



**HAL**  
open science

## **Deep learning approach to detect and visualize sexual dimorphism in monomorphic species**

Nicolas J. Silva, André C. Ferreira, Liliana R. Silva, Samuel Perret, Sonia Tieo, Julien Renoult, Rita Covas, Claire Doutrelant

► **To cite this version:**

Nicolas J. Silva, André C. Ferreira, Liliana R. Silva, Samuel Perret, Sonia Tieo, et al.. Deep learning approach to detect and visualize sexual dimorphism in monomorphic species. *Animal Behaviour*, 2025, 225, pp.123223. <10.1016/j.anbehav.2025.123223>. <hal-05288535>

**HAL Id: hal-05288535**

**<https://hal.science/hal-05288535v1>**

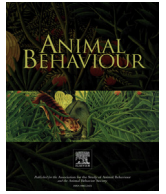
Submitted on 29 Sep 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License



## Methods

# Deep learning approach to detect and visualize sexual dimorphism in monomorphic species



Nicolas J. Silva <sup>a,\*</sup>, André C. Ferreira <sup>a,b,c,d</sup>, Liliana R. Silva <sup>b,c</sup>, Samuel Perret <sup>a</sup>,  
Sonia Tio <sup>a</sup>, Julien P. Renoult <sup>a</sup>, Rita Covas <sup>b,c,e,1</sup>, Claire Doutrelant <sup>a,e,1</sup>

<sup>a</sup> CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>b</sup> CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos, Campus de Vairão, Universidade do Porto, Vairão, 4485-66, Portugal

<sup>c</sup> BIOPOLIS Program in Genomics, Biodiversity and Land Planning, CIBIO, Campus de Vairão, Vairão, 4485-661, Portugal

<sup>d</sup> Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland

<sup>e</sup> FitzPatrick Institute of African Ornithology, DST-NRF Centre of Excellence, University of Cape Town, Rondebosch, 7701, South Africa

## ARTICLE INFO

## Article history:

Received 16 December 2024

Initial acceptance 10 February 2025

Final acceptance 12 March 2025

Available online 30 May 2025

MS. number: 24-00731R

## Keywords:

computer vision  
dimorphism detection  
EfficientNet  
Grad-CAM  
monomorphism

Sex recognition is facilitated by dimorphism in some traits. However, humans often fail to find the traits that allow distinction between sexes in other species. Deep learning has shown great application potential in identifying cryptic differences between sexes, but it has rarely been used to assess sexual dimorphism. In this study, the ability of a fine-tuned classification neural network, which is known as EfficientNet, to find differences between sexes in a species that appears monomorphic to humans, such as the sociable weaver, *Philetairus socius*, was evaluated. In addition, the benefits of the Grad-CAM visualization technique were assessed to understand which parts of the head of the individuals are used by the network to differentiate the sexes. We trained 10-fold cross-validation models on more than 4500 pictures of the head from more than 1300 individuals. Results show that the network can predict the sex of sociable weavers with an accuracy of 76%, which is considerably higher than humans' performance (56%). Moreover, the model was similarly good at predicting females and males. When interpreting the probability of being classified to one sex, our results further reveal the effect of the interaction of sex with age on the confidence score of the models, which shows that younger males are less masculine than older ones, and older females are more masculine than younger ones. Finally, using Grad-CAM, we found that the model mostly used the bill region to predict the sex of individuals. Overall, this work shows the potential application of artificial intelligence as a noninvasive sexing tool, surpassing human capabilities and aiding in pinpointing potential cryptic dimorphic body parts that have yet to be identified. Half of the world's bird species appear sexually monomorphic to humans, and re-evaluation of species dimorphism with this type of methods could deepen our understanding of topics such as sex-specific selection on animal traits, behavioural differences and sex ratio variation across time.

© 2025 The Authors. Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Sexual dimorphism can be present in few or several species' traits. It results from selection on those traits (particularly sexual selection and predation), drift and phylogenetic trajectories (Badyaev & Hill, 2003). Exploring sexual dimorphism in different species is important to determine the mechanism by which evolutionary pressures may vary within and between the sexes. In general, sexual dimorphism is based on differences in size, shape or coloration (Corl et al., 2010; Hedrick & Temeles, 1989; Kuntner &

Coddington, 2020; Mori et al., 2022; Warren et al., 2013), and it is common in taxa such as birds (Andersson, 1982), fish (Kodric-Brown & Brown, 1984), mammals (Caro, 2009) and invertebrates (Allen et al., 2011). Humans usually assess these traits through direct observations of the visual aspect of the individuals (Badyaev & Hill, 2000, 2003). However, their ability to detect dimorphism is limited by the capacity of our visual system, which may differ from that of animals, and only the most conspicuous dimorphism can be reliably identified by human observers. This issue limits our ability to detect different types of dimorphism and to study their adaptive value (Hung et al., 2017; Ó Marcaigh et al., 2021; Villafuerte & Negro, 1998).

\* Corresponding author.

E-mail address: [silva.nicolas.j@gmail.com](mailto:silva.nicolas.j@gmail.com) (N. J. Silva).

<sup>1</sup> Joint last authors.

In recent years, different approaches have been developed to overcome the issue that hinders sexual dimorphism detection. For example, Eaton (2005) used spectrophotometric measurements of reflectance and, considering the colour discriminatory abilities of birds between 300 and 700 nm, showed that more than 90% of the 139 presumed sexually monochromatic bird species from the order Passeriformes were dichromatic from an avian visual perspective. Similar measures were used in a jumping spider, *Phintella vittata*, revealing a UV-B sexual dimorphism in this species (Li et al., 2008). More recently, computer vision has been shown to be effective in measuring coloration. For example, a support vector machine, which is a machine learning tool, was used to classify the sex of zebrafish, *Danio rerio*, based on caudal fin coloration (Hosseini et al., 2019). The support vector machine requires human interference to predefine the features to be learned, thereby identifying and measuring the specific visual features that allow distinction and classification of sexes. Knowing the sexually dimorphic trait before the analyses has the main drawbacks of potentially missing important parts and features of the animals, which might be useful to animals for sex classification but cryptic to humans.

Deep learning methods, particularly convolutional neural networks (CNNs), do not require the predefinition of relevant features, and can overcome this limitation by automatically determining the features that are optimal to perform a given classification task (Angermueller et al., 2016; Chollet, 2021; Christin et al., 2019; Jordan & Mitchell, 2015). At present, CNNs are increasingly used in the fields of ecology and evolution to solve category classification problems (Christin et al., 2019; Ferreira et al., 2020; Hansen et al., 2020; Norman et al., 2023). In particular, CNNs have been previously used in species with known sexually dimorphic traits to increase the processing speed and efficiency of sex recognition; however, to date, CNNs are mostly used in humans (Dwivedi & Singh, 2019). For example, deep learning was used on dentomaxillofacial features extracted from radiographs to determine the sex of children at an early age, when sexual dimorphism is not well marked yet (Franco et al., 2022). Similarly, on human skulls, a deep learning network analysing computed tomography scans of human brains can be utilized to classify women and men with an accuracy of 95% (Bewes et al., 2019). In animals, Chen et al. (2021) showed that deep learning can be used to classify females and males in 30 *Pseudopoda* spider species based on their genitals (Chen et al., 2021). Hosseini et al. (2019) also used CNNs to classify sex based on colour on zebrafish pictures. However, CNNs have been rarely used to find sexual dimorphism in traits that appear monomorphic to the human eye, but deep learning revealed a previously unnoticed sexual dimorphism in cranial shape in grey wolf, *Canis lupus* (MacLeod & Kolska Horwitz, 2020). In *Pseudopoda* spider species, Chen et al. (2021) also used deep learning to identify sex-specific patterns and colours of the back of individuals (which are sexually monomorphic for humans) and found sexual dimorphism in 94% of their data set. Finally, a recent study on giant pandas, *Ailuropoda melanoleuca*, has successfully classified their sexes using CNNs with an accuracy of 77% (Wang et al., 2019).

