



EDITED BY
Anand Krishnan
Jawaharlal Nehru Centre for Advanced
Scientific Research, Bengaluru, India.

***CORRESPONDENCE**

Sagar Rajpurkar
✉ sagarvr@wii.gov.in

RECEIVED 26 September 2024
ACCEPTED 24 January 2025
ONLINE EARLY 02 February 2025
PUBLISHED 11 March 2025

CITATION

Parab, T. P., Hermon, G. B., Bhawe, M., Nagarale, S., Chattopadhyay, A. & Rajpurkar, S. (2025). Improved detection and classification of cryptic '*Phylloscopus*' warblers with integrated computer vision and deep learning method. *Journal of Wildlife Science*, 2(1), 10-19.
<https://doi.org/10.63033/JWLS.GGIG1171>

COPYRIGHT

© 2025 Parab, Hermon, Bhawe, Nagarale, Chattopadhyay & Rajpurkar. This is an open-access article, immediately and freely available to read, download, and share. The information contained in this article is distributed under the terms of the Creative Commons Attribution License (CC BY 4.0 ©), allowing for unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited in accordance with accepted academic practice. Copyright is retained by the author(s).

FUNDING

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

PUBLISHED BY

Wildlife Institute of India, Dehradun, 248 001
INDIA

PUBLISHER'S NOTE

The Publisher, Journal of Wildlife Science or Editors cannot be held responsible for any errors or consequences arising from the use of the information contained in this article. All claims expressed in this article are solely those of the author(s) and do not necessarily represent those of their affiliated organisations or those of the publisher, the editors and the reviewers. Any product that may be evaluated or used in this article or claim made by its manufacturer is not guaranteed or endorsed by the publisher.

Improved detection and classification of cryptic '*Phylloscopus*' warblers with integrated computer vision and deep learning method

Tushar P. Parab^{1,2} , Glen Bennet Hermon³ , Maitreyee Bhawe¹ , Shashank Nagarale¹ , Arnab Chattopadhyay¹ & Sagar Rajpurkar^{1,4*}

¹Wildlife Institute of India, Chandrabani, Dehradun, 248001

²Academy of Scientific and Innovative Research (AcSIR), New Delhi, India

³Unscene Innovation, 78 Green Park, Dehradun, 248001

⁴Saurashtra University, Rajkot, Gujarat, 360005

Abstract

The rapid advancement of image recognition technology, especially with the help of YOLO algorithms, has greatly improved the ability to identify and recognize different species. This research focuses on the *Phylloscopus burkii* cryptic species complex, distinguished by little phenotypic diversity making precise identification challenging. In order to make it easier to distinguish between members of this complex, we suggest an enhanced YOLOv5-based model that employs a part-based approach that keeps part detection and identification classification apart. Our goals included determining whether images taken in natural settings can produce statistically reliable identifications and examining the variation of image characteristics of the species within the burkii complex. A dataset of 289 annotated images was augmented to build 2,890 photographs for training purposes. After training on Amazon Elastic Cloud, the model achieved a mean average precision (mAP) of 94 and a recall of 97.3. The model demonstrated strong precision and recall metrics across species, and the results showed that it worked best at a confidence threshold of 0.50. *P. valentini* and *P. tephrocephalus*, in particular, showed considerable overlap, which sometimes resulted in misidentifications, although *P. poliogenys* and *P. whistleri* demonstrated excellent accuracy. We recommend training models based on species geographic distributions to improve identification accuracy. This strategy might successfully lower the rate of erroneous identifications, enhancing the caliber of data produced by citizen science projects. In the end, our results support current efforts to conserve biodiversity by offering a solid framework for automating species identification and guiding ecological study. Our goal is to improve species monitoring and data collection precision by utilizing innovative methods based on machine learning.

Keywords: Cryptic species, deep learning, image recognition, *Phylloscopus burkii* complex, YOLOv5.

Introduction

Studies on bird populations help us understand key indicators for environmental monitoring (Koskimies, 1989), and shifts in bird communities help track climate change (Lindström *et al.*, 2013). These surveys that monitor biodiversity often depend on numbers and counts (Larsen *et al.*, 2012). Through citizen science initiatives all around the world, the appeal of adopting a "Big Data" strategy over the past few years has made it possible to collect enormous amounts of data online. But all too frequently, this data lacks sufficient detail to clarify how the observations were made (Kelling *et al.*, 2018), and identification errors that have occurred (see de Freitas *et al.*, 2022). The difficulty of working with morphologically cryptic species is made apparent by this significant error rate in field identification, especially for smaller polytypic species like *Phylloscopus burkii*.

Traditional polytypic *P. burkii* is a classic example of cryptic speciation since distinct species have almost similar morphology and plumage (Päckert *et al.*, 2004). This group of sister species experienced a complex evolutionary history, resulting in both extant vicariance, vertical and possibly horizontal parapatry, and local sympatry throughout the Sino-Himalayan range (Alstrom & Olsson 1999, Martens *et al.*, 1999) with seven currently recognized species (Päckert *et al.*, 2004; Martens 2010). The distribution of four of these species (Figure 1) clearly indicates overlapping species ranges, although few of the species are almost completely separated altitudinally (Alström, 2020a, b, c, d). Martens *et al.* (2002, 2011) have earlier reported the presence of three of these

