CLLMM RESEARCH CENTRE | GOYDER INSTITUTE FOR WATER RESEARCH

Shorebird body condition monitoring through machine learning, targeted field surveys and community science in the CLLMM region

April 2026

Steven Delean, Ruth Cope, Darcy Whittaker and Thomas Prowse

The CLLMM Research Centre has the objective of developing and delivering locally driven, innovative, and impactful research that reflects community and First Nations priorities to provide an evidence base to address critical needs of the CLLMM region. The CLLMM Research Centre will also bring together our First Nations, local community and scientists to create and share knowledge, and support our well-informed and engaged community and empower our future generations to be part of the solution.

The Goyder Institute for Water Research Coorong, Lower Lakes and Murray Mouth (CLLMM) Research Centre, based in Goolwa SA, was established in 2023 through funding by the Australian Government.

Australian Government

Department of Climate Change, Energy, the Environment and Water

The Goyder Institute for Water Research is a partnership between the South Australian Government through the Department for Environment and Water, CSIRO, Flinders University, and Adelaide University.

Government of South Australia
Department for Environment and Water

CSIRO

Flinders University

Adelaide University

Title page photo credit: 'Red-necked stint feeding' – Darcy Whittaker (2025)

Enquires should be addressed to: CLLMM Research Centre
Level 1, 92 Barrage Road
Goolwa, SA 5214
e-mail: info@cllmmresearchcentre.org

This report may be cited as:

Delean S, Cope R, Whittaker D, and Prowse T (2026). *Shorebird body condition monitoring through machine learning, targeted field surveys and community science in the CLLMM region*. A report prepared by Adelaide University for the Goyder Institute for Water Research, CLLMM Research Centre, Goolwa. https://doi.org/10.25909/33432661.

Disclaimer

Table of Contents

1 Executive Summary

This project investigates the feasibility of using community-sourced photographs combined with automated machine learning to assess the body condition of migratory shorebirds in the Coorong, Lower Lakes, and Murray Mouth (CLLMM) region of South Australia.

Body condition was assessed using the Abdominal Profile Index (API), scored on an ordinal scale of 1 (lean) to 5 (plump), with an additional class 6 for "fluffed" (non-assessable) birds. Body condition API was scored on a dataset of 697 images of Sharp-tailed Sandpiper (*Calidris acuminata*), Curlew Sandpiper (*Calidris ferruginea*), Red-necked Stint (*Calidris ruficollis*) and Common Greenshank (*Tringa nebularia*) collected prior to this study. A state-of-the-art, three-phase machine learning pipeline, trained by the expert-elicited body condition assessments, was developed to automate body condition scoring from photographs. The new pipeline represents a substantial methodological advance over previous approaches, offering greater accuracy, transparency, generalisability, and opportunity for scientific audit.

Overall model accuracy was high, with the lowest for Class 1 (lean) birds, attributable to underrepresentation of this class in the training dataset, and confidence scores were consistently high for correctly classified images (median 0.5–0.8 for classes 2–5). The field survey and citizen science project encompasses two migratory seasons (2024–2025 and 2025–2026) and targets species including Sharp-tailed Sandpiper, Red-necked Stint, Curlew Sandpiper, and Common Greenshank. A community science programme was established, distributing educational materials to regional conservation groups and government environmental organisations across South Australia. Community submissions via BirdLife Australia's "Birdata" portal were modest. Data gaps were supplemented by 1,947 standardised images across 19 targeted surveys between September 2025 and March 2026, contributed by a professional nature photographer. Applying the trained model to 1,947 unlabelled images revealed meaningful spatiotemporal variation in body condition across species and regions of the CLLMM. The training dataset was class-imbalanced, particularly underrepresenting lean (Class 1) individuals, which contributes to boundary classifier uncertainty and potential central-tendency bias, limiting the ability to score shorebirds as lean. Training labels are currently standardised to a single expert, and we recommend expanding expert assessments to quantify inter-rater reliability.

Priority recommendations include expanding the expert-labelled image database across all body condition classes and species, computing inter-rater agreement metrics, and evaluating fluffed-bird classifier performance.

2 Acknowledgements

We acknowledge the cultural connections to the land and the connected waters, as well as the cultural significance of waterbirds that exist among the people of the Ngarrindjeri and Boandik Nations across the Coorong, Lower Lakes, and Murray Mouth region. We recognise that First Nations peoples' spiritual, social, cultural, and economic practices come from their lands and waters, and they continue to maintain their cultural heritage, economies, languages, and laws, which are of ongoing importance. We are committed to fostering and nurturing ongoing relationships with the First Nations community in the CLLMM region, as we continue to work together towards a diverse and thriving waterbird community.

This project was funded by the Goyder Institute for Water Research's CLLMM Research Centre, which is funded by the Australian Government. We thank the CLLMM Research Centre staff for their support in initiating, managing, and supporting the project. We thank the Goyder Institute for Water Research, the Research Advisory Committee, and two anonymous reviewers for their constructive feedback on the report.

We thank BirdLife Australia, the CLLMM Research Centre, Friends of Shorebirds South-East, SA Shorebirds Foundation, Friends of Adelaide International Bird Sanctuary, Tolderol Working Group, Nature Glenelg Trust, Limestone Coast Landscape Board, and Lower Limestone Coast DEW National Park Rangers for their support in distributing educational materials to community members. We particularly thank Joey Rumbold, for community member feedback during the early stages of material development. We also thank Joris Driessen, Senior Analyst at BirdLife Australia, for facilitating the inclusion of an image link in Birdata and for extracting Birdata submissions for this project.

3 Introduction

The physical condition of birds is a direct reflection of whether they are obtaining sufficient resources from the local environment to maintain good fitness, which, in turn, influences survival and breeding success. The nutritional and energetic state of migratory birds, for example, is reflected in their fat reserves and muscle composition, which can be observed by their body condition, and is a paramount determinant of individual fitness, reproductive success, and overall survival (Landys et al. 2005).

Body condition has been measured across a wide range of species to assess overall fitness (Lacy et al. 2024; Bierlich et al. 2024). The primary fat storage location for waterbirds is subcutaneously along the abdomen. Therefore, it is possible to assess shorebird body condition using an abdominal profile index (API), and other studies have demonstrated a clear positive relationship between APIs scored in the field and i) measured body mass of captured birds (Féret et al. 2005; Wiersma and Piersma 1995) and ii) the resultant fitness outcomes of birds (Swift et al. 2020), both of which reflect body condition. For shorebirds, a score is generally assigned between one (very lean) and five (very fat) (Wiersma and Piersma 1995; Figure 1).

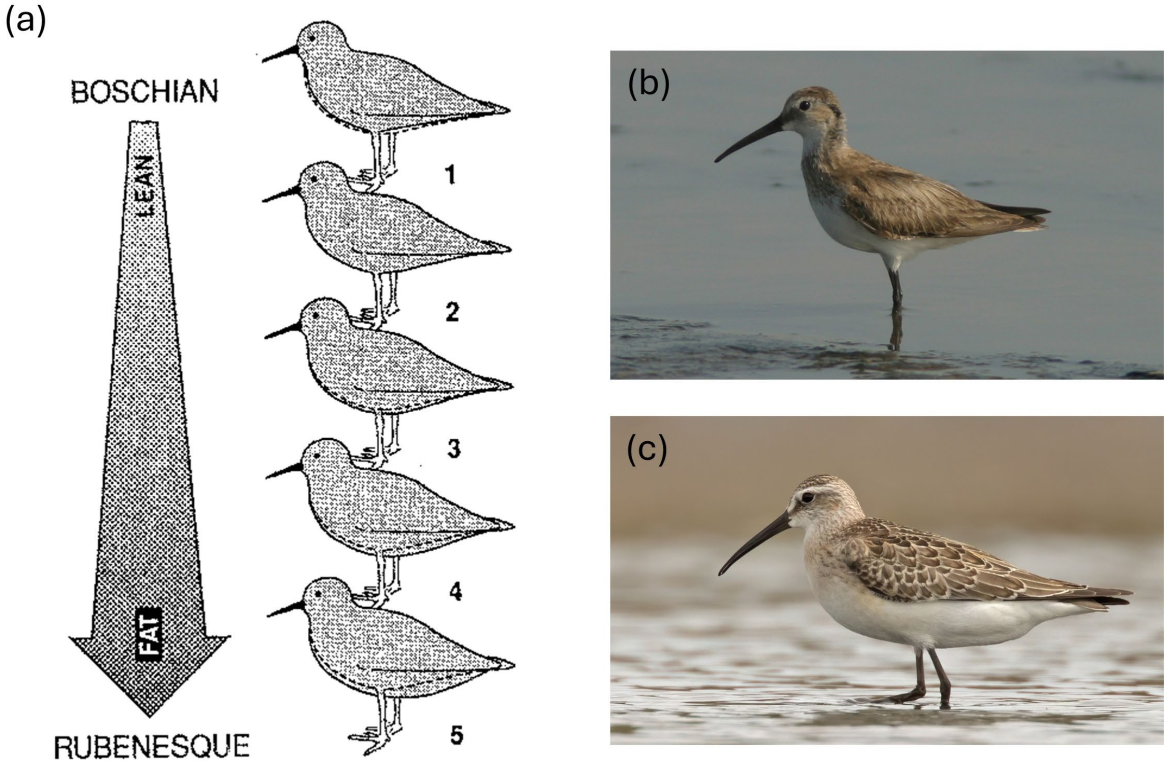


Figure 1. (a) Abdominal profile index for scoring shorebird body condition (based on Red Knots *Calidris canutus*). Reproduced from Figure 1 in Wiersma and Piersma (1995). Dashed line in birds 1, 3, 4, and 5 shows an abdominal profile for a bird with a score of 2. (b) Example image of a Curlew Sandpiper (*Calidris ferruginea*) from the Coorong assigned a score of 1 (c) Example image of a Curlew Sandpiper from the Coorong assigned a score of 5.