However, when using CNNs to classify the sex of animals, the features that were automatically built and used by the algorithm to perform classification are unknown (without additional analyses) or are difficult to represent in a way that is meaningful for humans (i.e. CNNs function as 'black boxes'; Alain & Bengio, 2016; Castelvocchi, 2016; Lei et al., 2018; Schwartz-Ziv & Tishby, 2017). In addressing these limitations, tools have been developed to understand what do CNNs learn. In computer vision, Gradient-weighted Class Activation Mapping (Grad-CAM) is a widely used tool (Selvaraju et al., 2017). Grad-CAM creates an activation thermal map for a specific class that is superimposed with the original image. Every channel of the picture (that is, the different

dimensions of an image pixels) is weighted with the class gradient of the channel. These Grad-CAMs colour the pictures to identify which pixels of the images were the most relevant for classification and the possibly most important features to discriminate sex. For example, in giant pandas, Grad-CAMs qualitatively showed sex differences in the eyes and nose of the individuals (Wang et al., 2019).

In this study, we developed a deep learning tool based on EfficientNet (Tan & Le, 2019) and used a monomorphic bird species, which is the sociable weaver, *Philetairus socius*, to explore whether species that appear sexually monomorphic to humans can be detected as sexually dimorphic by CNNs based on pictures (Maclean, 1973). In addition, we explored the use of Grad-CAM in identifying which parts of the individuals are used by the network to differentiate the sexes. In birds, delayed plumage maturation is commonly observed (Hawkins et al., 2012), making young males generally look more like females. Therefore, we tested whether the accuracy of sex identification varied with age, expecting younger males to be accurately classified less often than older males.

## METHODS

### *Species and Data Collection*

We worked with a population of sociable weavers inhabiting the Benfontein Nature Reserve, South Africa (28°52'00"S, 24°51'00"E, 1190 m above sea level), which were monitored since 1993 (Covas et al., 2002). Since 2008, the birds in the study colonies were caught every year, usually at the end of winter (August–September). During capture, a blood sample was drawn and used for genetic sexing (Covas et al., 2006). In addition, top-view pictures of the head of the birds were taken in accordance with an established protocol (Rat et al., 2015).

For most of the birds, breeding monitoring allowed us to know their exact age. However, individuals can move between colonies (especially females; Covas et al., 2006; van Dijk et al., 2015). For immigrant birds first caught as adults in the studied population, they were considered to have the average population dispersion age at capture. This age was calculated by adding the first date of capture to the average minimum age of first dispersion measured in the population (690 days for males and 727 days for females to the date of first capture; Silva et al., 2025).

### *Ethical Note*

Birds were captured and handled in accordance with the protocols approved by the Northern Cape Nature Conservation (permit FAUNA 1638/2015, 0825/2016, 0212/2017, 0684/2019 and 0059/2021) and the Ethics Committee of the University of Cape Town (permit 2014/V1/RC and 2018/V20/RC). The birds were captured using mist nets and were extracted directly after being caught. All efforts were made to minimize handling time of the birds (that is, task division of the processing steps of extracting the birds from the mist nets, ringing the birds, taking blood samples and taking pictures; the release of the birds was closely monitored to detect any uncommon flight behaviour). These procedures were performed during the nonbreeding season to avoid disturbing birds with chicks at the nests. A total of 1323 caught individuals were included in this study.

### *Data Sets*

A total of 4595 images (2246 for females and 2349 for males) depicting 1323 individuals (661 females and 662 males) photographed during six annual capture events (between August 2015

and September 2021; no captures in 2020) were obtained in this study. Only individuals with a fully developed adult plumage were selected. At each capture event, one to four pictures per individual were collected (females:  $2.63 \pm 0.81$  (mean  $\pm$  SD); males:  $2.46 \pm 0.88$  (mean  $\pm$  SD)). A Mask-RCNN model was trained and used to extract the birds' heads from full pictures (see Supplementary Material section 1 for the description of the model). All pictures were orientated with the throat of the individuals facing the bottom of the pictures, in a portrait style (Fig. 1a). Of the 1323 individuals included in our data set, 993 were assigned to their real age ( $773 \pm 560$  days old (mean  $\pm$  SD)) and 330 to an estimated age ( $1372 \pm 924$  days old (mean  $\pm$  SD)).

### Model Training

We used transfer learning from an EfficientNet B4 network (Tan & Le, 2019) pretrained on the ImageNet database (Deng et al., 2009) composed of more than 14 million pictures to classify objects from images. This method allowed us to use the weights of an already trained model and fine tune them for a particular task. Considering that our objective is also image classification (that is, classifying images as males or females), this pretrained model on the ImageNet network is an appropriate choice to perform transfer learning. During the fine tuning of the network, we implemented average pooling throughout the model's backbone to prevent overfitting during training. We replaced the fully connected part at the end of the EfficientNet B4 network with a dropout layer (rate: 0.4), a dense layer with 1024 neurons and a ReLU activation, a dense layer with 64 neurons and a ReLU activation and a final dense layer with one neuron and a sigmoid activation. The dropout layer was added to limit overfitting of the model by randomly setting to zero some neurons of the network (Srivastava et al., 2014). We compiled the model with an Adam optimizer (Kingma & Ba, 2014), a binary cross-entropy loss and a binary accuracy. All layers were set as trainable except for the BatchNormalisation layers. Considering the sigmoid last layer, the model outputs were numbered between zero and one, with  $[0; 0.5[$  being the sex prediction scores of females and  $[0.5; 1]$  being the sex prediction scores of males. For a better comparison between males and females, we reversed the score of females (one minus the predicted score). Therefore, for all individuals, a final score between  $[0; 0.5[$  signal indicated an erroneous sex prediction, and a score between  $[0.5; 1]$  indicated a correct sex prediction.

The hyperparameters of the training process were inferred by training a preliminary model using 90% of the individuals (595 males, 596 females and 4183 pictures) as the training data set and 10% as the validation data set (66 males, 66 females and 412

pictures). This preliminary model helped us to define the training parameters. The pictures were resized to  $400 \times 400$  pixels and passed through the network with a batch size of 16 pictures. Training was performed on 13 epochs with a learning rate of  $10^{-3}$  from epoch one to three to quickly learn the general patterns for sex classification. Then, the learning rate was set to  $10^{-4}$  to help the model to learn more precise features.