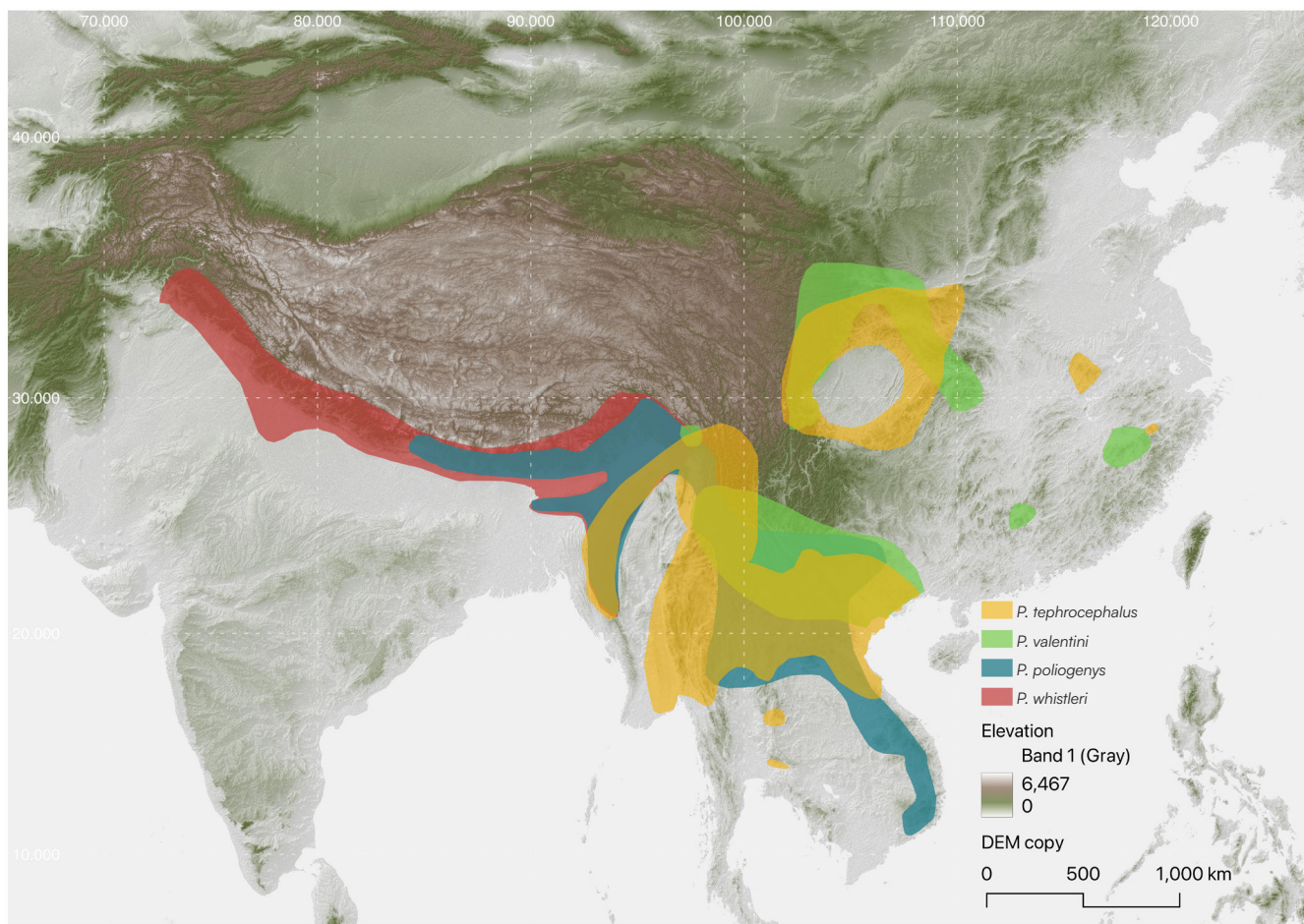


Figure 1. A distribution map of *P. tephrocephalus*, *P. valentini*, *P. poliogenys*, and *P. whistleri*, depicting their respective geographic ranges and areas of overlap. For understanding common zones where ranges converge, color-coded portions reflect each species' breeding grounds.

species along a single forested edge during the breeding season. Observation errors are most likely to occur in these regions where two or more of these species coexist.

In order to address these inaccuracies, image classification has evolved into an increasingly popular topic in deep learning and machine learning approaches (Alter & Wang, 2017). However, distinguishing between different bird species based on their distinctive shapes and appearances, considering background differences, controlling various lighting conditions in images, and adjusting to the dynamic postures of birds make it challenging to accurately identify bird species from images (Vo *et al.*, 2023). Our study used a part-based strategy to address this issue by splitting the identification process into two sections: part detection and identity classification. We provide an improved *You Only Look Once version 5* (further YOLOv5) based approach to part detection that can handle diverse ambient conditions and partial overlap between part objects.

The primary objective of this study was to determine whether the use of machine-learning and computer vision could make it easier to distinguish between members of the *P. burkii* complex based on morphological characteristics. Specific objectives of our investigation were to i) determine whether simple in-natural-setting photos could yield statistically accurate identification of visually similar co-occurring warblers of the *burkii* complex, and ii) record the degree to which any differences between groups show continuous or disjunct distributions in the image feature space; that might set up the foundation for developing a dependable and precise

method for automating the identification of at least some *P. burkii* species using morphological data. The results will facilitate quick and precise recognition of cryptic species, thereby reducing identification errors while monitoring.

Materials and Methods

Data Collection and Pre-processing

In addition to our own image library, we enhanced our dataset by collecting images for four species (*tephrocephalus*, *whistleri*, *valentini* from *P. burkii* complex and *poliogenys*-outgroup) from various online databases (iNaturalist, eBird, Wikimedia Commons) and creating respective sets of images for each species. These images are licensed under Creative Commons, allowing for their use in research and analysis while complying with copyright regulations. We specifically considered two important filtering criteria: a) subtle indicators like the supercilium, wing bars, bill and leg color, flanks, and most importantly the eye ring are crucial for identification in species with limited distinctive features, like *Phylloscopus* warblers. It was necessary to focus on side-profile photos to make sure that all of these features were readily apparent due to the complexity of these markers (this filter nearly cut the dataset in half). The images that passed the filter check were uploaded and annotated manually with the polygon tool (instances and label borders; Supplementary Figure S1-A) in *Supervisely* v1.0.23 software. A total of 289 photographs (71 personal observations and 218 sourced) were annotated, which consisted of *tephrocephalus*-

69, *whistleri*-73, *valentini*-59, and *poliogenys*-88, and later augmented to create 2890 photographs. In order to improve the model's robustness in identifying the target species under varying environmental conditions and its ability to generalize across diverse scenarios, we improved the dataset tenfold by flipping, applying slight rotations, adding random noise, and adjusting contrast and saturation in the images. We further analyzed the label size and label position to understand the ratio between the abscissa of the label center and the image width ('x'), as well as the ratio between the ordinate of the label center and the image height ('y') (Supplementary Figure S1-B). The correlogram of tagged samples and the bounding boxes are provided in Supplementary Figure S2. Images that were not included in the training validation dataset were used to generate the test dataset, which was further used to evaluate the model performance.

Dataset Splitting and Annotation

The 2,890 augmented images were split into training, validation, and test sets to ensure robust model training and unbiased evaluation. From these images, 2,312 images (80%) were allocated for training the YOLOv5 model, while 578 images (20%) were used for validation to monitor performance and fine-tune hyperparameters. Additionally, a separate dataset of 370 images, collected alongside the original dataset, was reserved exclusively for testing. This independent test set was not included in the training or validation phases, ensuring an unbiased evaluation of the model's capability.

All images in the training and validation sets were annotated using the polygon tool in Supervisely software to provide precise bounding boxes for the target species. The test dataset underwent the same annotation process, enabling accurate comparisons of predictions with ground truth during performance evaluation. Our dataset splitting strategy followed standard deep learning practice to ensure robust model training while accounting for independent assessment of the model's ability to generalize to unseen data. By using this structured approach, the model's performance was evaluated in a manner that closely simulates real-world scenarios.

The CNN Model (YOLO Architecture)

Convolutional neural networks are ideally suited for image classifications since they are composed of numerous convolution layers that record information from input photographs and offer a typical pattern for detecting the species evaluated after training the datasets (Szegedy *et al.*, 2015). Within CNNs, YOLO is a family of detection models in which the input image is divided into grids, and each cell in the grid detects the object. It incorporates test time augmentation, model ensemble, and the evolution of hyper-parameters. We used YOLOv5 (Redmon *et al.*, 2016) with the YOLOv5x6 variant (where x stands for "extra-large" and 6 for P6 layers in the model architecture), which performs exceptionally well and allows high-resolution images using larger models. The algorithm combines the feature extraction network to extract the feature information of the input bird image and then uses the aggregation network for detection to obtain the required semantic information (Liao & Tian, 2021). This architecture is also excellent at handling bounding boxes, which makes it ideal for managing multiple predictions using Intersection over Union (IoU) thresholding and non-max suppression, as well as for recognizing bird species that have bounding boxes. We used a Nvidia V100 GPU on an AWS EC2 instance, on which this model obtained a mean average precision (mAP) of 94 with a recall of 97.3 and a model size of 270MB.

The Amazon Web Services (AWS) Elastic Cloud Compute (EC2) Backend

The model was trained on a custom backend connected as

a compute cluster to Supervisely. We chose to proceed with this approach as the Supervisely platform already has the established service for Amazon Web Services (AWS)—Elastic Cloud Compute (EC2). Hence, all model training tasks were performed on the AWS backend. After the model was trained using annotated data to correctly identify the target species, it was validated with untrained annotated data to compare true and predicted results. Using cross-validation methods, hyperparameter tuning was carried out to improve model performance by optimizing learning rates, batch sizes, and network architecture. Real-time tracking of training progress and easy dataset management were made possible by the interface with Supervisely. Scalable resources were made available via AWS EC2, ensuring the effective processing of larger data sets. The model was further tuned for improved species identification accuracy after validation.

