial to have such precision. If the confidence level is too high, however, the total number of birds detected may be lowered as some birds that are valid but not scored with a high enough level of confidence might not be recorded. Hence, there is a need for a proper trade-off between the confidence threshold and detection rate when the system is being put into use. This trade-off is further illustrated by the Precision–Recall curve shown in Figure 7, which demonstrates the relationship between prediction precision and detection completeness.

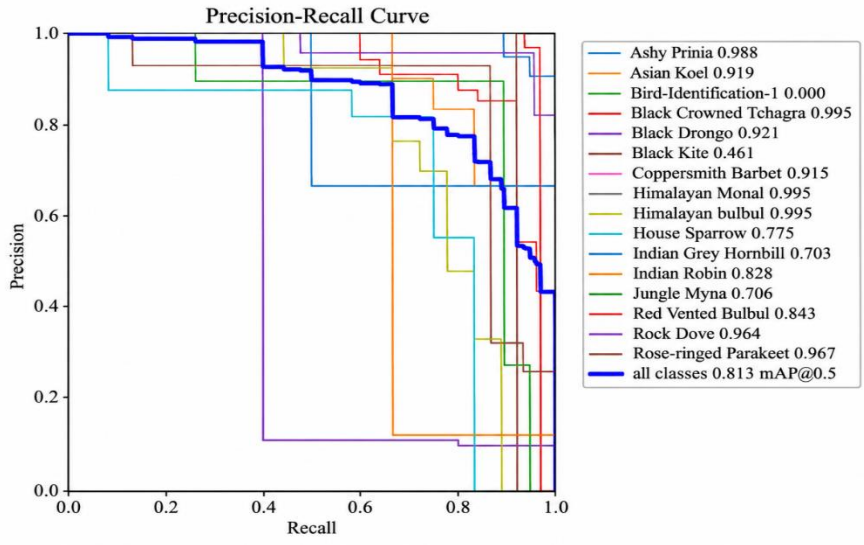

Fig. 7 Precision-Recall curve for all bird classes

The overall mAP@0.5 of about 0.813 from the Precision–Recall curve shows that the detector correctly optimized over all bird species, their ability to avoid false-positive and false-negative predictions. Precision and recall scores were close to one for several classes, which represent good localization and classification performance, but lower scores for a few species were caused by the small number of training samples and the high similarity between the classes visually. However, the global trend of the Precision–Recall curve shows that the proposed detector performs well under a broad range of

confidence thresholds and operating conditions. The effectiveness of the trained model is further demonstrated by the F1-Confidence curve, which combines precision and recall into a single performance indicator.

The F1 score takes into account both false positives and false negatives, unlike precision or recall scores. The result in Figure 8 indicates that the highest F1-score of around 0.77 was obtained with a confidence threshold around 0.56.

This is a point that balances the prediction accuracy and detection completeness most effectively. This threshold is what can be selected during real-time deployment so that the system is able to maximize the number of times birds are correctly detected, whilst minimizing false detections and false classification. Furthermore, the relatively smooth performance of the F1 curve also suggests there is no sudden drop in model performance across various confidence levels, highlighting the stability of the trained YOLOv8 detector.