Next, 10-fold cross-validation was performed (90% of the individuals as the training data set and 10% of the individuals as the test data set). At each fold, individuals from the test data set were swapped with an equivalent number of individuals from the training data set. This process allowed each individual to be in a test set once during training and to get one prediction of sex per image. Data augmentation was also performed as a random horizontal flip and a random contrast modification between  $[-0.2; 0.2]$ , both with a probability of 0.5. In addition, the impact of reducing by a quarter, half and three quarters the number of individuals in the training data set on the training process was evaluated (Table S1).

### Model Visualization

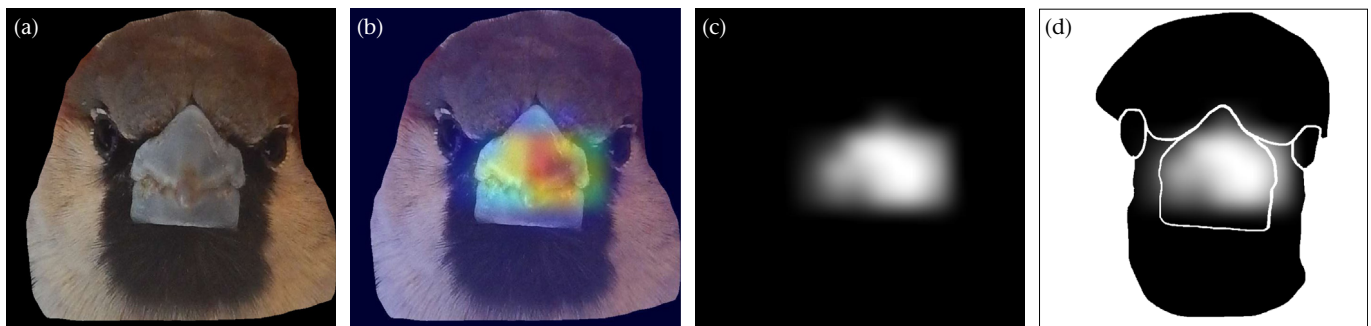
We used Grad-CAM (Selvaraju et al., 2017) to identify the specific areas of the images, which influenced the model's decision in classifying individuals as male or female. The Grad-CAM of the last 2D convolution layer (called 'top\_activation') was computed to determine the probability of being a male (considering that a sigmoid activation was used for the last layer). The different regions of the pictures were coloured from blue to red, indicating lower to higher importance, depending on the role they played in the network's decision (Fig. 1b). Analysis for this part was adapted from Chollet (2021).

### Statistical Analysis

All statistical analyses were conducted on R version 4.3.1.

### Sex recognition

The average accuracy of the models on the test data sets was saved ( $N = 10$  models) at the end of each fold during training. These accuracies were compared with those reached by human observers. In evaluating human performance, seven of the authors (A.C.F, L.R.S., S.P., S.T., J.P.R., R.C. and C.D.) were first given the same set of labelled pictures of 25 females and 25 males as an attempt to train themselves, for as long as they wanted, to distinguish between males and females. Then, they were given 100 random pictures without sex being annotated (50 males and 50 females),



**Figure 1.** Workflow of picture processing along with training and visualization. (a) The picture is segmented from a bigger picture using the first neural network to keep only the head of the individuals with their bib facing down. (b) The blue-to-red scale Grad-CAM superimposed with transparency to the original picture allows us to visualize which regions of an image the model is using to perform its classification. (c) Grey-scale Grad-CAM. (d) The cap, bill, bib and eyes are exported on their own and analysed separately.

and they had to determine the sex of the individuals. We performed a first Wilcoxon test to compare the accuracies of machines and humans.

In addition, we simulated a random sex classification of the 100 birds classified by humans using a binomial distribution with a probability of 0.5. Using Wilcoxon signed rank exact tests, we tested whether the accuracy reached by the CNNs and the one reached by humans were different from a random classification.

#### *Visualization of the features used for classification*

We quantitatively measured the parts of the pictures that were used by the models to classify males and females. For each of the 10-fold training, we randomly sampled, from the test data set, 10 pictures of males and 10 pictures of females that were correctly classified (that is, a predicted score by the deep learning model below 0.5 for females and above 0.5 for males). We manually segmented these 200 pictures to identify the four parts of interest on the head of the birds, namely the cap, bill, bib and eyes (Fig. 1d). We also computed Grad-CAM images with a grey-scale coloration (pure black indicates no activation, and pure white indicates very high activation; Fig. 1c). Based on these images, we measured (1) the proportion of activated pixels to estimate the surface proportion of each of the four regions used by the model for sex classification and (2) the average activation of the activated pixels from ] 0; 1] (one indicates pure white) using GIMP version 2.10.36. The first approach identified the regions of the head that are most used by the model by considering its surface, whereas the second approach determined the importance of each activated region among the activated pixels in the pictures. Both measurements are useful in identifying dimorphism because the first measurement determines how much of a region is used for classification, and the second measurement identifies the intensity of a region. For example, by relying on both measurements, we could distinguish the regions with a large proportion of pixels activated at a low intensity from regions with only a small portion of its surface activated but with a very high intensity. Given the sigmoid last layer, we predicted the probability for an individual to be a male. Therefore, a high difference in activation between males and females indicates that a trait is highly dimorphic. On the contrary, if the activation for a trait is similar between males and females, then the trait is not dimorphic.

We performed multiple Student *t* tests to determine whether the proportion of activated pixels per regions and the average activation were different between sexes, as well as to understand whether these two measures were different between regions within the same individual. *P* values were adjusted using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995).

#### *Effect of sex and age difference on classification performance*

We tested whether males and females differed in their likelihood of being correctly identified and whether age influenced the classification performance by running two mixed models.

First, we performed a generalized linear mixed model (GLMM; binomial error distribution with a link logit). As a response variable, the average predicted score per individual was binarized, with 1 indicating correctly classified individuals and 0 indicating wrongly classified ones. As fixed factors, sex was used, and the random factors identity of the individuals, cross-validation fold and year of data collection were added. Several pictures per individuals were used, which are not independent points in the model. In addition, the predicted score within a fold is not independent within the test data sets because each test data set of the cross-validation folds is processed using a different model during each iteration of the cross-validation. Furthermore, different people between years were

involved in the picture collection of the birds, so we controlled for intra- and interphotographer variation.

Second, we fitted a GLMM (beta error distribution with a link logit). The score predicted by the model (that is, the confidence score given by the deep learning model) was regarded as the response variable. This score indicates how confident the model is about its classification decision. A score closer to 1 indicates that the sex classification is relatively robust. Sex, age and their interaction were used as fixed factors to account for the expectation of males becoming more masculine with age, which is an effect not anticipated for females. As random factors, the identity of the individuals, the cross-validation fold and the year of data collection were included. A total of 27 pictures (18 males and two females,  $1009 \pm 825$  days old (mean  $\pm$  SD)) were predicted with a confident score of exactly 1. However, considering that the beta distribution only accepts continuous values in the range ]0; 1[, we subtracted a value of  $10^{-10}$  for these pictures to be included in the model. Age was scaled and mean centred. Considering that age is either the true age or the estimated age for immigrants, the model was initially computed with the individuals with the exact known age only (992 individuals, 3445 pictures) and then with all the individuals (1323 individuals, 4595 pictures). The two models were the same, except for the effect of age in females, which tends to be negative in the model that includes all individuals and is negative in the model including only individuals with exact age (Table S2). Therefore, we decided to be conservative and keep only the individuals with known exact age.