WORKFLOW PIPELINE

The workflow pipeline (Figure 2) outlines a methodical approach used in this machine-learning technique. First, the annotations are carefully examined and revised. This data was further used to train and evaluate the YOLOv5 machine-learning model. If the model converged, it was then tested on new footage. If the accuracy of the model was deemed adequate by the performance evaluation report, it was deployed; if not, the hyperparameters or dataset were adjusted, and the model was retrained to improve its detection ability. This iterative process guaranteed that the model was sturdy and dependable prior to its implementation in practical scenarios involving species identification.

Evaluation Metrics

Average confidence: It was calculated by adding the confidence values for true positives, true negatives, false positives, and false negatives, then dividing the result by the total number of instances.

Accuracy: It measured the frequency with which a model correctly predicts positive and negative outcomes across all possible scenarios.

$$\text{Accuracy} = (\text{True Positive} + \text{True Negative}) / (\text{True Positive} + \text{False Positive} + \text{True Negative} + \text{False Negative})$$

Precision (P) or True Positive Rate (TPR) or Sensitivity: An indicator of how frequently, a model makes a correct positive prediction out of all positive predictions.

$$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive})$$

Recall (R) or True Negative Rate (TNR) or Specificity: It is a measure that, when applied to the actual positive samples, indicates how frequently a model forecasts the accurate positive prediction based on the data's ground truth.

$$\text{Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative})$$

F1 Score: It computes the weighted average or harmonic mean of both Precision and Recall to merge them into a single value.

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

Intersection over Union (IoU): It metric that measures the overlap between the predicted bounding box and the ground truth bounding box to assess an object detector's accuracy for a given dataset.

$$\text{IoU} = (\text{Overlap area of Ground Truth \& Predicted bounding boxes}) / (\text{Total area of Ground Truth \& Predicted bounding boxes})$$

Generalized Intersection over Union (GIoU): It measures the distance between boxes and an enhancement over IoU that accounts for situations where there is no overlap by calculating the overlap between predicted and ground truth-bounding boxes.

$$GIoU = IoU - (\text{Area of minimum box enclosing both Ground truth \& Predicted bounding boxes}) / (\text{Area of minimum box enclosing both Ground truth \& Predicted bounding boxes})$$

Mean Average Precision (mAP): The trade-off between precision and recall is represented by the area under the precision-recall curve, which is the measure of average precision.

$$mAP = \frac{1}{n} \sum_{k=1}^n AP_k$$

In order to evaluate the model's performance in real-life scenarios, we generated a test set of frames using recently taken photos of the species under examination.

Results

Model training, validation, and testing

The model is most accurate at a confidence threshold of 0.75, suggesting that the performance of the model is good as the accuracy of the model holds up past the 0.5 confidence threshold mark (Figure 3A). Precision above 0.6 was estimated even at lower confidence thresholds of 0.2, suggesting that the model is mostly correct when it predicts a positive instance (Figure 3B). Through the precision and recall graph (Figure 3C), we can infer that the model has both good precision as well as a good recall. The model has a high recall value even at a high

precision value suggesting that the model is correct when it predicts a positive instance and is also good at identifying all the positive instances in the data. The model also has a good recall at a confidence threshold of 0.75 (Figure 3D), suggesting that the model is good at identifying the positive instances in the data and holds up the recall past the 0.5 confidence mark. The model performance confusion matrix for performing predictions and model training performance trends over 100 epochs of training are provided. The confusion matrix suggests that the model has less confusion by its capability to perform predictions close to the ground truth, thereby having a prominence of scores across the diagonal (Figure 4A). The model training performance trends suggest improving training curves across 100 epochs of training (Figure 4B). The GIoU values for the same are presented in Supplementary Table T1. Based on the confidence scores per detection, model performance on a few detection samples is provided in Figure 5.

Inference of test results

The model accuracy for the test dataset (n_train = 880, n_test=100) for *poliogenys* was 1, with a precision score of 1 and an average confidence of 0.86. Since *poliogenys* is the most distinct among the four species due to its characteristic fully grey head, this result can likely be attributed to this uniqueness. The model accuracy for the test dataset (n_train= 690, n_test=80) for *tephrocephalus* was 0.83, with a precision score of 0.6 and an average confidence of 0.62. The model was able to recognize 60 out of 80 test examples of *tephrocephalus* correctly, although there were a few cases where *valentini* was detected as *tephrocephalus* by this model. The model accuracy for the test dataset (n_train= 590, n_test=90) for *valentini* was 0.92, with a precision score of 1 and average confidence of 0.64. Due to a slight observable confusion that the model has between *valentini* and *tephrocephalus* due to the similarities between

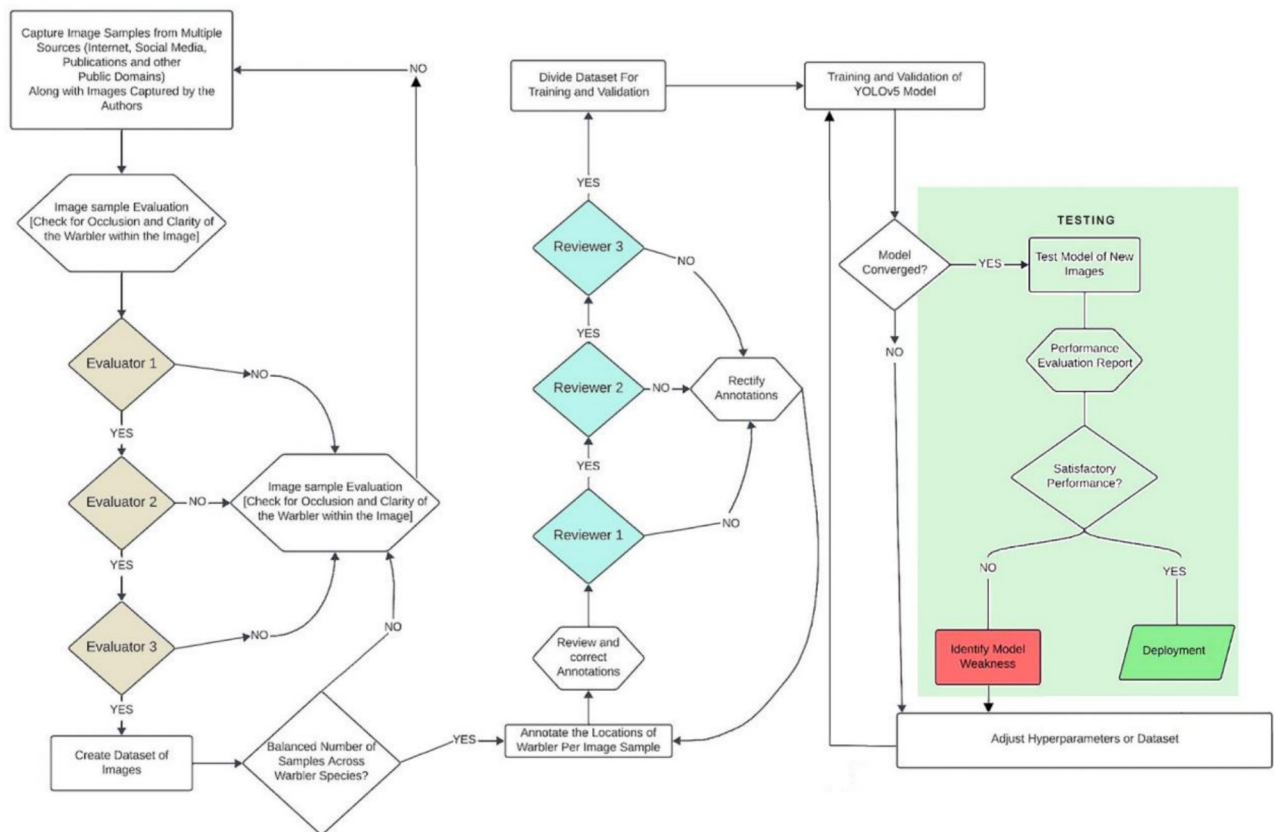


Figure 2. Workflow of image classification for annotations and subsequent model training and validation.