3.1 Applications of photography in measuring body condition

Monitoring the morphometrics and energetic reserves of bird populations can potentially provide insights into ecosystem health, so there is strong interest in methods that can measure condition at scales relevant to understanding these patterns. Historically, evaluating body condition has involved labour-intensive and invasive methods that require animal capture. However, contemporary research has focused on non-invasive optical techniques, such as photogrammetry, to extract biological measures from digital images. Osborne and Ryan (2021) used manual screening of digital photographs to moult-score albatross, which were also scored in the hand, demonstrating the effectiveness of using photographs whilst minimising capture stress on birds in the field. Photographs have also been used to moult score free-flying jaegers and skuas (von Bemmelen et al. 2018). Image classifiers have also been developed to count birds in their natural habitat from drone imagery (Hodgson et al. 2018). Hayes et al. (2021) have used machine learning methods to identify and count seabirds in drone imagery with high accuracy. Similarly, drone imagery can be used to estimate 2D and 3D body volumes of Australian sea lions, thereby accurately assessing body condition (Hodgson et al. 2020).

This shift toward the use of digital images has been accelerated by the proliferation of citizen science programmes, which harness the collective efforts of the public to gather opportunistic wildlife photographs on large scales (Tuia et al. 2022). Platforms such as iNaturalist have facilitated the rapid generation of extensive, crowdsourced ecological datasets, providing wide spatial and temporal resolution that circumvents the logistical constraints of traditional field surveys. Along with the development and deployment of cheap hardware for remotely capturing ecological data, these approaches have led to a proliferation of ecological data that poses significant challenges for timely processing and analysis (Blair et al. 2024).

As a consequence, research is increasingly focused on machine learning and computer vision methods to extract ecological and phenotypic information from vast data repositories (Blair et al. 2024; Lürig et al. 2021). Machine learning facilitates fine-grained taxonomic analyses by automatically localising specific, distinguishing morphological traits, such as plumage patterns, in visually similar bird species, enabling highly accurate, automated classification (Chowdhury et al. 2025). Furthermore, a recent study used vision transformers to predict the chronological age of wild primates solely from facial photographs, providing a non-invasive method to track development (Renoult et al. 2025).

Integrating crowdsourced citizen science imagery with advanced computer vision pipelines will enable ecologists to rapidly assess the body condition of bird populations across large geographic areas and facilitate temporal monitoring of populations and habitats. Ultimately, the automated inference of avian body condition from digital imagery can significantly advance evidence-based conservation and wildlife management of shorebirds.

3.2 Aims

This project aims to work with the community to create a spatiotemporal database of shorebird images, with which to:

1) Continue to improve an automated condition scoring system for different shorebird species.
2) Compare body condition trajectories over the non-breeding period for migratory shorebirds (e.g. sharp-tailed sandpiper, red-necked stint) in different wetlands.
3) Develop a new method for quantifying the body condition of migratory shorebirds to inform discussions regarding the management of this threatened functional group.

4 Methods

4.1 Community outreach and other data sources

Over two migratory bird seasons (2024–2025 and 2025–2026), images of shorebird body condition were taken by community members in the Coorong, Lower Lakes, and Murray Mouth region and surrounding South-East wetlands in South Australia, Australia. We deemed this community science initiative, 'Shorebird Snapshots'. To share the initiative with community members, a pamphlet was developed that educated the public on the project's purpose, how to take body condition photos, how to identify migratory species, and how to upload the image to Birdata (Figure 2). In October 2024, the educational material was distributed to a number of community groups and environmental government organisations (*including; Friends of the Coorong, Friends of Shorebirds South East, SA Shorebirds Foundation, Friends of Adelaide International Bird Sanctuary, Tolderol Working Group, Nature Glenelg Trust, Limestone Coast Landscape Board, Lower Limestone Coast DEW National Park Rangers, and circulated in South Australian bird groups on Facebook*). Additionally, three community science events were held at the CLLMM Research Centre in Goolwa to engage the local community in the project face-to-face (on 8th October 2024, 21st January 2025, and 7th February 2025; Figure 3a and b).

Geocoded images were submitted through Birdlife Australia's Birdata online portal using the identifying word 'SSproject'. Each survey entry included an image, species of migratory shorebird photographed (e.g. sharp-tailed sandpiper, red-necked stint), number of individuals, date image was taken, and location (latitude and longitude, BirdLife survey site name).

This is the first time a project of this nature has been implemented in the region. To address gaps in community-collected data, a professional nature photographer (Darcy Whittaker) was employed to capture images of shorebird body condition in the same areas surveyed by the community. During these surveys, the photographer captured side-on images of individual shorebirds (as per Figure 1) at locations across the CLLMM region, including Tolderol Game Reserve, Murray Mouth, and Coorong National Park (Long Point, Mark Point, and Parnka Point).

Figure 2. Extract from the 'Shorebird Snapshots' pamphlet that was distributed to Adelaide, Goolwa, The Coorong, and South-East community members.

Figure 3. (a) Presentation delivered to community members on the 8th of October to encourage community engagement in the Shorebirds Snapshot Project; (b) Stall at the *Showcasing Citizen Science* event, hosted by the CLLMM Research Centre on the 7th of February 2025, where the Shorebirds Snapshot initiative was promoted.

4.2 Machine learning

Annotating body condition class is a time-consuming, costly, and error-prone task, so we used a machine learning approach to assign body condition scores to a catalogue of images of shorebirds from the Coorong, Lower Lakes and Murray Mouth. To assess the body condition of shorebirds in the Coorong, only images where the bird was standing approximately side-on (as per Figure 1) were included in the image catalogue for analysis.

4.2.1 Historical data image data

Models were trained on sets of images collated historically that represent multiple species in different postures, environments, and lighting, and that include all body condition classes: the 1-5 ordinal classes from lean to plump, as well as a "fluffed" class 6. The "fluffed" class categorises birds with an altered physical shape and silhouette when the feathers are "puffed out". The species list included Sharp-tailed Sandpiper (*Calidris acuminata*), Curlew Sandpiper (*Calidris ferruginea*), Red-necked Stint (*Calidris ruficollis*) and Common Greenshank (*Tringa nebularia*).

Expert elicitation was used to assign body condition scores (1-6) to the birds in these images, creating a labelled image data set for model training. The images used in this report include a set of 381 images collated by Rowan Mott, which were dominated by shorebirds labelled as condition class 2. This dataset was supplemented by a second collection of 316 images that contained a greater proportion of class 5 and class 6 labelled shorebirds. The merged set of 697 images was used for model training (Table 1).

Table 1. Number of images in each body condition class across the two historical datasets used for model training.

DATASET	CLASS 1 ("LEAN")	CLASS 2	CLASS 3	CLASS 4	CLASS 5 ("PLUMP")	CLASS 6 ("FLUFFED")
N = 316	46	44	28	73	47	78
N = 381	4	276	73	8	0	20

4.2.2 Machine Learning Pipeline

To address the project aim of advancing an automated system to score shorebird body condition across a range of shorebird species, we have developed a new machine learning pipeline using state-of-the-art computer vision that isolates a bird's morphology from its natural background before assessing its condition. The general approach utilises advanced object detection and segmentation models to detect birds in images, generate exact bounding boxes and binary masks, and apply them to the image to remove the background. These masked images isolate morphology and are then processed by a pre-trained Vision Transformer (ViT), which captures localised geometric features to predict the bird's body condition as an ordinal score rather than a nominal category. In simple terms, the vision transformer step evaluates an image by breaking it into sections (i.e., patches), encoding the patches to talk to one another so they understand the whole image (i.e., self-attention), and adding tokens to the sequence of patches to identify image features (i.e., prediction). The image feature information is then passed to an ordinal model that determines whether the image passes class thresholds (i.e., "is it greater than class 1?, and if so, is it greater than class 2?, and so on") to finally predict the image's ordered rank. In our application, this will predict a body condition score. More specifically, the overall architecture can be broken into three phases: (1) detection and segmentation, (2) multi-task training, and (3) evaluation and prediction (Figure 4).