Mixed models were computed using lme4 package version 1.1–34 (Bates et al., 2014) and glmmTMB package version 1.1.8 (Brooks et al., 2017) for the GLMM Beta. Model plots were created using ggplot2 package version 3.4.2 (Wickham, 2014), ggeffects version 1.3.1 (D. Lüdtke, 2018) and sjPlot package version 2.8.15 (D. M. Lüdtke, 2023).

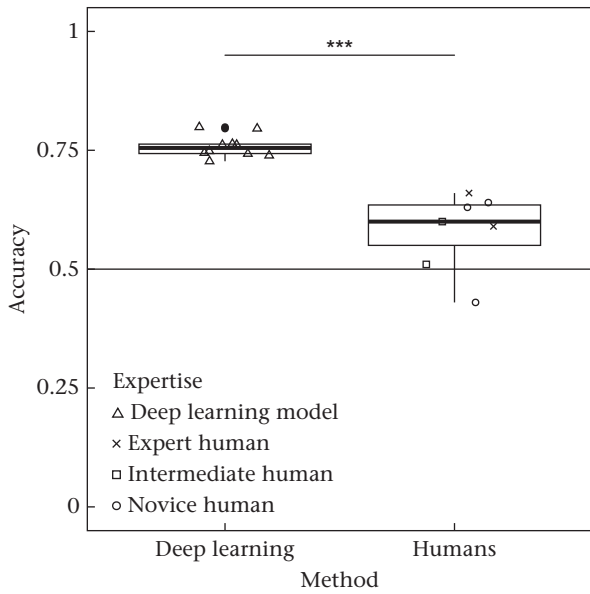
## RESULTS

### *Sex Recognition*

On average, we found that the models can classify males and females with an average accuracy of  $75.8\% \pm 2.4\%$  (mean  $\pm$  SD) at the picture level within the 10 folds (Fig. 2), indicating that these classifications were different from a random guess (i.e. accuracy = 0.5; Wilcoxon test,  $V = 0$ ,  $P < 0.001$ ). Human classifications were also found to be different from a random guess (Wilcoxon test,  $V = 0$ ,  $P < 0.001$ ), but they were less accurate  $58.0\% \pm 8.2\%$  ( $N = 7$  humans; Fig. 2) than the performance of deep learning models (Wilcoxon test,  $W = 70$ ,  $P < 0.001$ ; Fig. 2). These results indicate that deep learning can classify the sex of sociable weavers with higher accuracy than humans.

### *Visualization of the Features Used for Classification*

Males presented a higher proportion of activated pixels and a higher intensity in the activated pixels than females for the four regions delimited in the head of the birds (bill, bib, cap and eyes; Fig. 3a and b, Table S3). Among these four regions, the highest proportion of activated pixels and the highest intensity in the activated pixels were observed in the bill for males and in the bib for females (Table S4). This finding indicates that the four regions are used by the model to differentiate males from females, but the model focuses more on the bill of the individuals to perform its classification. The fact that the bib is also activated in females may indicate that the bib is a source of misclassification and that few sex differences can be found on that trait.



**Figure 2.** Accuracy of the individual classification in males and females in function of the classification method used. Deep learning accuracies were obtained from the predictions on the test data set at each of the 10 folds during training. Human accuracies were obtained from blind visual classification by humans. The classification methods were found to be significantly different from each other and from random classification. Triangles represent the accuracies obtained using deep learning models, and cross, rectangles and circles represent the accuracies obtained by the expert, intermediate and novice experimenters, respectively.

*Effect of Sex and Age Difference on Classification Performance*

No evidence as to whether sex has an effect on the probability to correctly classify a picture of an individual was found (estimate ± SE (95% CI):  $-0.19 \pm 0.17$  ( $-0.53; 0.14$ ),  $z = -1.12$ ,  $P = 0.260$ ; Table 1), indicating that no sex is better classified than the other one.

**Table 1**

Estimate of the GLMM parameters to test the effect of sex on the probability for the deep learning network to correctly classify sex from the pictures of individuals' head

Fixed effect	Estimate	SE	95% CI	Z	P
Intercept	2.28	0.15	1.99 ; 2.57	15.27	<0.001
Sex-M	-0.19	0.17	-0.53; 0.14	-1.12	0.261
Random effect					
Individual (N = 1323)	5.591				
Season (N = 6)	0.000				
Cross-validation fold (N = 10)	0.000				
Residuals	3.290				

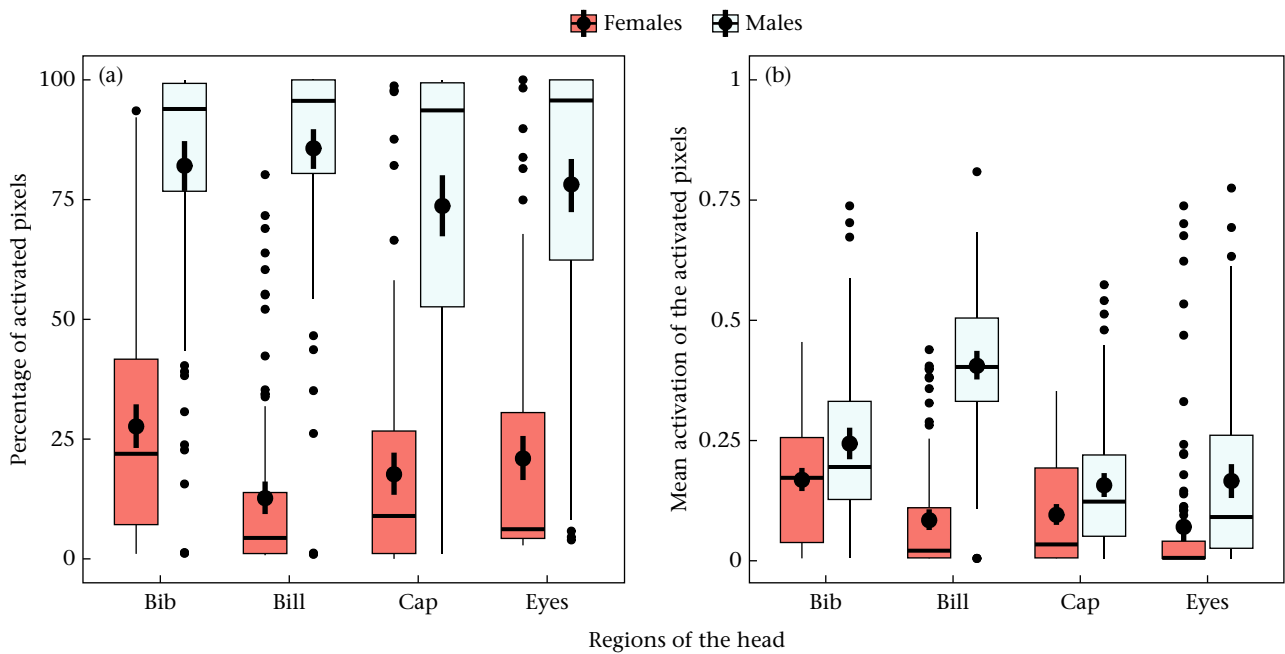
The model was performed on 4593 pictures, with individuals with the exact age and the estimated minimum age.