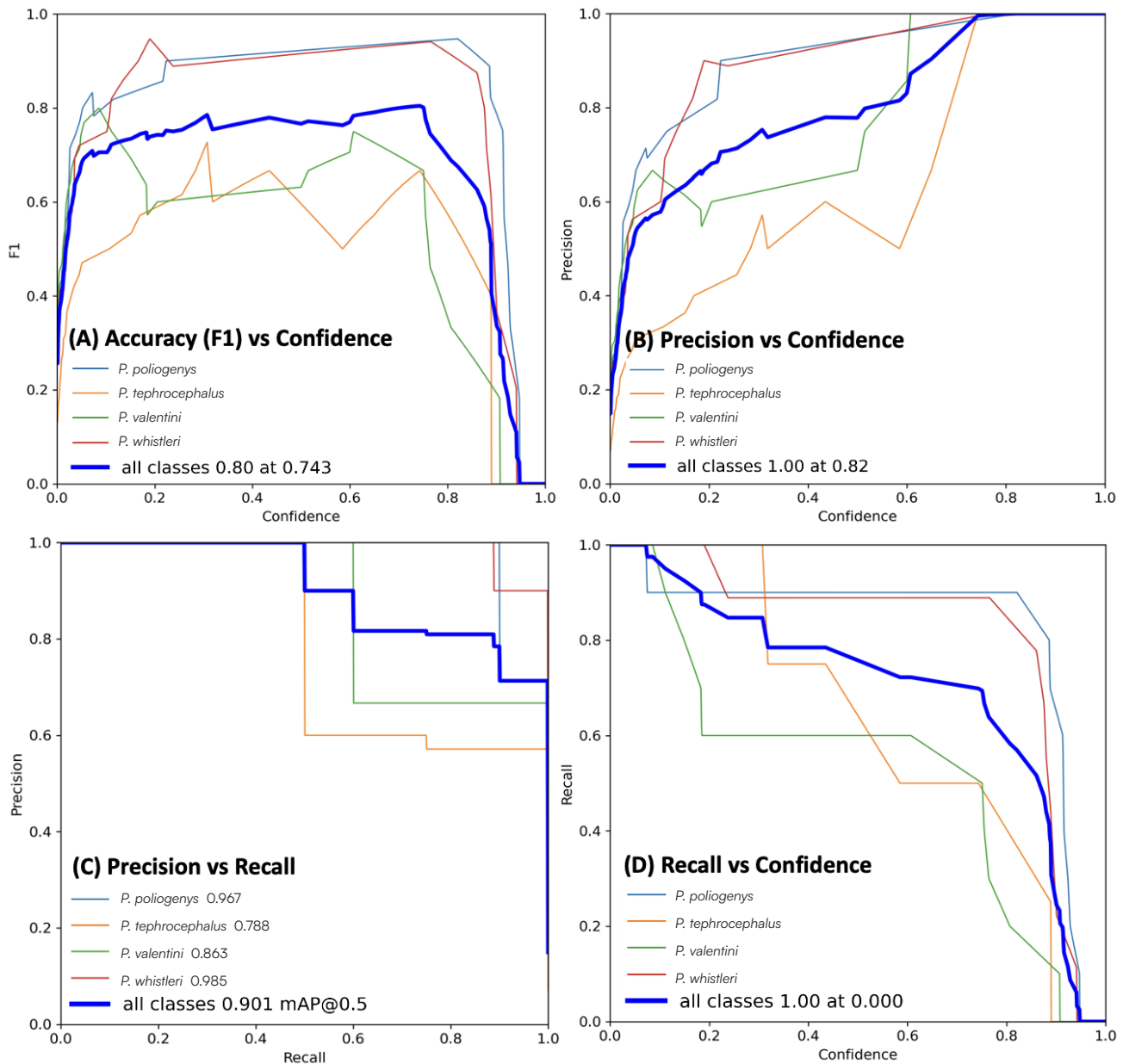


Figure 3. Based on the predictions that the model makes given an input image, for a set of test cases in the dataset, the following graphs are generated: (A) Graph showing the trade-off between the model's overall accuracy and confidence, (B) between the model's overall precision and confidence, (C) between the model's overall precision and recall, and (D) between the model's overall recall and confidence.

both species, the confidence scores for the valentini species have taken a major hit. Even though the model was able to detect valentini in 60 out of 90 test examples, the confidence scores were further reduced due to the model giving overlapping detections of tephrocephalus in 20 of these 60 correct detections, making only 40 purely correct detections for the valentini species. The model accuracy for the test dataset ($n_{\text{train}}=730$, $n_{\text{test}}=100$) for whistleri was 0.97, with a precision score of 1 and an average confidence of 0.77. The model showed no observable confusion for this species, asserting that all other confusion within the model can be overcome by giving more training examples for each species. Results of average confidence, model accuracy, precision, and average confidence are provided in Table 1 and Figure 6, and respective true negative, false negative, true positive, and false positive values for each species and calculations performed (for precision, recall, and F1 scores) are provided in the Supplementary Table T2.

Discussion

The rapid development of image recognition technology, especially YOLO algorithms (see Jiang *et al.*, 2022) has tremendously promoted the research on species detection and identification. While there are several other models like Deep Learning Ensembles (Chen *et al.*, 2019), Transfer Learning Models (Wang *et al.*, 2020), Support Vector Machines (Shalika & Seneviratne, 2016), and Generative Adversarial Networks (Zhang *et al.*, 2023) available for species recognition, the application of YOLO v5 models for studies on mammals (Tan *et al.*, 2022, Xie *et al.*, 2023), birds (Ou *et al.*, 2020, Liao & Tian, 2021, Yi *et al.*, 2023), reptiles (Pandey *et al.*, 2023, Afonso *et al.*, 2024), and amphibians (Takaya *et al.*, 2023) has shown significant advancements in terms of accuracy, speed, and model efficiency.