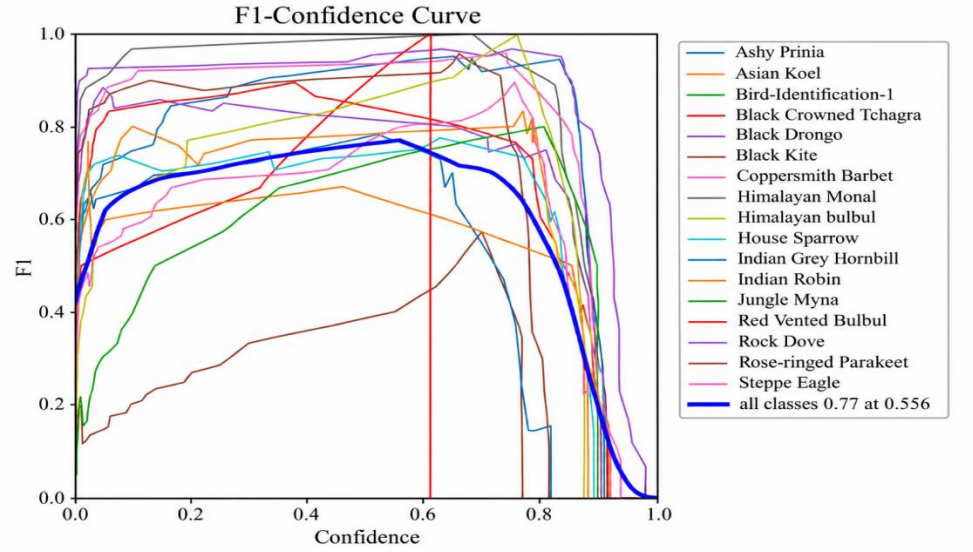

Fig. 8 F1-Confidence curve for all bird classes

The qualitative performance of the proposed system was evaluated using the training batch visualization. These are representative augmented images of the birds used for training the models, along with their bounding-box annotations and bird species labels. When analysed carefully, these visualisations have been able to show that the annotations have been placed correctly around each bird, even in the case of differences in object size, orientation, image quality and background environmental features.

The variety of the augmented samples displayed in these figures shows that the pipeline of preprocessing was effective to get the network to sample many variations in the real world of images, which helps it generalize better. Multiple bird species were placed in the same training batch, which also allowed the detector to learn class-specific discriminative features and, at the same time, enhance its object localization performance in cluttered scenes. Such visualizations are thus a qualitative piece of evidence that data cleaning played a vital role in the overall success of the trained model.

The final validation prediction results presented provide additional evidence regarding the practical applicability of the proposed bird identification system. The trained YOLOv8 model shows its ability to accurately localize birds with tightly-fitted bounding boxes while also providing high confidence scores for its predictions of bird species. The detector was able to correctly detect the birds under different illumination, vegetation, complex background, and different viewing perspectives, with excellent generalization ability. The detector was also able to reliably locate the birds, even in relatively small parts of the image or when the birds were partially obscured. It is especially helpful for applications like monitoring wildlife in an uncontrolled environment. The high level of confidence in the consistently high scores seen on the validation images suggests that the model was able to apply what it had learned to unseen data and that the model transferred the knowledge it learned from the training images to other images.

The quantitative evaluation results and qualitative visualization results show that the proposed bird identification model based on YOLOv8 has the advantages of reliability, accuracy and low calculation complexity, and can effectively realize automatic bird detection and species classification. The results achieved demonstrate the feasibility of the proposed approach for practical applications like biodiversity conservation, ecological surveys, smart wildlife monitoring, environmental research and real-time bird observation systems, and it serves as a solid basis for further development

of the approach using larger data sets, more species, higher image resolutions, and edge-computing platforms. The proposed bird identification system was also implemented in real-time on the computer vision device, which is based on the Raspberry Pi 5 microcomputer, to confirm the practicality of the system.

The weights of the YOLOv8 model trained on the desktop platform were transferred to the embedded platform, and the system was tested with video obtained from the Raspberry Pi Camera Module.

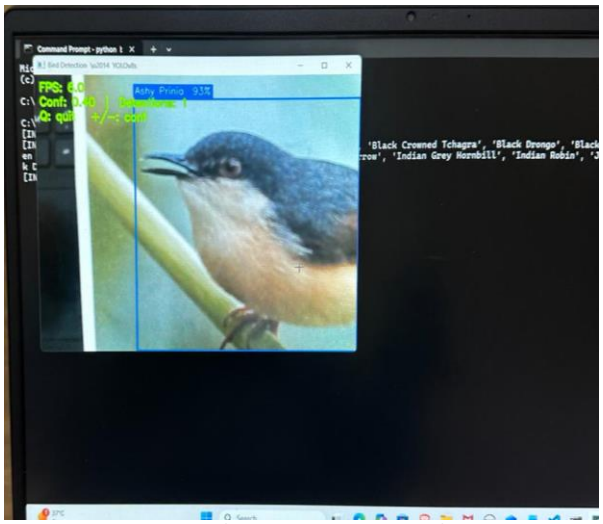

Fig. 9 The system successfully detected and classified an Ashy Prinia with a prediction confidence of approximately 93%

The deployment results showed that the proposed model could be deployed directly on the edge device to detect birds and classify them in real-time without depending on cloud computing resources. As shown in Figure 9, the system successfully detected and classified an Ashy Prinia with a prediction confidence of approximately 93%, while achieving an inference speed of nearly 6 Frames Per Second (FPS).