However, we found strong evidence for an association between the predicted score and age that differed between the sexes (estimate ± SE (95% CI):  $0.43 \pm 0.07$  ( $0.30; 0.56$ ),  $z = 6.47$ ,  $P < 0.001$ ; Fig. S1; Table S2 Model B), with a positive link for males and a negative link for females (Fig. 4, Table S2 Model B). Therefore, for males, younger individuals are less well classified than older ones, whereas it is the reverse for females.

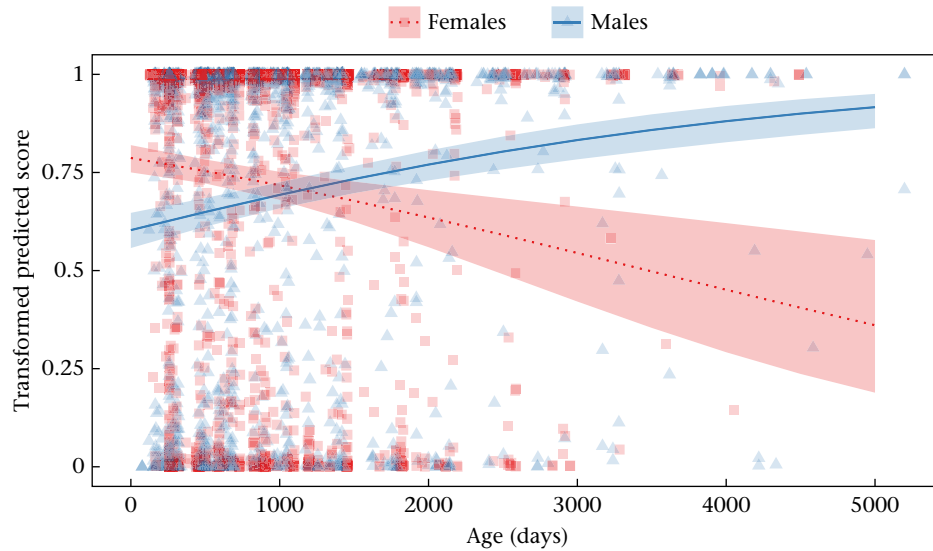
**DISCUSSION**

Our results indicate that a deep learning-based method surpasses human performance when classifying the sex in a bird species with limited sexual dimorphism. In addition, by using Grad-CAM, we can identify the differences used by the algorithm to differentiate between males and females. Finally, by investigating the source of misclassifications of the models, we found that age influenced differently females and males when predicting sex using CNNs.

Here, we reached only 76% accuracy, which is similar to the result obtained with giant panda faces (Wang et al., 2019) but insufficient for most studies in behavioural ecology and evolution.



**Figure 3.** Differences in the (a) percentage of activated pixels and (b) mean activation of the activated pixels, in function of sex and region of the head. Females are shown in red, and males are shown in light blue. The bill of the individuals is the main feature used by the models to perform its sex classification (higher percentage of activated pixels and higher activation intensity than females). Box plots represent the median and quantiles at 25% and 75%, respectively, and error bars represent the mean and 95% confidence intervals.



**Figure 4.** Score predicted by the deep learning model in function of age and sex. Females are represented in red with dashed line and squares, and males are represented in blue with solid line and triangles. The squares and triangles are the raw data used to fit the mixed model. The lines correspond to the estimated means that were obtained using the mixed model, and the shaded area corresponds to the 95% confidence intervals. Older males are better predicted than younger males, whereas the opposite is found for females. Female predicted scores were transformed in which 1 corresponds to the best predictions for both sexes.

However, in the present study, we only used a small portion of the bird's body from a single angle (head facing forward). In future work, using photos from multiple body parts of the birds will increase the model's performance. Notably, the images were taken approximately 1 month before the breeding season. In some bird species, plumage variation and dimorphism are minimized outside the breeding season. Therefore, collecting a similar data set during the breeding season is necessary to rule out the possibility that sexual dimorphism in this species may vary within the season and to reveal which traits could be relevant for individuals to become breeders (using Grad-CAMs). In addition, increasing the training sample size and including more pictures of more individuals taken from different ages can improve model accuracy. Based on our results (Table S1), the efficiency of our trained network decreased with the decrease of the training data set sample size, indicating that the full potential of using CNNs for sex classification in our species has not been reached. Here, standardized pictures were selected to train and test our models, and further tests could be performed with pictures taken under more natural conditions (e.g. at bird feeders; Ferreira et al., 2020). Hence, we suggest that in this and other monomorphic species, adequate training data sets and photos from different parts of the body will enable sexual recognition of individuals using deep learning models. Consequently, such methods could achieve sufficient performance to replace genetic sexing, which, in the long run, could save time and research funds. Furthermore, these methods could limit animal suffering, as pictures can be taken without capturing individuals (Ferreira et al., 2020), and could avoid capture and handling stress. Images of animals in the wild are relatively easy to collect, and an increasing number of studies use camera traps to detect animal presence, habitat use and behaviour. However, most individuals observed in these images and videos are unmarked, and for species that are monomorphic to the human eye, obtaining information on basic features such as the sex of the individuals is not feasible. The utilization of artificial intelligence to detect dimorphism that is not detectable by humans will address various biological questions. For example, artificial intelligence will enable a sex ratio estimation in groups of unmarked monomorphic species, thereby exploring the consequences of this variation on behavioural and evolutionary

questions (Grant & Grant, 2019). Other examples would include identifying previously unnoticed sex differences in specific behaviours. For example, the sex that contributes most to nest building in a large proportion of species is currently unknown (Healy et al., 2023). Similar questions can be asked about other behaviours such as territory defence, mobbing and group composition. In the case of the sociable weaver, this method will be crucial to study infanticide, which is a rare but remarkable behaviour in this species, prominently performed by migrant noncaptured individuals for which we have no information on sex, only image records (Covas et al., n.d.).