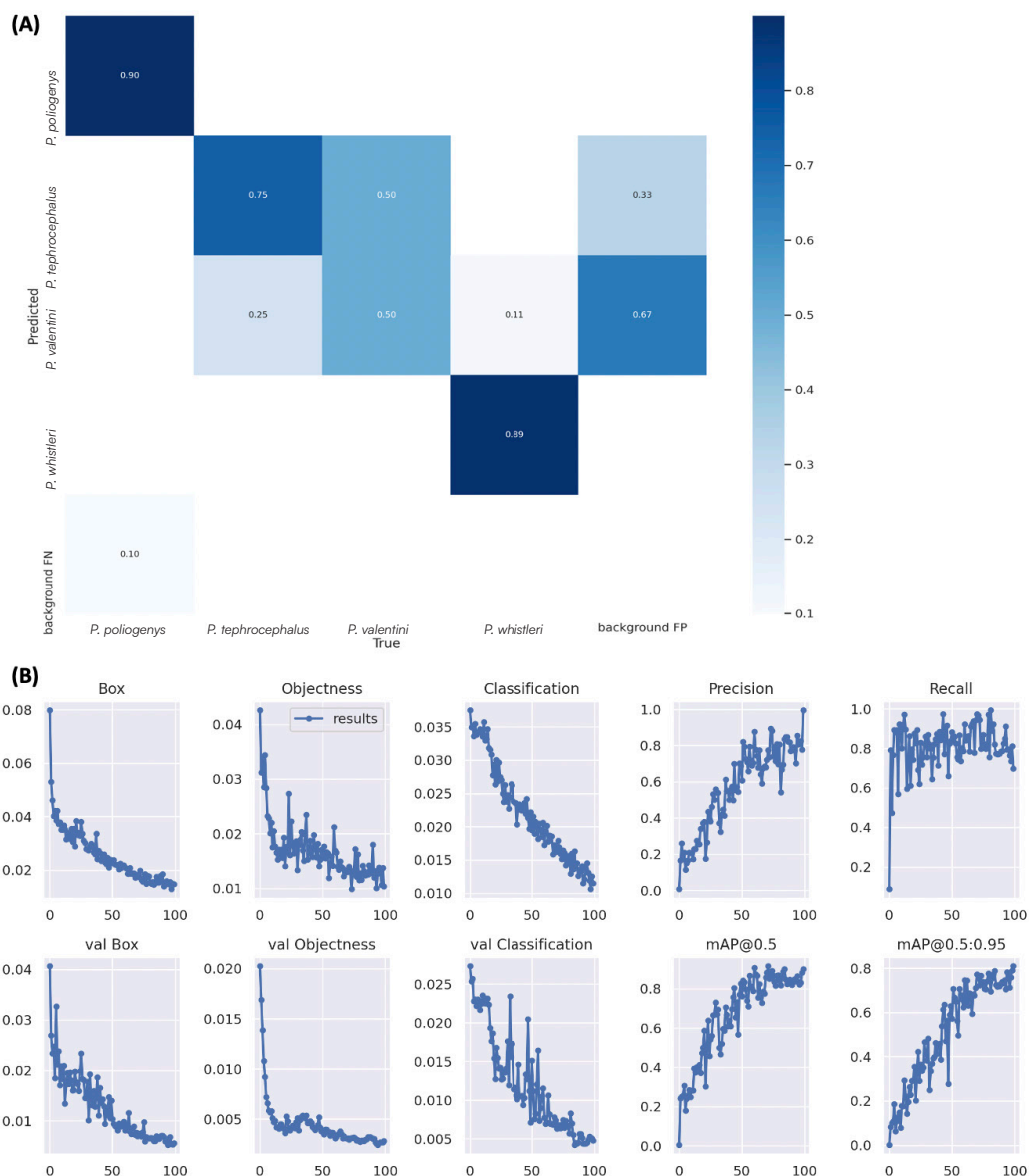


Figure 4. Comprehensive Evaluation of Object Detection Model Over 100 Training Epochs

(A) presents the confusion matrix, illustrating the classification accuracy of the detection model, where rows represent predicted species and columns show actual species. Correct predictions are displayed along the diagonal, while misclassifications are shown in off-diagonal elements.

(B) showcases various performance metrics, including loss metrics (Box, Objectness, Classification) and accuracy assessments (Precision, Recall, mAP@0.5, mAP@0.5:0.95) during training and validation phases.

However, cryptic species (*P. burkii* complex in our study) often possess minute variations in patterns that can be challenging for standard models. Thus, additional validation (see Ma & Yang, 2022) and specialized customization, such as enhanced data augmentation (Bati & Ser, 2023) and fine-tuning (Kim *et al.*, 2022) of model parameters, were necessary to improve detection accuracy and handle these subtle morphological differences in our study. Therefore, a comprehensive approach to preprocessing and model training was used to enhance species classification, which, in real-world scenarios, is both robust and adaptable. To maintain a clear and targeted research trajectory within this group of birds, we recommend that species belonging to different complexes be examined independently rather than merging them. This study addresses the challenges by proposing a deep learning method to improve the identification of bird species of the *burkii* complex, based on the YOLO-v5 algorithm.

In contrast to other CNN architectures like ResNet, YOLOv5 was selected for classification in this study as it can effectively handle the unique difficulties of our task while providing notable benefits in terms of flexibility, performance, and deployment. It integrates categorization and object detection into a single, effective pipeline. YOLOv5 does not require a separate detection step, which is especially helpful for applications requiring localized species identification, in contrast to ResNet, DenseNet, or other CNNs that are solely focused on classification. The distinctiveness of the species' morphology in the dataset may determine how well visual cues alone may distinguish species differences. The accuracy of image-based categorization may be constrained when species have notable overlap in appearance or modest visual differences. Utilizing complementary information, and combining visual and auditory signals may increase categorization accuracy. However, the primary reason for excluding acoustic classification was that it was not be feasible