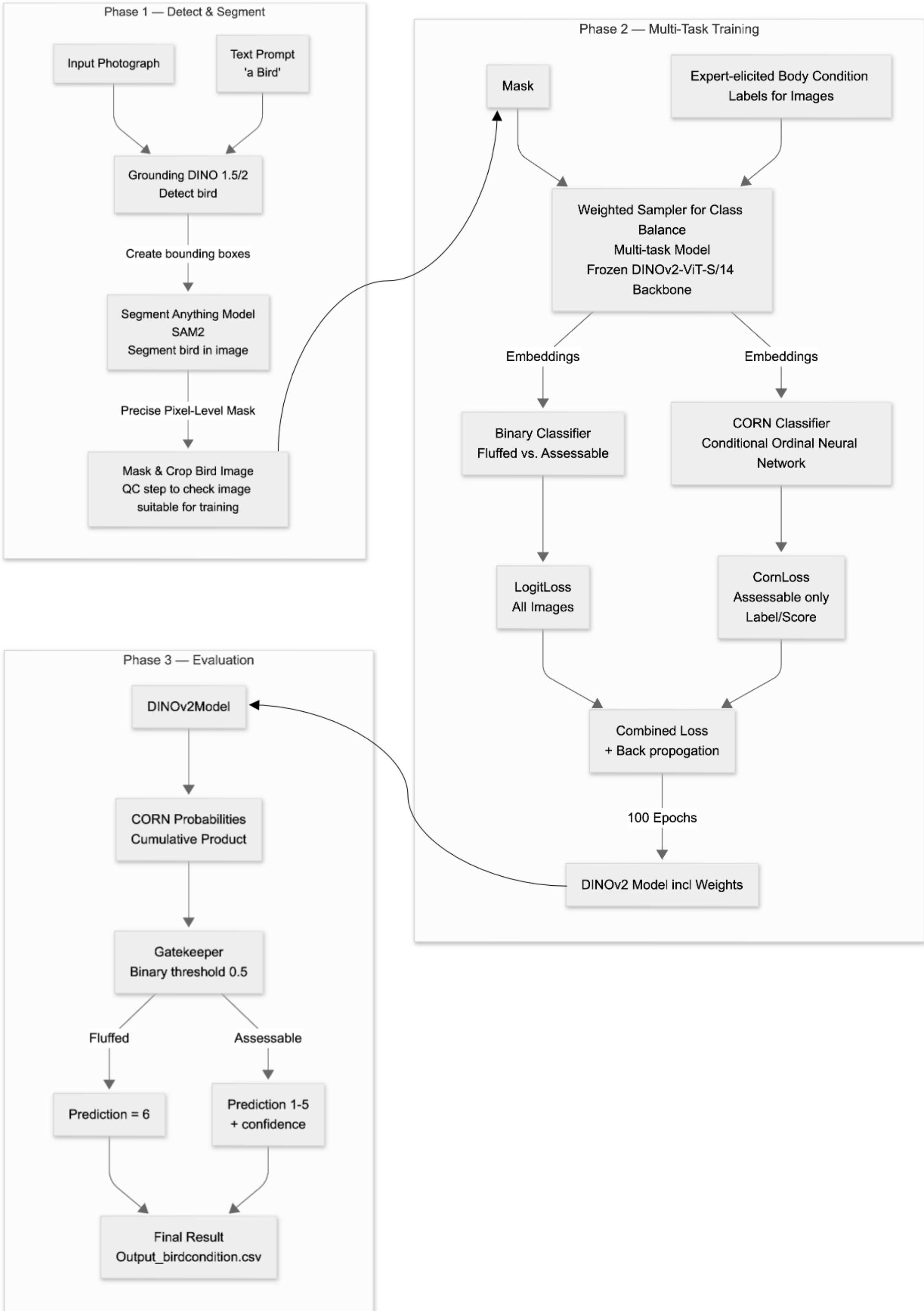


Figure 4. Machine learning pipeline for automated shorebird body condition scoring from photographs. Phase 1 detects shorebirds in input images, draws bounding boxes around them, and creates a pixel-level segmentation mask to separate them from the background. Phase 2 trains a multi-task vision transformer model to minimise a loss function for classifying shorebird body condition on a 1–5 ordinal scale, with a binary gatekeeper to detect fluffed (i.e., not assessable) birds (class 6) and save the neural network model weights. Phase 3 evaluates the conditional probabilities of the ordinal classifier, excluding fluffed individuals, and converts the values to interpretable 1–5 scores with a calibrated confidence metric.

4.2.2.1 Phase 1: detection and segmentation

The overall approach builds on the ViT-based foundation model DINOv2 (Oquab et al. 2024), a highly generalisable self-supervised learning model trained on more than 140 million images without labels. However, the first step of the pipeline localises the target, here a shorebird, using the Grounding DINO model, a variant of the DINO model that incorporates grounding information, which can help the model to better understand the spatial relationships between different parts of the image (Liu et al. 2024). This can be particularly useful for tasks such as body condition scoring, where the model needs to focus on specific body features (e.g., breast, abdomen) to make accurate predictions. Instead of providing weights or labels, the model input is simply a text prompt describing the target, such as "a bird", and the model uses this prompt in a “zero-shot” detection to locate the subject in an image. A bounding box containing the bird is then generated as output.

Next, the Segment Anything Model 2 (SAM 2; Ravi et al. 2024) is loaded to generate a pixel-level segmentation mask from the detected bounding box. This step isolates foreground bird pixels for downstream quality control and optional background removal. SAM 2 is a foundation model designed for prompt-directed visual segmentation for both static images and videos. It is trained on the massive SA-V dataset (Ravi et al. 2024), the largest video segmentation dataset to date, and is significantly more accurate and much faster than the original Segment Anything Model. SAM2 generates accurate, pixel-level segmentation masks for quality-control filtering (e.g., detecting occlusion or truncation at the boundary, ensuring the mask adequately covers the cropped image, and minimal image blur) before the image is passed to the classifier. To ensure that only birds in appropriate postures and orientations (e.g., clear side-on profiles) were assessed for body condition, the automated screening phase removed unsuitable images (such as birds in flight, facing the camera, or partially submerged). Rather than relying on manual removal, the pipeline utilises mask-derived quality control metrics during the segmentation phase to filter out heavily truncated or obscured crops. The output of this phase is the segments that can be fed into the DINOv2 backbone as an additional input channel (a ‘4th channel’; see below) to help the model focus on relevant features of the bird's body condition.

4.2.2.2 Phase 2: Multi-task training

Phase 2 introduces the cropped shorebird images, along with expert-elicited labels indicating the body condition score of the detected birds, to the DINOv2 model. The model learns to extract robust features from images that are invariant to domain shifts (i.e., changed information relative to pre-trained images), making it well-suited for tasks like body condition scoring, where the training data may not perfectly match the test data. The DINOv2 backbone does not know the specific definition of "body condition" and therefore must be fine-tuned on the specific labelled dataset to train the classification of shorebird body condition into ordinal classes (1 = lean to 5 = plump).

The labelled images used to train the model were not equally balanced across the five body condition classes (particularly at the lean and plump extremes). Class-balanced sampling with inverse frequency weighting (i.e., a weighted sampler) was therefore used to oversample the underrepresented extreme body condition classes to prevent classifier collapse at the boundary classes (Buda et al. 2018).

Whilst the explanation so far has focused on predicting ordinal body condition classes (1-5), images were sometimes scored by experts as class 6, meaning the shorebird was "fluffed up", preventing assessment of its current condition. To address this factor in the pipeline, we implemented a multi-task model at this stage to separate the "is this assessable?" decision from the "what is the score?" decision. The image is first assessed against a binary classifier as being "assessable" or "fluffed" using a Cross-Entropy loss function against the predicted class. If the model flags the bird as "fluffed", the image is immediately assigned as unsuitable, and the body condition score is set to "None", so the pipeline only proceeds to consider the 1-5 ordinal prediction if the bird is confirmed assessable.

Standard multi-class classification (Cross-Entropy) ignores the natural biological ordering of body condition scores. To respect this ordering of body condition classes in our classification step, the ViT's embeddings (i.e., how it codes the patches) are passed into a specialised Conditional Ordinal Regression for Neural Networks (CORN) model (Shi et al. 2023). This model evaluates $K-1$ independent conditional probabilities (e.g., "what is the probability the bird is > condition 1?", "> condition 2?", etc.), where K is the number of classes. CORN therefore guarantees mathematical rank consistency, meaning that the probability of exceeding each subsequent class boundary declines, which is consistent with the biological notion of relative body condition. The pipeline trains the binary gatekeeper and ordinal scorer jointly while ensuring the ordinal loss is never computed on "fluffed" birds. The model was trained over 100 epochs (i.e., training blocks).