Using Grad-CAMs, we found a sexual dimorphism on the bill of the individuals. No sexual differences on the head of the individuals were determined on the basis of morphological measurements or field observations, and the sexes appeared indistinguishable (Covas & Doutrelant, unpublished data; Maclean, 1973). Sexual dimorphism based on the bill of individuals have been described in other weaver species such as red-billed quelea, *Quelea quelea* (Walsh et al., 2012), red-billed buffalo weaver, *Bubalornis niger* (Maclean, 1993), African village weaverbird, *Ploceus cucullatus* (Collias & Collias, 1970) and white-browed sparrow-weaver, *Plocepasser mahali* (Leitner et al., 2009). However, the dimorphism described for these species based on bill length and coloration is visible using measurements, which is not the case for the sociable weavers. Thus, further investigation is necessary to identify the features that are used by the model to classify sex (e.g. shapes, colours and contrasts). In addition, other measures (e.g. morphological information) and alternative sensory views (e.g. UV images) should be investigated to help fine tune the existing model and allow better classification, especially for species that use UV to communicate. Here, the method used indicates a potentially sexually dimorphic part of the bird, which could help us in the future to implement measures in the field to quantify this dimorphism.

The mechanisms used by deep learning models to identify the critical features is usually treated as a 'black box' (Alain & Bengio, 2016; Castellevecchi, 2016; Lei et al., 2018; Schwartz-Ziv & Tishby, 2017). Grad-CAMs are the first steps to address this issue and understand the potential cryptic differences between categories. In dimorphic species, many studies have investigated the potential

adaptive value of conspicuous traits (e.g. coloration and size), linking them to fitness benefits (e.g. reproductive success and mating success). Our results highlight the potential to investigate similar questions on dimorphic traits that are cryptic to the human eye but can be detected by using a deep learning neural network, which could be under the same evolutionary pressures as conspicuous traits. However, whether birds use the bill to infer the sex of their conspecifics remains unknown. Side-view images of the birds are currently being taken, and future work would investigate bill dimorphism using these new images.

When dealing with classification problems, particularly in the field of computer science, performance metrics should be maximized (e.g. to obtain nearly 100% accuracy). However, in the field of biological sciences, the inefficiency of neural networks could indicate cryptic features of animals (or organs, cells, etc.), which might open the door to new, biologically relevant hypotheses. For example, deep learning was used in mandrills, *Mandrillus sphinx*, to assess face similarity of individuals, revealing higher similarity among kin and allowing the exploration of spatial associations among similar-looking individuals (Charpentier et al., 2022). Here, we took advantage of the imperfection of our models regarding sex determination to reveal that young males are less masculine than older ones and that older females look more masculine than younger ones, which has already been observed in primates (Kloth et al., 2015; Tiew et al., 2023). However, in birds, only a lower masculinity of younger males has been regularly observed with notably delayed maturation in plumage (Hawkins et al., 2012). To our knowledge, no quantitative description of female masculinization with age has been reported in a nonprimate species, although some descriptions exist. For example, older females gang-gang cockatoos, *Callocephalon fimbriatum*, can acquire orange coloration in the head, which is typical of males (Forshaw & Cooper, 2002). This finding is not surprising, particularly in apparently monomorphic species, where humans cannot directly obtain information about which regions of the individuals are attributed to a masculine or a feminine appearance, thereby hindering the investigation of how they change with age. In this study, for very old female sociable weavers, the relationship between the predicted score and age could arise from the male-biased sample size in the training data set (0 female and 21 males older than 3000 days; Fig. S2). However, removing these old individuals from the data set used for the mixed models did not change the observed relationship for younger individuals, except for a slight change in effect size (Table S5). Therefore, despite the unbalanced sex ratio of very old individuals, we can still be confident, for most individuals, of the relationship between the predicted score and age in relation to sex that we found. We ran an additional multi-input deep-learning model to account for age in the classification, but this approach did not improve the accuracy of the models (see Supplementary Material section 6). This limitation is due to the skewed distribution of age in our data set towards younger individuals (Fig. S3). Our approach and results indicate that similar to primates (Charpentier et al., 2022; Tiew et al., 2023), deep learning could be used in birds to study the mechanism by which sex and age (or other relevant traits) interact and affect the visual aspect of individuals. In addition, potential features expressing age may be present in the head of the individuals (e.g. age classification of giant pandas; Qi et al., 2022). On the basis of these results, the potential adaptive value of such interactions can be addressed, and the age of the individuals in the wild can be estimated.

## Conclusion

Our study shows that deep learning methods can surpass human abilities in detecting sexual dimorphism in species, paving the

way to explore broader research question such as those requiring information on population sex ratios or exploring hitherto unknown sexual differences in behaviours. Such methods also allowed us to identify the bill as a trait potentially indicating the sex of our study species. This approach will stimulate interest among researchers to reassess sexual monomorphism in their study species and how it varies in relation to individual attributes. Sexual dimorphism is a primary trait used in sexual selection studies, and a good quantification of species dimorphism could influence our understanding of the selective forces acting in each focal species, thereby uncovering new avenues for research.

## Author Contributions

**Nicolas J. Silva:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **André C. Ferreira:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Liliana R. Silva:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization. **Samuel Perret:** Writing – review & editing, Investigation, Data curation. **Sonia Tiew:** Writing – review & editing, Methodology. **Julien P. Renoult:** Writing – review & editing, Methodology. **Rita Covas:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization. **Claire Doutrelant:** Writing – review & editing, Project administration, Investigation, Funding acquisition, Conceptualization.

## Data Availability

Data and scripts to reproduce the analyses are available on the Figshare repository: <https://doi.org/10.6084/m9.figshare.28574504>.

## Declaration of Interest

The authors declare no conflicts of interest.

## Acknowledgments

We thank all the people that helped in field data collection along the years, especially Annick Lucas and Margaux Rat. We also thank Franck Théron for managing the long-term database. Access to Benfontein Nature Reserve was provided by De Beers Mining Corporation. This study was supported by funding from ANR to C.D. (France, grant 19-CE02-0014-01), ERC to R.C. (EU, Consolidator grant 866489) and by the DST-NRF Centre of Excellence at the Fitzpatrick Institute of African Ornithology University of Cape Town. N.J.S. was funded by ANR 19-CE02-0014-01, L.R.S by ERC-Consolidator 866489, R.C. by FCT (CEECIND/03451/2018) and A.C.F by University of Zurich Forschungskredit postdoc grant (K-74312-01-01 University of Zurich), Swiss Federal Commission for Scholarships and European Research Council (grant agreement no. 850859 awarded to Damien Farine). This project is part of the OSU OREME, and long-term Studies in Ecology and Evolution (SEE-Life) program of the CNRS.

## Supplementary Material

Supplementary material associated with this article is available at <https://doi.org/10.1016/j.anbehav.2025.123223>.

## References

Alain, G., & Bengio, Y. (2016). Understanding intermediate layers using linear classifier probes. *ArXiv Preprint ArXiv:1610.01644*.