Similarly, Figure 10 illustrates the successful identification of a Rock Dove with a confidence score of approximately 82% at an inference speed of about 4.7 FPS. The trained YOLOv8 model was robust enough to accurately bound the birds in both cases and correctly classify them into the corresponding species, as the class labels generated by the model matched the real-world results. The speed of inference of the Raspberry Pi is not as fast as high-performance GPUs, but it is fast enough for practical applications of wildlife monitoring that involve moderate movement of birds. The implementation of the proposed model on a small and low-power embedded platform proves its computational efficiency and applicability to edge-AI applications. The YOLOv8 model that was deployed was able to make an inference at about 4.7-6.0 Frames Per Second (FPS) on the tested bird detection scenarios.

The frame-processing interval, estimated from the observed FPS, was about 166.7-212.8 ms/frame. The system was used to detect and classify birds using local resources on the Raspberry Pi without making any inferences in the cloud. The results show the feasibility of implementing the suggested bird identification model on a low-cost edge computing platform for field-based wildlife monitoring. The experimental results obtained in real-time confirm that the developed system can be effectively deployed to monitor birds in an autonomous manner in forests, wildlife sanctuaries, agricultural fields and in biodiversity conservation projects. In addition, the hardware implementation shows that the proposed system provides hand-off-hand, cost-effective, and scalable way to achieve effective and correct bird detection and classification without depending on a continuous Internet connection.

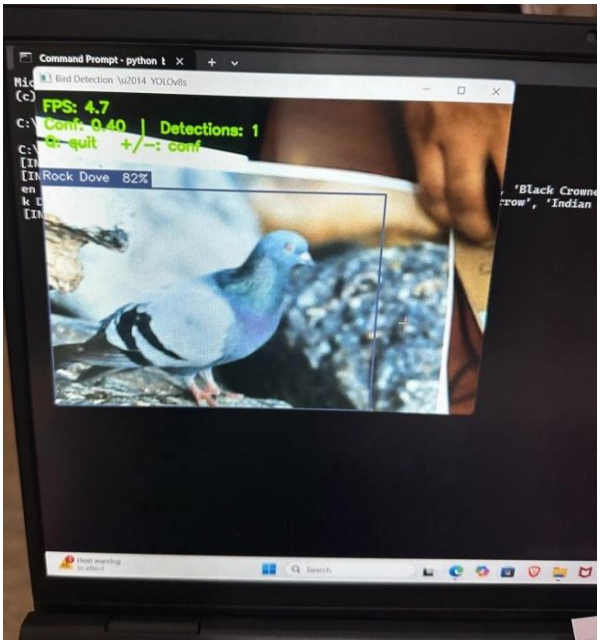

Fig. 10 The system identifying a Rock Dove with a confidence score of approximately 82%

5. Conclusion

This study was able to create an automated bird identification system by implementing the YOLOv8 (Object Detection Framework) to detect and classify 17 different species of birds. The number of images was increased from 1,131 with annotations on Roboflow to 2,376 images to train the proposed model. All the data augmentation techniques were found to be effective in creating more data variations in the training data, resulted in better generalization and robustness of the model in different environmental conditions. Experimental results demonstrated that the proposed model exhibited good precision, recall and mean Average Precision (mAP) to accurately detect and classify bird species in natural scenes with a high number of birds. The outcomes of training and validation indicated that the model converged with minimal overfitting, the confusion matrix demonstrated that

the model could classify a number of bird species with high accuracy and the evaluation curve using the confidence confirmed the ability of the model to classify the species of birds. Moreover, a qualitative validation of the prediction results showed good accuracy even in different scenarios (various background lights, light and viewing angles), demonstrating the robustness of the YOLOv8 approach in real-world applications of wildlife monitoring. The proposed system gave satisfactory results, but has scope for improvement. Possible future work includes increasing the number of bird species by adding more images for underrepresented classes to decrease class imbalance and to increase the classification accuracy, as well as expanding the dataset by adding more images for underrepresented classes. Further, the detection performance can be boosted by using more advanced data augmentation techniques (Mosaic, MixUp and Copy-Paste augmentation), longer training durations and higher-resolution images. Further, attention mechanisms, transformer-based object detection model or ensembling techniques could be used to enhance the feature extraction and classification accuracy. The trained model can then be extended to edge computing devices like Raspberry Pi, NVIDIA Jetson Nano and to mobile platforms for real time forest bird and wildlife identification, ecological monitoring and forest management. Moreover, the suggested system may be integrated with Internet of Things (IoT) sensors, surveillance cameras and/or even drones which monitor the environment and are capable of intelligent surveillance and may be beneficial for future environmental study and wildlife protection programmes, that is, monitoring, assessment and conservation of biodiversity, species diversity and intelligent environmental surveillance in real-time.