The first three channels of the DINOv2 backbone's input are the standard RGB (red/green/blue) channels of the image. These channels contain the image's colour information, which is crucial for feature extraction and classification. The fourth channel (mentioned above) added to our modified architecture is the mask generated by SAM 2 that highlights the bird in the image, allowing the model to focus on the relevant features for body condition scoring. This fourth channel is a data augmentation technique that expands the amount of information provided to the model for training. Research demonstrates that mixing background-removed images into the training set, as a form of data augmentation, forces classification models to focus explicitly on the physical characteristics and structural details of birds (Zhao and Lee 2025). For example, because a "fluffed" bird is characterised entirely by an altered physical shape and silhouette (its feathers are puffed out), removing complex background clutter isolates this exact morphological feature. It prevents the model from overfitting to irrelevant environmental features or noisy patterns that can confuse ViTs. Utilising background removal as an augmentation technique can achieve species classification accuracy of 99.5%, significantly outperforming models trained with traditional augmentations (i.e., rotation, flipping, translation) alone (Zhao and Lee

2025), and can improve performance when identifying birds in non-traditional body postures that differ from common standing positions.

By gradually increasing the masking probability, you can encourage the model to first learn general features and then focus more on the bird's morphology as training progresses, thereby further enhancing its ability to distinguish subtle differences in body condition. This means early epochs train on richer contextual images, whilst later epochs progressively target background-invariant features of the bird, which is recommended in fine-grained recognition tasks such as this one (Hacohen and Weinshall 2019). During training, masking was applied stochastically (with probability ramping from 0.2 to 0.6 so that by the final training epoch, masking probability is highest (meaning only 40% of training images were seen unmasked). The model was never trained solely on background-removed images, as this can sometimes degrade performance by stripping away all contextual diversity (Zhao and Lee 2025).

Model accuracy was assessed on the same images used to train the model and therefore represents a model refitting estimate, which is expected to be an optimistic estimate of model performance on new images. Therefore, it should be interpreted as an upper bound rather than an unbiased estimate of prediction accuracy. The limited number of expert-scored images precluded hold-out cross-validation on independent sets of images while maintaining adequate training sample sizes. Two aspects of our approach minimise this limitation: the random background-removal augmentation regularises the model and reduces its dependence on the specific training images, and the conditional ordinal form of the model makes the best possible use of the labelled data because of the ordering of the condition classes.

4.2.2.3 Phase 3: Evaluation and prediction

The final stage of the pipeline takes the integrated outputs from phase two, which merge the labelled input data with the model-estimated probabilities, mark QC-failed images as unsuitable, and retain a record of every input image, including detection failures. The predictions of body condition are ranked to determine how many conditional probabilities exceed 0.5, indicating how many "conditions" are supported, and thereby arriving at a predicted score. Importantly, a confidence score is also calculated to quantify model certainty across all binary tasks (i.e., how far the probabilities are from the ambiguous 0.5 threshold) and is scaled from 0 (uncertain) to 1 (certain), allowing confidence in predictions to be evaluated, as well as flagging borderline morphologies for manual review.

The entire machine learning pipeline has been constructed using bespoke Python scripts. The tasks require specialised high-performance computing resources, including GPUs for parallel processing, to effectively train the model. This is because the multi-task learning framework involves training on a large, complex dataset, which can be computationally intensive, and because transformer-based image processing is highly parallel. The use of GPUs enables faster processing and training times, enabling us to achieve better results in less time than with traditional CPU-based training methods. To this end, our research group has set up virtual machines on the Australian Research

Data Cloud (Nectar Research Cloud) with the necessary compute resources and the specific Python software environments required to train our new model architecture.

4.2.2.4 Post-hoc assessment of image features that discriminate body condition

Attention mechanisms in ViTs compute the relative importance of different image patches. This allows the model to explicitly focus on discriminative local regions of the image, such as the contour of the breast and underbelly, while suppressing background noise and irrelevant body parts such as the beak or wings (Chowdhury et al. 2025). The Prompt-CAM approach injects learnable "prompts" (i.e., Visual Prompt Tuning) into the frozen DINOv2 backbone during model training, forcing the model to produce clear, visual heatmaps that pinpoint the exact physical traits driving body condition classification. Tests on bird datasets demonstrate that Prompt-CAM successfully isolates specific physical zones, such as a bird's belly or breast, to explain its predictions (Chowdhury et al. 2025).

We have developed an additional pipeline that runs independently of the 3-phase method described above and trains a Prompt-CAM model using DINOv2 with register tokens and class-specific prompt vectors tailored to our body condition assessments, extending the approach by Chowdhury et al. (2025). This provides crucial biological auditability, providing a visual attention map for manual review to assess how the model scored the bird based on its underside, rather than image artifacts.

A limitation of this approach is that we forfeit the CORN ordinal classifier used in the formal model training and inference described above. This is necessary as there is currently no method to incorporate class-specific prompts in an ordinal model, nor to implement a multi-task approach that separates a binary gatekeeper for detecting “fluffed” birds from the ordinal body condition classifier. Instead, the approach reverts to using the Cross-Entropy loss function and stochastic gradient descent to evaluate the nominal 1-6 categories. This means that for post-training inference, we generated class-specific heatmaps to inspect images produced by the model’s decision-making process from its multiple parallel attention pathways, or 'heads'. Rather than pooling these pathways, we visualise the attention map of the highest-ranked head for a given test image.

Consequently, this method should be viewed as an add-on step that can aid biological interpretation, with the caveat that it will not be as accurate as our ordinal model. Further validation of the approach is required, noting that a single training run over 100 epochs on a data set of 697 images takes ~40 hours on a high-performance computing server.

4.2.3 Advances from the previous approach to predict shorebird body condition

The following section explains key differences between the methods presented in this report and our previous approach (Jackson et al. 2022). There have been significant advances in computer vision models (ViT’s), which we have harnessed to provide a state-of-the-art pipeline that simplifies and modularises the estimation of shorebird body condition, designed to be more flexible and adaptable across species.

While the DINOv2 backbone we have used is pre-trained on massive unlabelled datasets (i.e., self-supervised learning) to generate robust features, the classification step in our pipeline still requires specific labelled data to map those features to body condition scores. In contrast, our previous approach used labelled images for the nominal classification of body condition classes (Cross-Entropy Loss) but supplemented this with an Unsupervised Domain Adaptation (UDA) module that generated pseudo-labels (using K-means and cosine similarity) for unlabelled data to improve performance. Our new approach relies on the generalisable features learned by DINOv2 during its self-supervised pre-training to handle domain shifts, which reduces or eliminates the need for the complex UDA/pseudo-labelling machinery used previously (Oquab et al. 2024).

The previous approach used the YOLO (You Only Look Once) architecture for object detection. Specifically, the yolov3.weights from the YOLOv3 (Redmon and Farhadi 2018) model were used to detect a bird, which is restricted to the specific object classes for which YOLO was originally trained. The yolov3.weights and the pre-trained suitability prediction model components are replaced by the more advanced, generalisable Grounding DINO foundation model and objective metrics that offer higher accuracy and transparency. Grounding DINO is an "open-set" detector that pairs a Transformer-based architecture with language understanding, allowing it to detect arbitrary objects using text prompts. This makes it significantly more robust to domain shifts (e.g., changes in lighting or background) compared to the older YOLOv3 architecture.

The previous approach used a trained "black box" neural network, ResNet50 (He et al. 2016), to predict a binary "suitable/unsuitable" label. The new approach uses the high-precision segmentation masks generated by SAM 2. By calculating objective geometric properties, such as the percentage of the crop the bird fills (foreground proportion) or whether the mask touches the edge (truncation), we obtain a scientifically auditable metric for "suitability" without training a separate model.

The previous approach then predicted the bird body condition label and estimated confidence scores using a complex approach that attempted to train a self-tuning neural network to extract information from new, unfamiliar images with associated expert-elicited labels. Pairs of raw and segmented images for each bird were passed through the ResNet50 backbone, and the network was optimised using a standard Cross-Entropy Loss function. Because the model struggled when the background or lighting changed at new camera sites (domain shift), it tried to learn from unlabelled data. It did this by finding the "average look" (centroids/prototypes) of birds for each of the five body-condition classes and measuring how closely the new bird matched those 5 "average birds" using Unsupervised Domain Adaptation (UDA) & K-Means. It then assigned a fake label to the image based on the closest match, using that guess to further train itself. The new approach completely bypasses the background problem by using SAM 2 to deterministically mask the background to pure black before normalisation, isolating the bird. There is no need for domain adaptation because the confusing background effects on classification are minimised by the progressive masking proportion applied during model training. The patch-based attention strategy of DINOv2 inherently learns fine-grained morphological details more effectively than ResNet50, eliminating the need for bilinear pooling or unsupervised adaptation. Finally, instead of treating body condition classes as unordered and guessing a single independent category, the ordinal CORN model respects the natural order of the body condition classes and evaluates a sequence of conditional boundaries consistent with the biological reality of the measure.