- Allen, C. E., Zwaan, B. J., & Brakefield, P. M. (2011). Evolution of sexual dimorphism in the lepidoptera. *Annual Review of Entomology*, 56, 445–464. <https://doi.org/10.1146/annurev-ento-120709-144828>
- Andersson, M. (1982). Sexual selection, natural selection and quality advertisement. *Biological Journal of the Linnean Society*, 17(4), 375–393. <https://doi.org/10.1111/j.1095-8312.1982.tb02028.x>
- Angermueller, C., Pärnamaa, T., Parts, L., & Stegle, O. (2016). Deep learning for computational biology. *Molecular Systems Biology*, 12(7), 1–16. <https://doi.org/10.15252/msb.20156651>
- Badyaev, A. V., & Hill, G. E. (2000). Evolution of sexual dichromatism: Contribution of carotenoid-versus melanin-based coloration. *Biological Journal of the Linnean Society*, 69(2), 153–172. <https://doi.org/10.1006/bijl.1999.0350>
- Badyaev, A. V., & Hill, G. E. (2003). Avian sexual dichromatism in relation to phylogeny and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 34(1), 27–49. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132441>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bewes, J., Low, A., Morphett, A., Pate, F. D., & Henneberg, M. (2019). Artificial intelligence for sex determination of skeletal remains: Application of a deep learning artificial neural network to human skulls. *Journal of Forensic and Legal Medicine*, 62(December 2018), 40–43. <https://doi.org/10.1016/j.jflm.2019.01.004>
- Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal*, 9(2), 378–400. <https://doi.org/10.32614/rj-2017-066>
- Caro, T. (2009). Contrasting coloration in terrestrial mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1516), 537–548. <https://doi.org/10.1098/rstb.2008.0221>
- Castelvocchi, D. (2016). Can we open the black box of AI? *Nature News*, 538(7623), 20–23. <https://doi.org/10.1038/538020a>
- Charpentier, M. J. E., Poirotte, C., Roura-Torres, B., Amblard-Rambert, P., Willaume, E., Kappeler, P. M., Rousset, F., & Renoult, J. P. (2022). Mandrill mothers associate with infants who look like their own offspring using phenotype matching. *Elife*, 11, 1–25. <https://doi.org/10.7554/eLife.79417>
- Chen, Q., Ding, Y., Liu, C., Liu, J., & He, T. (2021). Research on spider sex recognition from images based on deep learning. *IEEE Access*, 9, 120985–120995. <https://doi.org/10.1109/ACCESS.2021.3109120>
- Chollet, F. (2021). *Deep learning with Python*. Simon & Schuster.
- Christin, S., Hervet, E., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632–1644. <https://doi.org/10.1111/2041-210X.13256>
- Collias, N. E., & Collias, E. C. (1970). The behaviour of the west African village weaverbird. *Ibis*, 112(4), 457–480. <https://doi.org/10.1111/j.1474-919X.1970.tb00818.x>
- Corl, A., Davis, A. R., Kuchta, S. R., Comendant, T., & Sinervo, B. (2010). Alternative mating strategies and the evolution of sexual size dimorphism in the side-blotched lizard, *Uta stansburiana*: A population-level comparative analysis. *Evolution*, 64(1), 79–96. <https://doi.org/10.1111/j.1558-5646.2009.00791.x>
- Covas, R., Brown, C. R., Anderson, M. D., & Brown, M. B. (2002). Stabilizing selection on body mass in the sociable weaver *Philetairus socius*. *Proceedings of the Royal Society B: Biological Sciences*, 269(1503), 1905–1909. <https://doi.org/10.1098/rspb.2002.2106>
- Covas, R., Dalecky, A., Caizergues, A., & Doutrelant, C. (2006). Kin associations and direct vs indirect fitness benefits in colonial cooperatively breeding sociable weavers *Philetairus socius*. *Behavioral Ecology and Sociobiology*, 60(3), 323–331. <https://doi.org/10.1007/s00265-006-0168-2>
- Covas, R., Silva, L. R., Ferreira, A. C., D'Amelio, P. B., Fortuna, R., Duval, D., Paquet, M., & Doutrelant, C. (n.d.). Infanticide in a colonial cooperative bird is not motivated by direct reproductive benefits. In preparation.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., & Li, K. (2009). ImageNet: Constructing a large-scale image database. *Journal of Vision*, 9(8). <https://doi.org/10.1167/9.8.1037>, 1037–1037.
- Dwivedi, N., & Singh, D. K. (2019). Review of deep learning techniques for gender classification in images. In *Harmony search and nature inspired optimization algorithms: Theory and applications, ICHSA 2018* (pp. 1089–1099). Singapore: Springer. [https://doi.org/10.1007/978-981-13-0761-4\\_102](https://doi.org/10.1007/978-981-13-0761-4_102)
- Eaton, M. D. (2005). Human vision fails to distinguish widespread sexual dichromatism among sexually “monochromatic” birds. *Proceedings of the National Academy of Sciences of the United States of America*, 102(31), 10942–10946. <https://doi.org/10.1073/pnas.0501891102>
- Ferreira, A. C., Silva, L. R., Renna, F., Brandl, H. B., Renoult, J. P., Farine, D. R., Covas, R., & Doutrelant, C. (2020). Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution*, 11(9), 1072–1085. <https://doi.org/10.1111/2041-210X.13436>
- Forshaw, J. M., & Cooper, W. T. (2002). *Australian Parrots*. Melbourne: Lansdowne.
- Franco, A., Porto, L., Heng, D., Murray, J., Lygate, A., Franco, R., Bueno, J., Sobania, M., Costa, M. M., Paranhos, L. R., Manica, S., & Abade, A. (2022). Diagnostic performance of convolutional neural networks for dental sexual dimorphism. *Scientific Reports*, 12(1), 1–12. <https://doi.org/10.1038/s41598-022-21294-1>
- Grant, P. R., & Grant, B. R. (2019). Adult sex ratio influences mate choice in Darwin's finches. *Proceedings of the National Academy of Sciences*, 116(25), 12373–12382. <https://doi.org/10.1073/pnas.1903838116>
- Hansen, O. L. P., Svenning, J. C., Olsen, K., Dupont, S., Garner, B. H., Iosifidis, A., Price, B. W., & Høye, T. T. (2020). Species-level image classification with convolutional neural network enables insect identification from habitus images. *Ecology and Evolution*, 10(2), 737–747. <https://doi.org/10.1002/ece3.5921>
- Hawkins, G. L., Hill, G. E., & Mercadante, A. (2012). Delayed plumage maturation and delayed reproductive investment in birds. *Biological Reviews*, 87(2), 257–274. <https://doi.org/10.1111/j.1469-185X.2011.00193.x>
- Healy, S. D., Tello-Ramos, M. C., & Hébert, M. (2023). Bird nest building: Visions for the future. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1884), Article 20220157. <https://doi.org/10.1098/rstb.2022.0157>
- Hedrick, A. V., & Temeles, E. J. (1989). The evolution of sexual dimorphism in animals: Hypotheses and tests. *Trends in Ecology & Evolution*, 4(5), 136–138. [https://doi.org/10.1016/0169-5347\(89\)90212-7](https://doi.org/10.1016/0169-5347(89)90212-7)
- Hosseini, S., Simianer, H., Tetens, J., Brenig, B., Herzog, S., & Sharifi, A. R. (2019). Efficient phenotypic sex classification of zebrafish using machine learning methods. *Ecology and Evolution*, 9(23), 13332–13343. <https://doi.org/10.1002/ece3.5788>
- Hung, H. Y., Yeung, C. K. L., Omland, K. E., Yao, C. T., Yao, C. J., & Li, S. H. (2017). Himalayan black bulbuls (*Hypsipetes leucocephalus nigerimus*) exhibit sexual dichromatism under ultraviolet light that is invisible to the human eye. *Scientific Reports*, 7(October 2016), 1–9. <https://doi.org/10.1038/srep43707>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kingma, D. P., & Ba, J. L. (2014). Adam: A method for stochastic optimization. *ArXiv Preprint ArXiv:1412.6980*.
- Kloth, N., Damm, M., Schweinberger, S. R., & Wiese, H. (2015). Aging affects sex categorization of male and female faces in opposite ways. *Acta Psychologica*, 158, 78–86. <https://doi.org/10.1016/j.actpsy.2015.04.005>
- Kodric-Brown, A., & Brown, J. H. (1984). Truth in advertising: The kinds of traits favored by sexual selection. *American Naturalist*, 124(3), 309–323. <https://doi.org/10.1086/284275>
- Kuntner, M., & Coddington, J. A. (2020). Sexual size dimorphism: Evolution and perils of extreme phenotypes in spiders. *Annual Review of Entomology*, 65(1), 57–80. <https://doi.org/10.1146/annurev-ento-011019-025032>
- Lüdecke, D. (2018). ggeffects: Tidy data frames of marginal effects from regression models. *Journal of Open Source Software*, 3(26), 772. <https://doi.org/10.21105/joss.00772>
- Lüdecke, D. M. (2023). Package ‘sjPlot’ (issue 1). <https://CRAN.R-project.org/package=sjPlot>.
- Lei, D., Chen, X., & Zhao, J. (2018). Opening the black box of deep learning. *ArXiv Preprint ArXiv:1805.08355*, 1–27. <http://arxiv.org/abs/1805.08355>.
- Leitner, S., Mundy, P., & Voigt, C. (2009). Morphometrics of white-browed sparrow-weavers *Plocepasser mahali* in south-western Zimbabwe. *Ostrich*, 80(2), 99–102. <https://doi.org/10.2989/OSTRICH.2009.80.2.6.833>
- Li, J., Lim, M. L. M., Zhang, Z., Liu, Q., Liu, F., Chen, J., & Li, D. (2008). Sexual dichromatism and male colour morph in ultraviolet-B reflectance in two populations of the jumping spider *Phintella vittata* (Araneae: Salticidae) from tropical China. *Biological Journal of the Linnean Society*, 94(1), 7–20. <https://doi.org/10.1111/j.1095-8312.2008.00968.x>
- Maclean, G. L. (1973). The sociable weaver, part 1: Description, distribution, dispersion and populations. *Ostrich*, 44(3–4), 176–190. <https://doi.org/10.1080/00306525.1973.9639158>
- Maclean, G. L. (1993). Roberts birds of southern Africa. In *John Voelcker Bird Book Fund, Cape Town*.
- MacLeod, N., & Kolska Horwitz, L. (2020). Machine-learning strategies for testing patterns of morphological variation in small samples: Sexual dimorphism in gray wolf (*Canis lupus*) crania. *BMC Biology*, 18(1), 1–26. <https://doi.org/10.1186/s12915-020-00832-1>
- Mori, E., Mazza, G., & Lovari, S. (2022). Sexual dimorphism. *Encyclopedia of Animal Cognition and Behavior*, 6389–6395. <https://doi.org/10.1007/978-3-319-47829-6>
- Norman, D. L., Bischoff, P. H., Wearn, O. R., Ewers, R. M., Rowcliffe, J. M., Evans, B., Sethi, S., Chapman, P. M., & Freeman, R. (2023). Can CNN-based species classification generalise across variation in habitat within a camera trap survey? *Methods in Ecology and Evolution*, 14(1), 242–251. <https://doi.org/10.1111/2041-210X.14031>
- Ó Marcaigh, F., Kelly, D. J., Analuddin, K., Karya, A., Lawless, N., & Marples, N. M. (2021). Cryptic sexual dimorphism reveals differing selection pressures on continental islands. *Biotropica*, 53(1), 121–129. <https://doi.org/10.1111/btp.12852>
- Qi, Y., Su, H., Hou, R., Zang, H., Liu, P., He, M., Xu, P., Zhang, Z., & Chen, P. (2022). Giant panda age recognition based on a facial image deep learning system. *Ecology and Evolution*, 12(12), Article e9507. <https://doi.org/10.1002/ece3.9507>
- Rat, M., van Dijk, R. E., Covas, R., & Doutrelant, C. (2015). Dominance hierarchies and associated signalling in a cooperative passerine. *Behavioral Ecology and Sociobiology*, 69(3), 437–448. <https://doi.org/10.1007/s00265-014-1856-y>
- Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 618–626). <http://arxiv.org/abs/1610.02391>.
- Shwartz-Ziv, R., & Tishby, N. (2017). Opening the black box of deep neural networks via information. *ArXiv Preprint ArXiv:1703.00810*, 1–19. <http://arxiv.org/abs/1703.00810>.