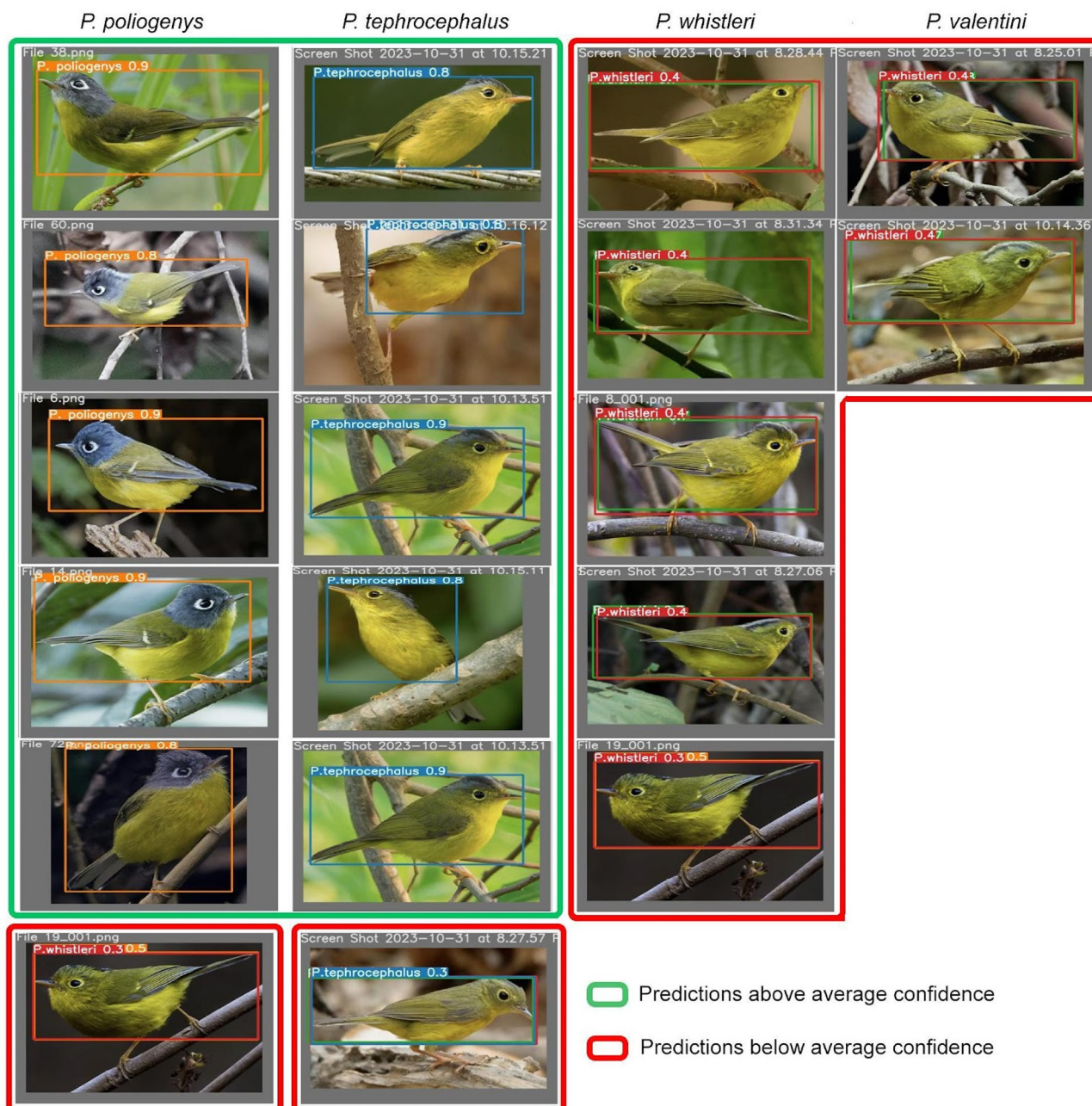


Figure 5. Sample set of model detections over the test dataset, categorized based on confidence scores for each detection. Links to the photographs featured in this figure, along with the names of the respective photographers, are provided in the supplementary section.

Table 1. The average performance metrics, obtained from the confusion matrix findings over multiple runs, for test datasets, including accuracy, precision, recall, and F1 score.

Species	Avg. Confidence	Accuracy	Precision	Recall	F1 Score
<i>P. poliogenys</i> (ougroup)	0.86	1	1	1	1
<i>P. tephrocephalus</i>	0.59	0.83	0.6	0.75	0.67
<i>P. valentini</i>	0.53	0.92	1	0.67	0.8
<i>P. whistleri</i>	0.76	0.97	1	0.9	0.95

to gather high-quality acoustic data for these species within the limitations of our analysis. It is more difficult to get reliable and consistent audio recordings of this species since the calls often consist of short pulses, and species identification primarily relies on songs.

The results of this study indicate that statistically valid morphology-based identifications of cryptic species can be obtained from basic images taken in natural settings. The YOLOv5 model's strong recall, accuracy, and precision rates reveal how machine learning algorithms use the unique morphological features available in these images even if they have subtle variations. The model demonstrated strong performance with high accuracy, precision, and recall at the 0.50 confidence threshold and maintained precision even at lower thresholds, which makes it ideal for such tasks. But in order to make it adaptive, we are currently focusing on increasing the training dataset as the model slightly confuses *valentini* (average confidence = 0.64). The model operates effectively

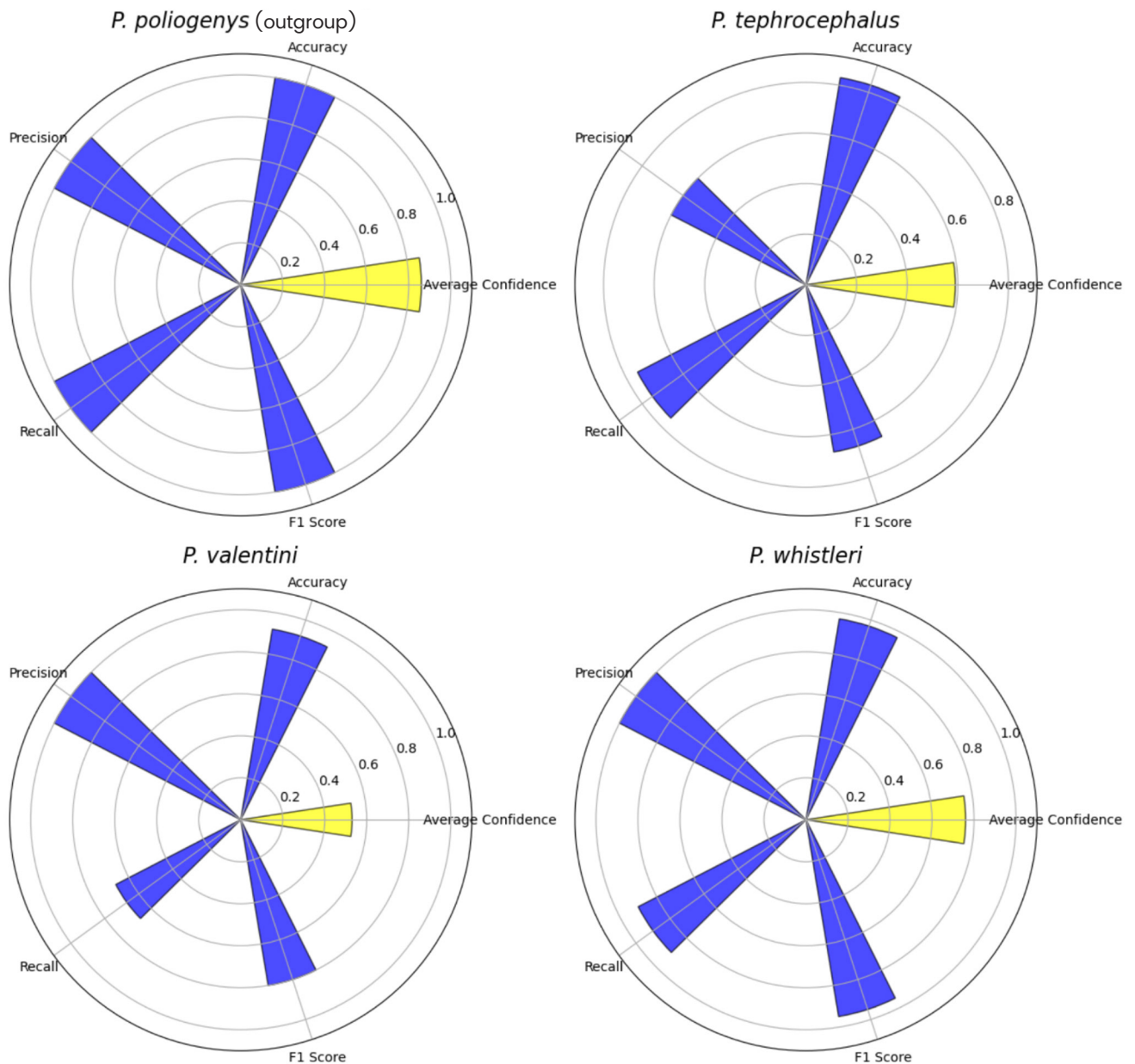


Figure 6. Accuracy, Precision, Recall, F1 Score, and Average Confidence distributions for different species are shown in a heat plot, emphasizing the variation in performance across metrics.

with *tephrocephalus* and has excellent detection and differentiation ability between *poliogenys* and *whistleri* (average confidence = 0.086 and 0.77, respectively). The test dataset presents difficulties for tackling *valentini*, though, often leading to the simultaneous detection of *tephrocephalus* and *valentini*. Resolving these issues will possibly improve the model and guarantee accurate species identification in ecological research (Shah *et al.*, 2023).