Conflicts of Interest

There is no conflict of interest associated with this review article.

Funding Statement

No funding is received for this article.

References

[1] Judith E. Krauss, "Unpacking SDG 15, its Targets and Indicators: Tracing Ideas of Conservation," *Globalizations*, vol. 19, no. 8, pp. 1179-1194, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Shaoyang Liu, Jianjun Bai, and Jun Chen, "Measuring SDG 15 at the County Scale: Localization and Practice of SDGs Indicators based on Geospatial Information," *ISPRS International Journal of Geo-Information*, vol. 8, no. 11, pp. 1-27, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] Vânia Proença et al., "Global Biodiversity Monitoring: From Data Sources to Essential Biodiversity Variables," *Biological Conservation*, vol. 213, pp. 256-263, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[4] Petteri Vihervaara et al., "How Essential Biodiversity Variables and Remote Sensing Can Help National Biodiversity Monitoring," *Global Ecology and Conservation*, vol. 10, pp. 43-59, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] Alison Johnston, Eleni Matechou, and Emily B. Dennis, "Outstanding Challenges and Future Directions for Biodiversity Monitoring using Citizen Science Data," *Methods in Ecology and Evolution*, vol. 14, no. 1, pp. 103-116, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[6] Nana Wix, and Michael Reich, "Time-Triggered Camera Traps Versus Line Transects-Advantages and Limitations of Multi-Method Studies for Bird Surveys," *Bird Study*, vol. 66, no. 2, pp. 207-223, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[7] Kevin Darras et al., "Autonomous Sound Recording Outperforms Human Observation for Sampling Birds: A Systematic Map and user Guide," *Ecological Applications*, vol. 29, no. 6, pp. 1247-1265, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Emran Alotaibi, and Nadia Nassif, "Artificial Intelligence in Environmental Monitoring: In-Depth Analysis," *Discover Artificial Intelligence*, vol. 4, no. 1, pp. 1-26, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[9] M.A. Obisa, E.C. Too, and B.O. Osero, *A Review of Artificial Intelligence, Machine Learning, and Deep Learning and their Applications in Detecting Wildlife Animals*, Springer, pp. 1-32, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Sandhya Sharma, Kazuhiko Sato, and Bishnu Prasad Gautam, "A Methodological Literature Review of Acoustic Wildlife Monitoring using Artificial Intelligence Tools and Techniques," *Sustainability*, vol. 15, no. 9, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Samson O. Ooko et al., "IoT and Machine Learning for Smart Bird Monitoring and Repellence: Techniques, Challenges, and Opportunities," *IoT*, vol. 6, no. 3, pp. 1-21, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[12] Theresa Jose, and J. Albert Mayan, "Lorawan and Artificial Intelligence Integrated Smart Acoustic Sensor Network for Bird Species Identification and Deterrence System for Farm Protection," *Measurement*, vol. 256, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[13] Dan Stowell et al., "Automatic Acoustic Detection of Birds through Deep Learning: The First Bird Audio Detection Challenge," *Methods in Ecology and Evolution*, vol. 10, no. 3, pp. 368-380, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[14] Monika Kothiyal et al., "The Conservation of Forest Biodiversity of India through the Integration of Law and AI/ML," *International Conference on Data Mining and Information Security*, Phnom Penh, Cambodia, vol. 1387, pp. 129-147, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[15] Vikram Delu, Yomto Mayi, and Shantabala Devi Gurumayum, *Conservation Technologies: AI, Drones and Beyond in Wildlife Conservation and Mitigating Illegal Wildlife Trade*, Wildlife Trade in India, Springer, pp. 311-322, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[16] Dan Stowell et al., "Bird Detection in Audio: A Survey and a Challenge," *2016 IEEE 26th International Workshop on Machine Learning for Signal Processing*, Vietri Sul Mare, Italy, pp. 1-6, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[17] Alberto Peña Fernández et al., "Real-Time Monitoring of Broiler Flock's Welfare Status using Camera-based Technology," *Biosystems Engineering*, vol. 173, pp. 103-114, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[18] Ben G. Weinstein et al., "A General Deep Learning Model for Bird Detection in High-Resolution Airborne Imagery," *Ecological Applications*, vol. 32, no. 8, pp. 1-23, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[19] Stefan Kahl et al., "BirdNET: A Deep Learning Solution for Avian Diversity Monitoring," *Ecological Informatics*, vol. 