5 Results

5.1 Community outreach and other shorebird image data sources

As a result of community engagement, a total of 24 photo entries were submitted to the Shorebirds Snapshot project during the reporting period by three unique users (Table 2), representing 44% of the total migratory shorebirds uploaded to Birdata in the region. Of the total photos submitted, the Sharp-tailed Sandpiper was the most recorded migratory shorebird, followed by the Red-necked Stint. Single records were also obtained for Common Greenshank, Curlew Sandpiper, and Red Knot. Photos were extracted from Birdata twice during the project, corresponding to consecutive migratory shorebird seasons: 1 September 2024 to 31 May 2025, and 1 September 2025 to 30 March 2026.

Table 2. Summary of shorebird surveys submitted by the public between 2024 and 2026 across the Coorong to Victorian border, extracted from Birdata. Values represent nested subsets, where surveys containing migratory shorebird (MS) photos are a subset of total surveys, and surveys submitted to the Shorebirds Snapshot (SS) project are a subset of those containing MS photos. Values in square brackets [*n*] indicate the number of unique photos submitted each period, as multiple photos can be uploaded per survey.

SURVEY PERIOD	TOTAL NUMBER OF SURVEYS	SURVEYS CONTAINING MS PHOTOS	OF THESE: SUMBITTED FOR SS PROJECT	TOTAL NUMBER MS PHOTO USERS	NUMBER OF SS PROJECT USERS	PERCENTAGE OF MS PHOTOS SUBMITTED TO SS PROJECT
2024–2025	85	16 [28]	7 [*17*]	5	2	60.7%
2025–2026	98	13 [27]	3 [*7*]	4	1	25.9%

The community outreach data was supplemented with images collected by researchers working in the CLLMM region. In total, 1947 photographs were collated by Darcy Whittaker across 19 targeted photo surveys (Figure 6) conducted between September 2025 and March 2026, providing a robust, standardised dataset for evaluating shorebird body condition (Table 3 and Figure 5). Consistent with the community photo dataset, Red-necked Stint and Sharp-tailed Sandpiper were the most frequently photographed species at these sites (Table 3). For later analysis, locations were grouped into regions: Tolderol (Tolderol Game Reserve and Mosquito Point); North Coorong (Murray Mouth, Pelican Point Road, Mark Point Boat Ramp, Long Point, Coorong N50W); Central Coorong (Parnka Point, Villa De Yumpa, Coorong S21C), and South Lagoon (Morella Basin, Loop Road).

Table 3. Number of images of three shorebird species (n = 1947) that were used to score body condition (Source - Darcy Whittaker). Images have been grouped into four regions as described in the caption of Figure 5.

SURVEY PERIOD	REGION	RED-NECKED STINT	SHARP-TAILED SANDPIPER	CURLEW SANDPIPER
Late 2025	Tolderol	158	94	2
	North Coorong	685	22	41
	Central Coorong	180	-	-
	South Lagoon	26	-	-
Early 2026	Tolderol	481	118	81
	North Coorong	-	-	-
	Central Coorong	48	-	1
	South Lagoon	8	2	-

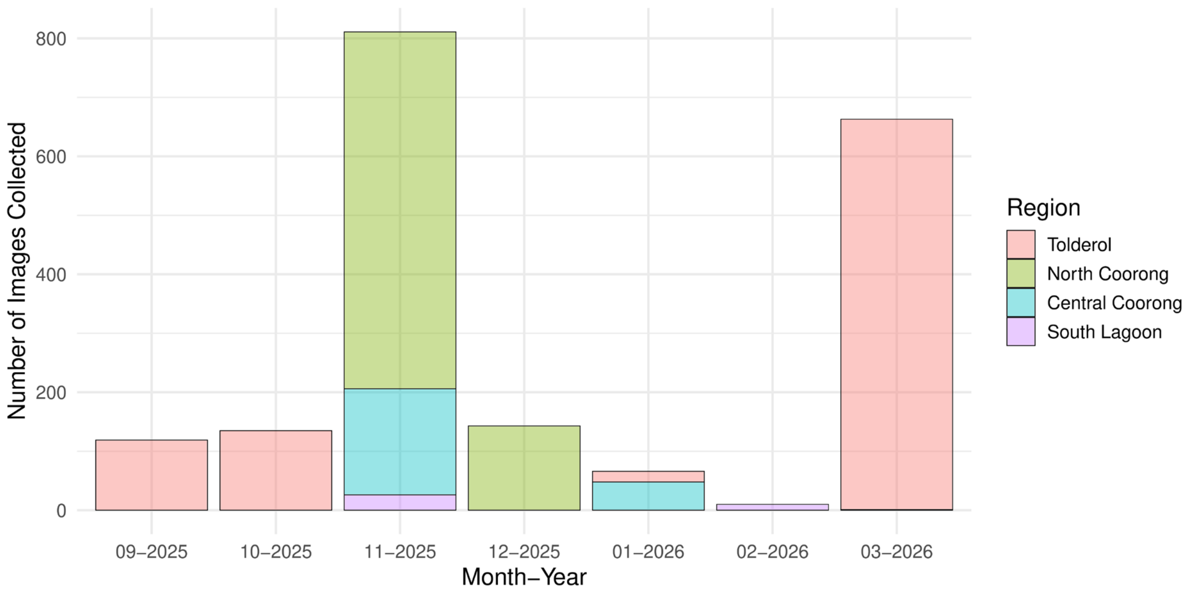


Figure 5. Stacked histogram showing the number of images collected each month from September 2025 to March 2026, grouped by region (Tolderol, North Coorong, Central Coorong, South Lagoon).

Figure 6. Locations of the shorebird photographs supplied by Darcy Whittaker that were scored for body condition of three shorebird species (n = 1947; Source - Darcy Whittaker). Locations were grouped into regions for subsequent analysis: Tolderol (Tolderol Game Reserve and Mosquito Point); North Coorong (Murray Mouth, Pelican Point Road, Mark Point Boat Ramp, Long Point, Coorong N50W); Central Coorong (Parnka Point, Villa De Yumpa, Coorong S21C), and South Lagoon (Morella Basin, Loop Road).

5.2 *Classification of body condition*

5.2.1 Model validation against the expert-labelled body condition class

Body condition classification was based on the DINOv2 model, combined with multi-task binary and ordinal classifiers trained on 697 images scored by experts for body condition. Model-based prediction probabilities were used to produce a ranked body condition class, and these predictions were compared against the expert-elicited "ground truth" body condition value. The model's classification accuracy was evaluated using a confusion matrix that cross-classified predictions with ground-truth labels. The results show a predominance of scores along the diagonal of the confusion matrix, indicating high model accuracy (Figure 7). Accuracy was lower when classifying the lean, or poorer condition, end of the scale; the predicted body condition class was often higher than the expert-elicited score. So, the model was failing to recognise birds in the leanest condition category. The model's overall accuracy was also supported by the fact that, when misclassifications occurred, they were, in most cases, predictions into the next adjacent category, and there were very few instances of multiple class deviations from the expert assessment (Figure 7).

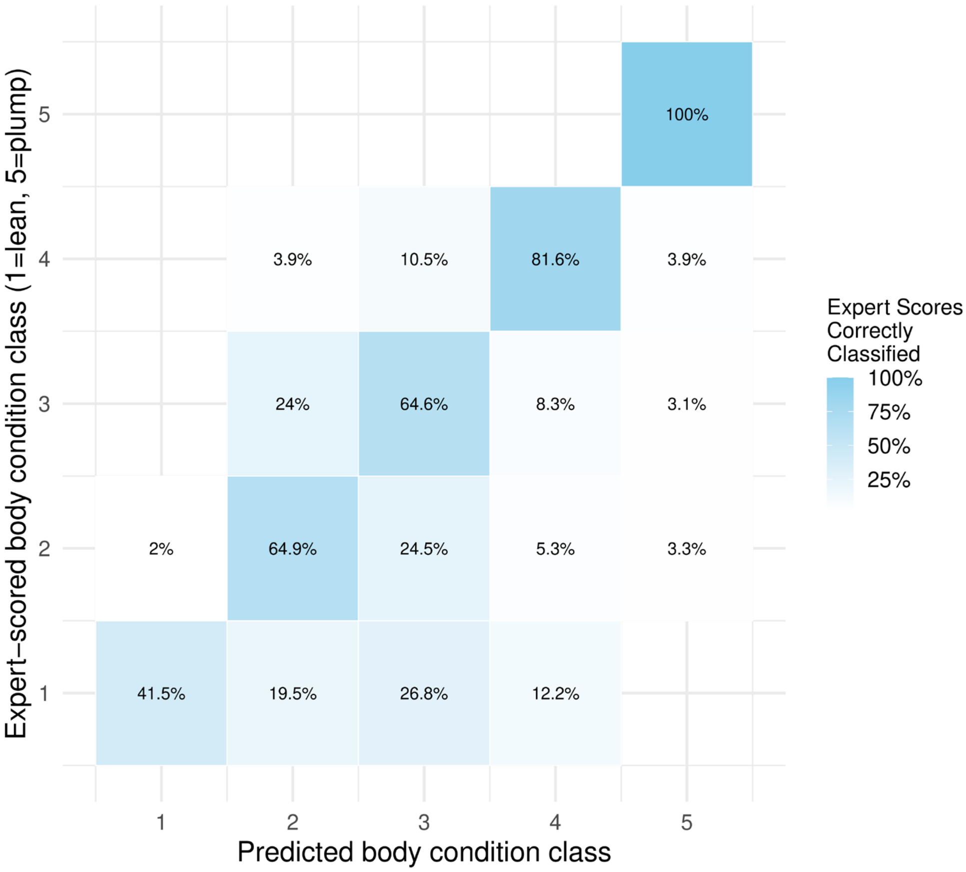


Figure 7. Classification accuracy of expert-elicited shorebird body condition scores (n = 697) measured against Vision transformer model-predicted classes. Body condition scores were calculated from conditional ordinal cumulative probabilities of body condition exceeding a given class, converted to interpretable 1–5 scores. Accuracy is indicated by high diagonal percentages, with deviations only into adjacent classes. Model accuracy was lowest when classifying poor condition, due to limited labelled images of 'lean' shorebirds (class 1) available for training.