Even though the image only included *P. valentini*, there have been instances where multiple bounding box detections occurred in the same image, generating bounding boxes for both *P. valentini* and *P. tephrocephalus*. This might have been influenced by the low imagecount of both *P. valentini* and *P. tephrocephalus* in the current dataset, as well as similarity in several features between the two species. To address this, we recommend (a) training the dataset by doubling epochs and (b) extending the dataset with more diverse and representative examples of both *P. valentini* and *P. tephrocephalus*. One way to reduce the model's confusion in differentiating between species

like *valentini* and *tephrocephalus* is to train it independently depending on the geographic distributions of each species. Identification can be made much simpler by concentrating on the unique elevational gradients and ecological requirements (see Yang *et al.*, 2023, Parab *et al.*, 2023) preferred by each species (Alstrom, 2020). By using a training technique tailored to geographic distribution, the model can acquire the distinct traits linked to the preferred settings of individual species (see Opaev & Kolesnikova, 2019), resulting in increased recognition accuracy. Enhancing species identification in regions where overlap exists requires incorporating geographic distribution data into the training process (Alström, 2020a, b, c, d).

Although other applications like Merlin offer users a list of potential species, mostly aimed at skilled birders who use their knowledge to make the ultimate decision, our method starts the process of developing a more reliable model for the few focal species that are being considered. Our goal was to create a model that could make a single, solid prediction, allowing even inexperienced birders to progressively come to trust the

model's accuracy and gain more knowledge by using it. We foresee adding a feedback mechanism that displays confidence scores next to predictions as a possible future improvement. This would provide an instructional component currently lacking in other applications and allow users to assess the model's forecasting precision, thereby fostering confidence.

In conclusion, the percentage of incorrect identifications on citizen science platforms can be considerably decreased by using specific models created for species identification. We may enhance the quality of the data gathered by citizen scientists by improving the precision of these models. This development is critical to successful conservation efforts because accurate species identification directly affects the quality of data (Sharma *et al.*, 2019) and our understanding of the ecosystem. The application of such models will ultimately result in improved ecological research outputs and more informed decision-making regarding biodiversity conservation.

Limitations and future work

The test dataset for this study did not specifically include non-target species, such as co-occurring *Phylloscopus* warbler species and other co-occurring non-warbler species. In order to assess YOLOv5's performance under controlled circumstances, the main goal was to concentrate on target species identification. To preserve clarity and consistency in performance evaluation, we opted to focus only on target species for the present study; however, the inclusion of non-target species would have provided more insights into the framework's generality. We acknowledge that adding non-target species would be useful for evaluating the resilience of the model in practical applications, and suggest that this be the primary target for future research.

TO DOWNLOAD SUPPLEMENTARY MATERIAL CLICK [HERE](#).

Acknowledgement

We acknowledge the help and support provided by the Director, Wildlife Institute of India and the Director, Academy of Scientific and Innovative Research, for carrying out this study. No fieldwork permissions were required for this study.

DECLARATION OF USE OF GENERATIVE AI

The authors declare that they have used generative artificial intelligence (Claude AI v3) for rephrasing and in the writing of descriptions of figures and tables. The authors thoroughly reviewed and edited the content generated by AI tool and take full responsibility for the content of the publication.

CONFLICT OF INTEREST

The author declares no conflict of interest.

DATA AVAILABILITY

The code used for the analyses in this study is available at the following GitHub repository: https://github.com/Wild-Inst-Ind/yolov5_Warbler.

AUTHORS' CONTRIBUTION

T.P: conceptualization, investigation, methodology, writing original draft, annotations & formal data analysis, software, visualization, formatting, writing – review & editing. G.B.H: conceptualization, investigation, methodology, annotations and formal data analysis, software, writing – review and editing, validation.

M.B and S.N: Annotations, Data Augmentation, Image Sorting and preprocessing.

A.C: Annotations.

S.R: conceptualization, investigation, methodology, annotations and formal data analysis, software, visualization, formatting, writing – review and editing, validation.

References

- Afonso, A. L., Lopes, G. & Ribeiro, A. F. (2024). Lizard Body Temperature Acquisition and Lizard Recognition Using Artificial Intelligence. *Sensors*, 24(13), 4135. <https://doi.org/10.3390/s24134135>
- Alström, P. (2020a). Bianchi's Warbler (*Phylloscopus valentini*), version 1.0. In: J. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E. (eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi.org/10.2173/bow.biawar1.01>
- Alström, P. (2020b). Gray-cheeked Warbler (*Phylloscopus poliogenys*), version 1.0. In: J. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E. (eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi.org/10.2173/bow.gycwar2.01>
- Alström, P. (2020c). Gray-crowned Warbler (*Phylloscopus tephrocephalus*), version 1.0. In: J. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E. (eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi.org/10.2173/bow.gycwar1.01>
- Alström, P. (2020d). Whistler's Warbler (*Phylloscopus whistleri*), version 1.0. In: J. del Hoyo, J., Elliott, A., Sargatal, J., Christie, D. A. & de Juana, E. (eds.), *Birds of the World*. Cornell Lab of Ornithology, Ithaca, NY, USA. <https://doi.org/10.2173/bow.whiwar2.01>
- Alström, P. & Olsson, U. (1999). The golden-spectacled warbler: a complex of sibling species, including a previously undescribed species. *Ibis*, 141(4), 545-568. <https://doi.org/10.1111/j.1474-919X.1999.tb07363.x>
- Alter, A. L. & Wang, K. M. (2017). An exploration of computer vision techniques for bird species classification. <https://cs229.stanford.edu/proj2017/final-reports/5161697.pdf>
- Bati, C. T. & Ser, G. (2023). Effects of Data Augmentation Methods on YOLO v5s: Application of Deep Learning with Pytorch for Individual Cattle Identification. *Yuzuncu Yil University Journal of Agricultural Sciences*, 33(3), 363-376. <https://doi.org/10.29133/yyutbd.1246901>
- Chen, Y., Wang, Y., Gu, Y., He, X., Ghamisi, P. & Jia, X. (2019). Deep learning ensemble for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(6), 1882-1897. <https://doi.org/10.1109/JSTARS.2019.2915259>
- de Freitas, E. L., Campagna, L., Butcher, B., Lovette, I. & Caparroz, R. (2022). Ecological traits drive genetic structuring in two open-habitat birds from the morphologically cryptic genus *Elaenia* (Aves: Tyrannidae). *Journal of Avian Biology*, 2022(4), e02931. <https://doi.org/10.1111/jav.02931>
- Jiang, P., Ergu, D., Liu, F., Cai, Y. & Ma, B. (2022). A Review of Yolo algorithm developments. *Procedia computer science*, 199, 1066-1073. <https://doi.org/10.1016/j.procs.2022.01.135>
- Kelling, S., Johnston, A., Fink, D., Ruiz-Gutierrez, V., Bonney, R., Bonn, A. & Guralnick, R. (2018). Finding the signal in the noise of Citizen Science Observations. *bioRxiv*, 326314. <https://doi.org/10.1101/326314>
- Kim, J. H., Kim, N., Park, Y. W. & Won, C. S. (2022). Object detection and classification based on YOLO-V5 with improved maritime dataset. *Journal of Marine Science and Engineering*, 10(3), 377. <https://doi.org/10.3390/jmse10030377>
- Koskimies, P. (1989). Birds as a tool in environmental monitoring. In: *Annales Zoologici Fennici*. Finnish Zoological Publishing Board, formed by the Finnish Academy of Sciences, Societas Scientiarum Fennica, Societas pro Fauna et Flora Fennica and Societas Biologica Fennica Vanamo. pp.153-166.
- Larsen, F. W., Bladt, J., Balmford, A. & Rahbek, C. (2012). Birds as biodiversity surrogates: will supplementing birds with other taxa improve effectiveness? *Journal of Applied Ecology*, 49(2), 349-356. <https://doi.org/10.1111/j.1365-2664.2011.02094.x>