61, pp. 1-10, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[20] Shazzadul Islam et al., "Bird Species Classification from an Image using VGG-16 Network," *Proceedings of the 7th International Conference on Computer and Communications Management*, Association for Computing Machinery, New York, NY, United States, pp. 38-42, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[21] Penchala Abhinav, and Sunitha Dhavale, "Bird Species Classification and Prediction using Machine Learning Algorithms," *2024 IEEE Pune Section International Conference*, Pune, India, pp. 1-5, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] John Stephen Buslon Malarejes et al., "A Comparative Study of Bird Species Classification using K-Nearest Neighbors, Convolutional Neural Networks, and Support Vector Machines," *2025 IEEE Open Conference of Electrical, Electronic and Information Sciences*, Vilnius, Lithuania, pp. 1-6, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[23] Kotha Lavanya et al., "Ensemble Machine Learning Approach for Biological Bird Species Detection and Classification," *2024 3rd Edition of IEEE Delhi Section Flagship Conference*, New Delhi, India, pp. 1-6, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] S. Shree Pranav, V. Tiruvikraman, and S. Lokesh, "SA-BirdClassifiernet: A Hybrid Attentiondriven Deep Learning Model for Real-Time Classification and Monitoring of Migratory Birds," *2026 Second International Conference on Intelligent Systems for Communication, IoT and Security*, Coimbatore, India, pp. 1-6, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[25] Vamsi Krishna Karanam et al., "Bird Species Recognition using YOLOv8: A Deep Learning Approach for Habitat Conservation and Preservation," *International Conference on Computer Vision and Robotics*, Goa, India, vol. 1644, pp. 331-341, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] Shaokai Zhang et al., "A Novel Bird Sound Recognition Method based on Multifeature Fusion and a Transformer Encoder," *Sensors*, vol. 23, no. 19, pp. 1-20, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[27] Thomas Pellegrini, "Densely Connected CNNs for Bird Audio Detection," *2017 25th European Signal Processing Conference*, Kos, Greece, pp. 1734-1738, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[28] Hailemariam Abebe Endalamaw, Chuan-Kai Yang, and Cheng-Hung Hsu, "A Real-Time Bird Sound Recognition App Via Deep Learning Techniques," *Multimedia Tools and Applications*, vol. 85, no. 2, pp. 1-34, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[29] Jaume Segura-Garcia et al., "5G AI-IoT System for Bird Species Monitoring and Song Classification," *Sensors*, vol. 24, no. 11, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[30] Francisco A. Delgado-Rajó, and Carlos M. Travieso-Gonzalez, "Flexible Hybrid Edge Computing IoT Architecture for Low-Cost Bird Songs Detection System," *Ecological Informatics*, vol. 90, pp. 1-11, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[31] Hannah F.R. Hereward et al, "Raspberry Pi Nest Cameras: An Affordable Tool for Remote Behavioral and Conservation Monitoring of Bird Nests," *Ecology and Evolution*, vol. 11, no. 21, pp. 14585-14597, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Mason Youngblood, "A Raspberry Pi-based, RFID-Equipped Birdfeeder for the Remote Monitoring of Wild Bird Populations," *Ringing and Migration*, vol. 34, no. 1, pp. 25-32, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[33] K.M. Abubeker et al., "An Edge-Enabled Visual Sensing and LoRAWAN Telemetry Framework for Real-Time Parakeet Detection and Deterrence in Banana Plantations," *IEEE Sensors Journal*, vol. 26, no. 11, pp. 17734-17741, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[34] L. Rossi Labianca et al., "On-Edge Artificial Intelligence Architecture for Real-Time Bird Identification for Ecosystem Protection," *International Symposium on Distributed Computing and Artificial Intelligence*, Springer, Cham, vol. 1630, pp. 55-66, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[35] Zhiyan Liu et al., "Integrated Sensing and Edge AI: Realizing Intelligent Perception in 6G," *IEEE Communications Surveys and Tutorials*, vol. 28, pp. 2725-2770, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[36] Axel Frederick Félix-Jiménez et al., "Integration of YOLOv8 Small and MobileNet V3 Large for Efficient Bird Detection and Classification on Mobile Devices," *AI*, vol. 6, no. 3, pp. 1-27, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[37] Jianchao Ma et al., "An Improved Bird Detection Method using Surveillance Videos from Poyang Lake based on YOLOv8," *Animals*, vol. 14, no. 23, pp. 1-21, 2024. [CrossRef] [Google Scholar] [Publisher Link]