- Silva, N. J., Silva, L. R., D'Amelio, P. B., Ferreira, A. C., Lucas, A., Covas, R., & Doutrelant, C. (2025). Long-term evaluation of the association between dominance, bib size, sex and age. *bioRxiv*, 2025-02. <https://doi.org/10.1101/2025.02.13.638048>
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning* (pp. 6105–6114).
- Tieo, S., Dezeure, J., Cryer, A., Lepou, P., Charpentier, M. J. E., & Renoult, J. P. (2023). Social and sexual consequences of facial femininity in a non-human primate. *iScience*, 26(10), Article 107901. <https://doi.org/10.1016/j.isci.2023.107901>
- van Dijk, R. E., Covas, R., Doutrelant, C., Spottiswoode, C. N., & Hatchwell, B. J. (2015). Fine-scale genetic structure reflects sex-specific dispersal strategies in a population of sociable weavers (*Philetairus socius*). *Molecular Ecology*, 24(16), 4296–4311. <https://doi.org/10.1111/mec.13308>
- Villafuerte, R., & Negro, J. J. (1998). Digital imaging for colour measurement in ecological research. *Ecology Letters*, 1(3), 151–154. <https://doi.org/10.1046/j.1461-0248.1998.00034.x>
- Walsh, N., Dale, J., McGraw, K. J., Pointer, M. A., & Mundy, N. I. (2012). Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird. *Proceedings of the Royal Society B: Biological Sciences*, 279(1726), 58–66. <https://doi.org/10.1098/rspb.2011.0765>
- Wang, H., Su, H., Chen, P., Hou, R., Zhang, Z., & Xie, W. (2019). Learning deep features for giant panda gender classification using face images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 279–285). <https://doi.org/10.1109/ICCVW.2019.00037>
- Warren, I. A., Gotoh, H., Dworkin, I. M., Emlen, D. J., & Lavine, L. C. (2013). A general mechanism for conditional expression of exaggerated sexually-selected traits. *BioEssays*, 35(10), 889–899. <https://doi.org/10.1002/bies.201300031>
- Wickham, M. H. (2014). Package 'ggplot2.' In *Create elegant data visualisations using the grammar of graphics. Version*, 2(1), 1–189. <https://ggplot2.tidyverse.org/>.