Confidence values for the body condition class predictions where the bird was correctly classified against the expert-elicited score indicated strong model reliability (generally above 0.5 with median scores 0.5-0.8), particularly for body condition classes 2 through to 5, whereas certainty was markedly lower for the lean class 1 body condition (Figure 8). In contrast, images misclassified against the expert assessment, represented by off-diagonal classifications in the confusion matrix (Figure 7), generally had substantially lower overall confidence in the model scores than correctly classified cases (Figure 8). Once again, the lowest confidence was reserved for misclassified body condition scores in the lean class 1 category.

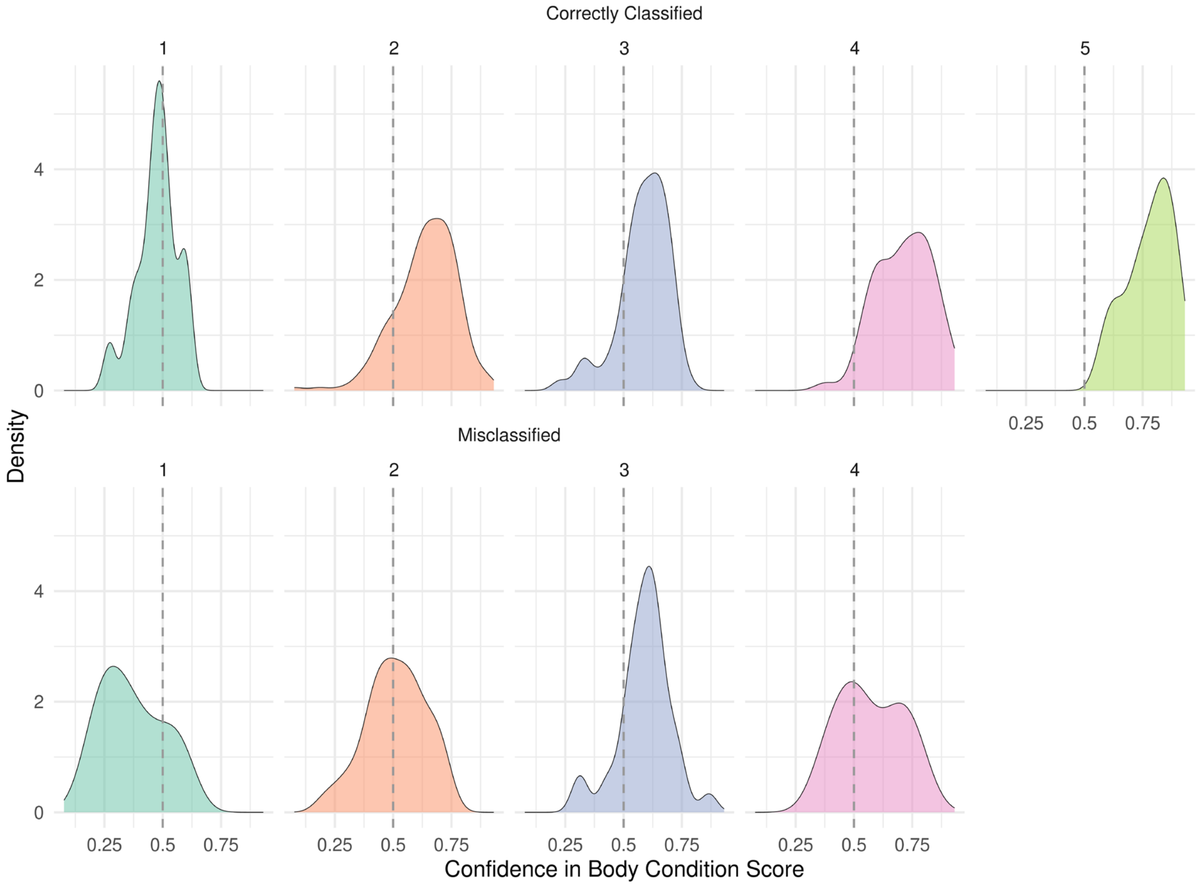


Figure 8. Density plots of the distribution of confidence scores for each shorebird body condition class for images correctly classified (top row) and images misclassified (bottom row) by the vision transformer model. A score close to 0 indicates low certainty in predicting body condition class, while a score close to 1 indicates high certainty (vertical dashed lines shown at 0.5 to assist visualising distribution shifts). Confidence was low for class 1 predictions, consistent with low classification accuracy, and was also (expectedly) relatively lower for all misclassified images.

5.2.2 Inferring body condition for unlabelled images

Applying the trained DINOv2 + CORN model to the 1,947 unlabelled images collated from the CLLMM region over 2025-2026 indicated substantial variation in predicted body condition among species, as well as differences across sampling periods and geographic regions of the CLLMM.

Red-necked Stints, the most numerous species recorded in the dataset, scored a median body condition of class 3 (i.e., a bird with minor surplus fat reserves, defined here as an intermediate condition) across regions in 2025, although in the North and Central Coorong, there was a substantial proportion of individuals in lower (class 2) and even lean (class 1) condition in 2025 (Figure 9). In 2026, the median condition score for Red-necked Stints at Tolderol was class 2, with a substantive number of individuals in poorer (class 1) and intermediate (Class 3) condition. However, the modal condition was estimated as class 2 at Tolderol, which contrasted with the modal class for this species in the Central Coorong, which indicated that most birds were in intermediate condition (Figure 9). Sharp-tailed sandpipers were scored as being in substantively above intermediate condition (Class 4) at Tolderol in 2025, whereas in early 2026 that distribution had shifted markedly to lower-condition individuals (Figure 9). Low-condition individuals were also more common in the North Coorong in 2025 for this species. Curlew Sandpiper had a median score between classes 3 and 4 in both 2025 and 2026, although the modal distribution was class 4 in 2026, rather than class 3 in 2025 (Figure 9).