- Liao, Z. & Tian, M. (2021, October). A bird species detection method based on YOLO-v5. In: *Proceedings SPIE 11933, 2021 International Conference on Neural Networks, Information and Communication Engineering*, 119330C. 15 October 2021, Qingdao, China. pp.65-75. <https://doi.org/10.1117/12.2615310>
- Lindström, Å., Green, M., Paulson, G., Smith, H. G. & Devictor, V. (2013). Rapid changes in bird community composition at multiple temporal and spatial scales in response to recent climate change. *Ecography*, 36(3), 313-322. <https://doi.org/10.1111/j.1600-0587.2012.07799.x>
- Ma, D. & Yang, J. (2022). Yolo-animal: An efficient wildlife detection network based on improved yolov5. In: *2022 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*. Xi'an, China. pp.464-468. <https://doi.org/10.1109/ICICML57342.2022.10009855>
- Martens, J. (2010). A preliminary review of the leaf warbler genera *Phylloscopus* and *Seicercus*. Systematic notes on Asian birds: 72. *Bulletin of the British Ornithologists' Club*, 5, 41-116.
- Martens, J., Eck, S., Päckert, M. & Sun, Y. H. (1999). The Golden-spectacled Warbler *Seicercus burkii*-a species swarm (Aves: Passeriformes: Sylviidae). Part 1. *Zoologische Abhandlungen-Statistisches Museum Fur Tierkunde In Dresden*, 50, 281-327.
- Martens, J., Eck, S., Päckert, M. & Sun, Y. H. (2002). Methods of systematic and taxonomic research on passerine birds: the timely example of the *Seicercus burkii* complex (Sylviidae). *Bonner Zoologische Beiträge*, 51(2/3), 109-118.
- Martens, J., Tietze, D. T. & Päckert, M. (2011). Phylogeny, biodiversity, and species limits of passerine birds in the Sino-Himalayan region—a critical review. *Ornithological Monographs*, 70(1), 64-94. <https://doi.org/10.1525/om.2011.70.1.64>
- Opaev, A. & Kolesnikova, Y. (2019). Lack of habitat segregation and no interspecific territoriality in three syntopic cryptic species of the golden-spectacled warblers *Phylloscopus* (*Seicercus*) *burkii* complex. *Journal of Avian Biology*, 50, e02307. <https://doi.org/10.1111/jav.02307>
- Ou, Y. Q., Lin, C. H., Huang, T. C. & Tsai, M. F. (2020). Machine learning-based object recognition technology for bird identification system. In: *2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan)*. Taoyuan, Taiwan. pp.1-2. <https://doi.org/10.1109/ICCE-Taiwan49838.2020.9258061>
- Päckert, M., Martens, J., Sun, Y. H. & Veith, M. (2004). The radiation of the *Seicercus burkii* complex and its congeners (Aves: Sylviidae): molecular genetics and bioacoustics. *Organisms Diversity & Evolution*, 4(4), 341-364. <https://doi.org/10.1016/j.ode.2004.06.002>
- Pandey, S. K., Kumar, A., Yadav, D. P., Sinha, A., Hassan, M. M., Singh, N. K. & Garg, N. (2023). Health evaluation and dangerous reptile detection using a novel framework powered by the YOLO algorithm to design high-content cellular imaging systems. *The Journal of Engineering*, 2023(12), e12335. <https://doi.org/10.1049/tje2.12335>
- Parab, T., De, K., Singh, A. P. & Uniyal, V. P. (2023). Effects of weather on behavioural responses of two warbler (*Phylloscopus*) species in the Great Himalayan National Park Conservation Area. *Ornithology Research*, 31(2), 111-118. <https://doi.org/10.1007/s43388-023-00121-9>
- Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. pp.779-788. <https://doi.org/10.1109/CVPR.2016.91>
- Shah, C., Alaba, S. Y., Nabi, M. M., Prior, J., Campbell, M., Wallace, F., Ball, J. E. & Moorhead, R. (2023). An enhanced YOLOv5 model for fish species recognition from underwater environments. In: *Proceedings SPIE 12543, Ocean Sensing and Monitoring XV*, 1254300. 12 June 2023, Orlando, Florida, United States. <https://doi.org/10.1117/12.2663408>
- Shalika, A. U. & Seneviratne, L. (2016). Animal classification system based on image processing & support vector machine. *Journal of Computer and Communications*, 4(1), 12-21. <https://doi.org/10.4236/jcc.2016.41002>
- Sharma, N., Colucci-Gray, L., Siddharthan, A., Comont, R. & Van der Wal, R. (2019). Designing online species identification tools for biological recording: the impact on data quality and citizen science learning. *PeerJ*, 6, e5965. <https://doi.org/10.7717/peerj.5965>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. In: *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR2015)*. 7-12 June, Boston. pp. 1-9.
- Takaya, K., Taguchi, Y. & Ise, T. (2023). Individual identification of endangered amphibians using deep learning and smartphone images: case study of the Japanese giant salamander (*Andrias japonicus*). *Scientific Reports*, 13(1), 16212. <https://doi.org/10.1038/s41598-023-40814-1>
- Tan, M., Chao, W., Cheng, J. K., Zhou, M., Ma, Y., Jiang, X., Ge, J., Yu, L. & Feng, L. (2022). Animal detection and classification from camera trap images using different mainstream object detection architectures. *Animals*, 12(15), 1976. <https://doi.org/10.3390/ani12151976>
- Vo, H. T., Thien, N. N. & Mui, K. C. (2023). Bird detection and species classification: using YOLOv5 and deep transfer learning models. *International Journal of Advanced Computer Science and Applications*, 14(7). <https://doi.org/10.14569/IJACSA.2023.01407102>
- Wang, X., Li, P. & Zhu, C. (2020). Classification of wildlife based on transfer learning. In: *Proceedings of the 2020 4th International Conference on Video and Image Processing*. Xi'an, China. pp.236-240. <https://doi.org/10.1145/3447450.3447487>
- Xie, Y., Jiang, J., Bao, H., Zhai, P., Zhao, Y., Zhou, X. & Jiang, G. (2023). Recognition of big mammal species in airborne thermal imaging based on YOLO V5 algorithm. *Integrative Zoology*, 18(2), 333-352. <https://doi.org/10.1111/1749-4877.12667>
- Yang, W., Liu, T., Jiang, P., Qi, A., Deng, L., Liu, Z. & He, Y. (2023). A forest wildlife detection algorithm based on improved YOLOv5s. *Animals*, 13(19), 3134. <https://doi.org/10.3390/ani13193134>
- Yi, X., Qian, C., Wu, P., Maponde, B. T., Jiang, T. & Ge, W. (2023). Research on fine-grained image recognition of birds based on improved YOLOv5. *Sensors*, 23(19), 8204. <https://doi.org/10.3390/s23198204>
- Zhang, Q., Yi, X., Guo, J., Tang, Y., Feng, T. & Liu, R. (2023). A few-shot rare wildlife image classification method based on style migration data augmentation. *Ecological Informatics*, 77, 102237. <https://doi.org/10.1016/j.ecoinf.2023.102237>