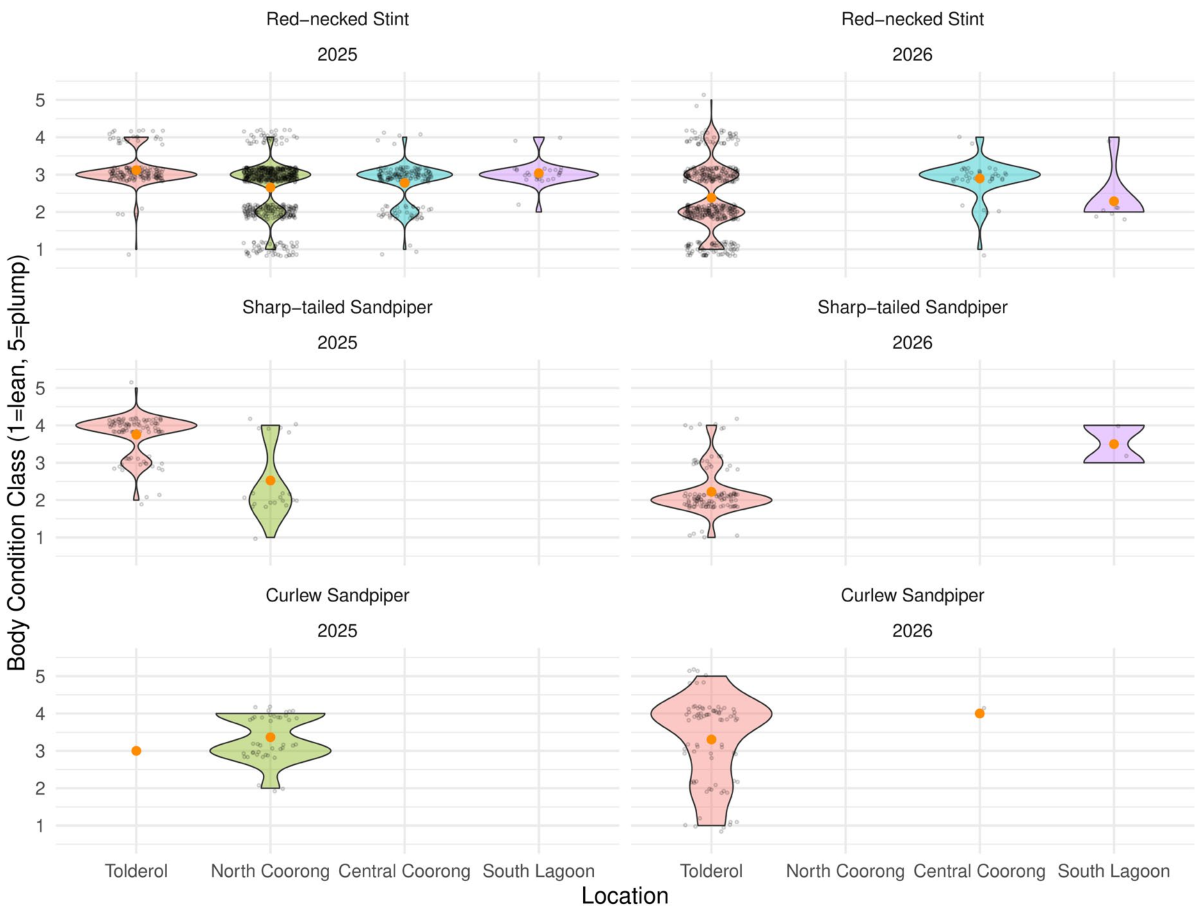


Figure 9. Violin plots showing the distribution of body condition scores for three species of shorebirds photographed over late 2025 to early 2026 (Source: Darcy Whittaker). Violins are widest at the highest modes in the discrete distribution of body condition scores. Points show class scores for individual images and are jittered around both the body condition class and the region (shown in different colours) to aid visualisation. Orange solid points are class means for each species in each region.

Overall, these results indicate: (a) the capability of the automated body condition scoring pipeline to classify birds across a wide range of condition classes, (b) differences in the distribution of condition scores both spatially from the Lower Lakes through the North Coorong and down to the Southern Lagoon, and (c) between-year differences, which in this case reflect seasonal variability as the 2025 scores were recorded for images taken during the third and fourth quarters of 2025, while the 2026 data was recorded during the first quarter of 2026 (Figure 9).

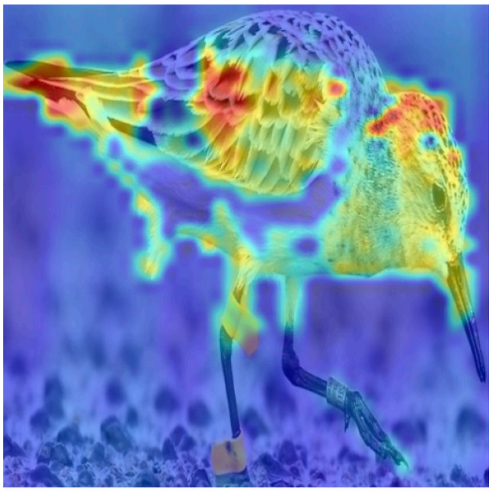
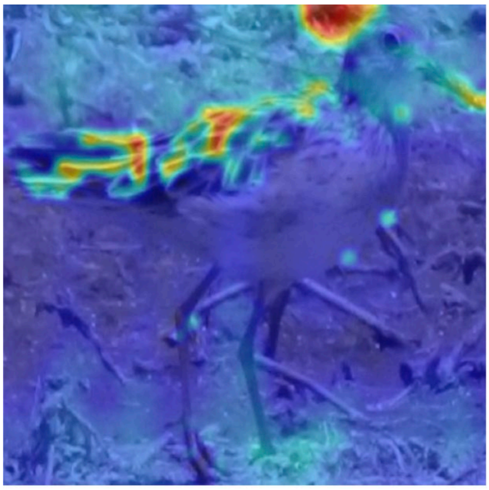
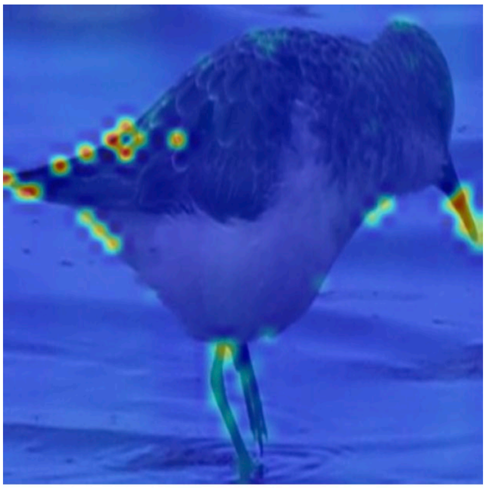
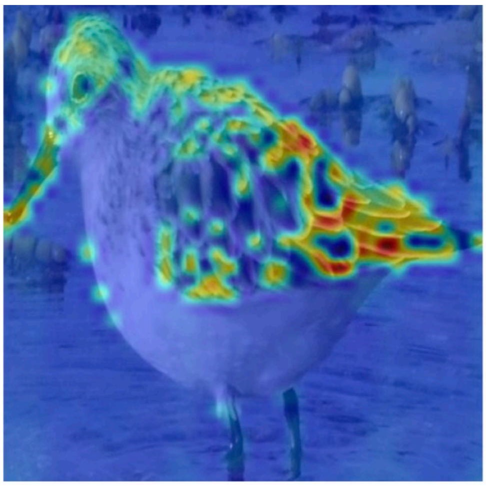
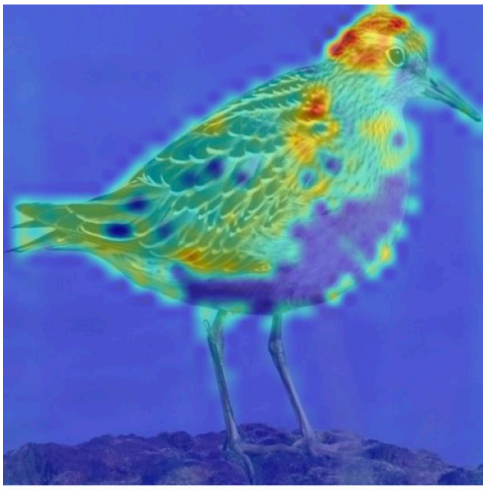
Images of shorebirds that were assessed as lean (Class 1), intermediate (Class 3) and plump (Class 5) for two of the most sampled species, Red-necked Stint and Sharp-tailed Sandpiper, provide context for the body condition assessment (Table 4).

Table 4. Example images of Red-necked Stint and Sharp-tailed Sandpiper classified by experts' scores of relative body condition for classes 1, 3, and 5.

BODY CONDITION CLASS	RED-NECKED STINT	SHARP-TAILED SANDPIPER
Class 1 (Lean)		
Class 3 (Intermediate)		
Class 5 (Plump)		

The utility of attention maps derived from the Prompt CAM machine learning model for the same six individuals shown in Table 5 shows variable results in model predictions. The attention maps provide hotspot assessments of components of the image that were targeted during training of the prompt CAM model. Importantly, for both Class 3 and Class 5 body condition individuals, there are clear hotspot areas across the underside of the birds, indicating that the areas experts use to assess relative body condition are also identified by the model in condition scoring. Clearly, several other areas of the images are drawing the model's attention and are not consistent with the body condition differentiation. It's worth noting again that the Prompt-CAM model uses a Cross- Entropy loss function to evaluate unordered categories of body condition and includes class 6 birds that cannot be assessed for condition. These two factors have a substantive impact on model accuracy. For these reasons, whilst this method is a useful additional visual guide to interpreting scores, the images should be interpreted with some caution, whilst noting the very promising result of hotspots of attention on the underside of the bird in scoring body condition.

Table 5. Attention maps for images of Red-necked Stint and Sharp-tailed Sandpiper highlight hotspot areas (red pixels) of the shorebird images that are most informative for assessing differences in body condition using the trained vision transformer and conditional ordinal regression neural network model. Panels show the same images as presented in the previous figure.

BODY CONDITION CLASS	RED-NECKED STINT	SHARP-TAILED SANDPIPER
Class 1 (Lean)		
Class 3 (Intermediate)		
Class 5 (Plump)		

6 Discussion

Using shorebird body condition to assess habitat quality is an appealing proxy measure because, in theory, the physical condition of the bird is a direct reflection of whether it is gaining sufficient resources from the local environment to maintain good fitness (Swift et al. 2020). An improvement in the body condition of migratory shorebirds just prior to departure is further evidence that there are sufficient resources to support the extreme pre-migration fattening that some species require.

Given the advances we have made in developing a robust approach for automating the scoring of body condition in shorebirds, an interesting area for future study would be to compare the body condition of shorebirds found in the Coorong to that in other parts of Australia, especially if the model could be further improved by training on a larger dataset or expert-scored images. If body condition in the Coorong is lower than elsewhere during a given year, this would suggest that habitat condition in the Coorong was poor. However, we acknowledge that the ability to detect change in body condition using this method will be further improved by resolving predictions for shorebirds in the lean (i.e., the poorest) condition. This additional goal will improve the assessment of migration impacts and enable monitoring of the health of the Coorong and the wider region in terms of habitat quality.

6.1 Recommendations

There are several specific recommendations to improve the accuracy and robustness of our body condition scoring pipeline, based on the latest research in computer vision and machine learning for ecological applications. These recommendations include:

- Expand the database of expert-labelled images to ensure a more balanced training information base, particularly at the extremes of body condition. As outlined in the methods section, the conditional ordinal regression neural network (CORN) model is sensitive to class imbalance at extreme ranks, such as when body condition classes 1 and 5 are underrepresented. This outcome can cause those boundary values to collapse toward zero, producing a unimodal probability distribution centred on the middle class (Class 3; Beckham and Pal 2017). Central tendency bias (or hedging) describes the case where model predictions systematically default to the middle of the ordinal range, with low boundary confidence across class levels, and is a common problem in ordinal regression. Class-balanced sampling with inverse frequency weighting was used to oversample underrepresented extreme body condition classes. However, boundary classifier collapse cannot be resolved by loss modifications alone if the training signal at those boundaries is unclear (Buda et al. 2018).
- When expanding the database of expert-labelled images across body condition classes and shorebird species, it will also be important to compute inter-rater agreement (e.g., Cohen's weighted Kappa) for the expert annotations since the degree of ambiguity about where the boundary lies between each pair of condition classes in the ground truth data directly informs how the loss function being minimised during model fitting can be adapted to better discriminate

between class boundaries. Currently, training labels are standardised to a single expert, providing no information about observer bias or inter-observer agreement. The aim would be to monitor whether hybrid loss functions that incorporate inter-rater reliability correctly sharpen boundary confidence without causing overfitting to the training set of images. If successful, this could further enhance the model's ability to distinguish between closely related body condition classes, a common challenge in body condition assessments.

- The classification model was constructed to incorporate independent cross-validation for model fitting conditional on sufficient sample sizes. Therefore, future work will evaluate an out-of-sample estimate obtained via a hold-out set or k-fold cross-validation with a larger sample of labelled images.
- Low participation in the citizen science program, as evidenced by the small number of community-submitted photographs, suggests that increased visibility and accessibility are key areas for improvement. Increased promotion through social media campaigns to reach wider networks, making the submission of photographs easier to reduce barriers to engagement (e.g., designing an app that integrates with smartphone cameras), and workshops in regional communities should be considered collectively.
- It will be important to evaluate the false negative rate of the binary "fluffed" bird classifier, given that undetected fluffed individuals entering the ordinal pipeline could cause 'middle-class' inflation that adaptations of the loss function alone are unlikely to resolve.
- To improve boundary discrimination, especially for distinguishing between closely related body condition classes, a two-step process could be implemented: first training the model's output layer (the 'heads'), and later gradually unfreezing deeper parts of the model to refine its ability to detect subtle differences in shorebird morphology.

7 References

Beckham C, and Pal C (2017). Unimodal probability distributions for deep ordinal classification. *arXiv*, 705.05278

Van Bemmelen RSA, Clarke RH, Pyle P, and Camphuysen K (2018). Timing and duration of primary molt in Northern Hemisphere Skuas and Jaegers. *The Auk,* 135 (4), 1043–1054.

Bierlich KC, Karki S, Bird CN, Fern A, and Torres LG (2024). Automated body length and body condition measurements of whales from drone videos for rapid assessment of population health. *Marine Mammal Science*, 40(4), e13137.

Blair JD, Gaynor KM, Palmer MS, and Marshall KE (2024). A gentle introduction to computer vision-based specimen classification in ecological datasets. *Journal of Animal Ecology*, 93(2), 147–58.

Buda, M, Maki A, and Mazurowski MA (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259.

Chowdhury A, Paul D, Mai Z, Gu J, Zhang Z, Mehrab KS, Campolongo EG, Rubenstein D, Stewart CV, Karpatne A, and Berger-Wolf T (2025). *Prompt-CAM: Making vision transformers interpretable for fine-grained analysis*. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4375–4385).

Féret M, Bêty J, Gauthier G, Giroux J-F, and Picard G (2005). Are abdominal profiles useful to assess body condition of spring staging greater snow geese? *Ornithological Applications*, 107(3), 694–702.

Hacohen G, and Weinshall D (2019). On the power of curriculum learning in training deep networks. *arX*iv, 1904.03626.

Hayes MC, Gray PC, Harris G, Sedgwick WC, Crawford VD, Chazal N, Crofts S, and Johnston DW (2021). Drones and deep learning produce accurate and efficient monitoring of large-scale seabird colonies. *Ornithological Applications*, 123(3), duab022.

He K, Zhang X, Ren S, and Sun J (2016). *Deep residual learning for image recognition*. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–78.

Hodgson JC, Holman D, Terauds A, Koh LP, and Goldsworthy SD (2020). Rapid condition monitoring of an endangered marine vertebrate using precise, non-invasive morphometrics. *Biological Conservation*, 242, 108402.

Hodgson JC, Mott R, Baylis SM, Pham TT, Wotherspoon S, Kilpatrick AD, Raja Segaran R, Reid I, Terauds A, and Koh LP (2018). Drones count wildlife more accurately and precisely than humans. *Methods in Ecology and Evolution*, 9(5),1160–1167.

Jackson MV, Mott R, Prowse TAA, Delean S, Shu Y, Liu L, Hunt BJ, Sanchez-Gomez S, Brookes J, and Cassey P (2022). *Recommended habitat quality measures for key waterbird species in the Coorong*. Report by CSIRO for the Goyder Institute for Water Research. CSIRO, Canberra.

Lacy H, De Cuyper A, Dalerum F, Tosoni E, Clauss M, Ciucci P, and Meloro C (2024). Estimating body condition of apennine brown bears using subjective scoring based on camera trap photographs. *Mammal Research*, 69(3), 355–64.

Landys MM, Piersma T, Guglielmo CG, Jukema J, Ramenofsky M, and Wingfield JC (2005). Metabolic profile of long-distance migratory flight and stopover in a shorebird. *Proceedings of the Royal Society B: Biological Sciences*, 272, 295–302.

Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, Jiang Q, Li C, Yang J, Su H, Zhu J (2024). *Grounding dino: Marrying dino with grounded pre-training for open-set object detection*. In European conference on Computer Vision. pp. 38–55. Springer Nature, Cham, Switzerland.

Lürig MD, Donoughe S, Svensson EI, Porto A, Tsuboi M (2021). Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. *Frontiers in Ecology and Evolution*, 9, 642774.

Oquab M, Darcet T, Moutakanni T, Vo H, Szafraniec M, Khalidov V, Fernandez P, Haziza D, Massa F, El-Nouby A, Assran M (2023). Dinov2: Learning robust visual features without supervision. *ArXiv*, 2304.07193.

Osborne A, and Ryan PG (2021). Using Digital Photography to Study Moult Extent in Breeding Seabirds. *Ostrich*, 1–4.

Ravi N, Gabeur V, Hu YT, Hu R, Ryali C, Ma T, Khedr H, Rädle R, Rolland C, Gustafson L, and Mintun E (2025). *Sam 2: Segment anything in images and videos*. In International Conference on Learning Representations. pp. 28085–28128.

Redmon J, and Farhadi A (2018). YOLOv3: an incremental improvement. *arXiv*, 1804.02767.

Renoult JP, Karpinski R, Sauvadet L, Kreyer M, Mbadoumou R, Roura-Torres B, Baniel A, Charpentier MJ. Vision (2025). Vision transformers for age prediction from facial images in a wild primate. *Methods in Ecology and Evolution*,16(12), 2858–2872.

Xintong S, Cao W, and Raschka S (2023). Deep neural networks for rank-consistent ordinal regression based on conditional probabilities. *Pattern Analysis and Applications*, 26(3), 941–955.

Swift RJ, Rodewald AD, Johnson JA, Andres BA, and Senner NR (2020). Seasonal survival and reversible state effects in a long-distance migratory shorebird. *Journal of Animal Ecology*, 89(9), 2043–2055.

Tuia D, Kellenberger B, Beery S, Costelloe BR, Zuffi S, Risse B, Mathis A, Mathis MW, Van Langevelde F, Burghardt T, and Kays R (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*,13(1), 792.

Popko W, and Piersma T (1995). Scoring abdominal profiles to characterize migratory cohorts of shorebirds: an example with Red Knots. *Journal of Field Ornithology*, 66 (1), 88–98.

Yu-Xiang Z, and Lee Y (2025). Bird species classification using image Background removal for data augmentation. *Computers, Materials and Continua*, 84(1), 791–810.

CLLMM RESEARCH CENTRE

GOYDER INSTITUTE FOR WATER RESEARCH