Check for updates

Received: 28 March 2023 | Accepted: 24 May 2023

DOI: 10.1111/2041-210X.14187

Methods in Ecology and Evolution

RESEARCH ARTICLE

RFIDeep: Unfolding the potential of deep learning for radio-frequency identification

Gaël Bardon[1,2] | Robin Cristofari[3] | Alexander Winterl[4] | Téo Barracho[2,5] | Marine Benoiste[2] | Claire Ceresa[1] | Nicolas Chatelain[2] | Julien Courtecuisse[2] | Flávia A. N. Fernandes[2,6] | Michel Gauthier-Clerc[7] | Jean-Paul Gendner[2] | Yves Handrich[2] | Aymeric Houstin[1,2,8] | Adélie Krellenstein[2,4] | Nicolas Lecomte[5] | Charles-Edouard Salmon[9] | Emiliano Trucchi[6] | Benoit Vallas[2,10] | Emily M. Wong[8,11] | Daniel P. Zitterbart[4,8] | Céline Le Bohec[1,2,12]

[1]Centre Scientifique de Monaco, Département de Biologie Polaire, Monaco, Principality of Monaco; [2]Université de Strasbourg, CNRS, IPHC UMR 7178, Strasbourg, France; [3]University of Turku, Department of Biology, Turku, Finland; [4]Friedrich-Alexander-University Erlangen-Nürnberg, Department of Physics, Erlangen, Germany; [5]University of Moncton, Canada Research Chair in Polar and Boreal Ecology and Centre d'Études Nordiques, Department of Biology, Moncton, New Brunswick, Canada; [6]Marche Polytechnic University, Department of Life and Environmental Sciences, Ancona, Italy; [7]University of Geneva, Faculty of Sciences, Geneva, Switzerland; [8]Woods Hole Oceanographic Institution, Applied Ocean Physics and Engineering Department, Woods Hole, Massachusetts, USA; [9]Beefutures, Nantes, France; [10]Terres australes et antarctiques françaises, Saint-Pierre, France; [11]Stanford University, Stanford, California, USA and [12]Université de Montpellier - Université Paul-Valéry Montpellier - EPHE, CNRS, CEFE UMR 5175, Montpellier, France

Correspondence
Gaël Bardon
Email: gbardon@centrescientifique.mc

Céline Le Bohec
Email: celine.le-bohec@cnrs.fr

Funding information
Academy of Finland, Grant/Award Number: 331320; Centre National de la Recherche Scientifique; Centre Scientifique de Monaco, Grant/Award Number: LIA-647 and RTPI-NUTRESS; Deutsche Forschungsgemeinschaft, Grant/Award Number: ZI1525/3-1 and ZI1525/7-1; Institut Polaire Français Paul Emile Victor

Handling Editor: Sarab Sethi

Abstract

1. Automatic monitoring of wildlife is becoming a critical tool in the field of ecology. In particular, Radio-Frequency IDentification (RFID) is now a widespread technology to assess the phenology, breeding and survival of many species. While RFID produces massive datasets, no established fast and accurate methods are yet available for this type of data processing. Deep learning approaches have been used to overcome similar problems in other scientific fields and hence might hold the potential to overcome these analytical challenges and unlock the full potential of RFID studies.
2. We present a deep learning workflow, coined "RFIDeep", to derive ecological features, such as breeding status and outcome, from RFID mark-recapture data. To demonstrate the performance of RFIDeep with complex datasets, we used a long-term automatic monitoring of a long-lived seabird that breeds in densely packed colonies, hence with many daily entries and exits.
3. To determine individual breeding status and phenology and for each breeding season, we first developed a one-dimensional convolution neural network (1D-CNN) architecture. Second, to account for variance in breeding phenology and technical limitations of field data acquisition, we built a new data augmentation

Daniel P. Zitterbart and Céline Le Bohec co-last authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

step mimicking a shift in breeding dates and missing RFID detections, a common issue with RFIDs. Third, to identify the segments of the breeding activity used during classification, we also included a visualisation tool, which allows users to understand what is usually considered a "black box" step of deep learning. With these three steps, we achieved a high accuracy for all breeding parameters: breeding status accuracy = 96.3%; phenological accuracy = 86.9%; and breeding success accuracy = 97.3%.

4. RFIDeep has unfolded the potential of artificial intelligence for tracking changes in animal populations, multiplying the benefit of automated mark-recapture monitoring of undisturbed wildlife populations. RFIDeep is an open source code to facilitate the use, adaptation, or enhancement of RFID data in a wide variety of species. In addition to a tremendous time saving for analysing these large datasets, our study shows the capacities of CNN models to autonomously detect ecologically meaningful patterns in data through visualisation techniques, which are seldom used in ecology.

KEYWORDS
artificial intelligence, behaviour, machine learning, RFID, wildlife monitoring

1 | INTRODUCTION

Electronic monitoring systems have been widely used over the past two decades to better understand animal populations without human disturbance (Fagerstone & Johns, 1987; Schooley et al., 1993). Radio-Frequency IDentification (RFID) technology allows the monitoring of uniquely identified individuals and automated recording of the presence of tagged individuals at chosen locations (Gibbons & Andrews, 2004). By placing RFID antennas along animal paths, at perches or narrow entries of the breeding site (Bonter & Bridge, 2011; Zydlewski et al., 2006), individual survival and breeding rates as well as behaviour and locations can be precisely estimated in, for example the classical capture-mark-recapture framework (Le Bohec et al., 2008). While RFID technology allows the recording of vast amounts of data, it also creates new challenges for data treatment, even if the data structure itself is rather simple (i.e. id, date and time, and location for each detection; Iserbyt et al., 2018). Because RFID data are not directly linked with biological parameters, one of the classic approaches is human expert interpretation (Afanasyev et al., 2015; Descamps et al., 2002). Even today, most of the information extraction and ecological interpretation from such detection data is done manually, although this remains extremely time-consuming and potentially biased by human interpretation. Additionally, the difficulty in manually processing potentially large numbers of detection data is increased by the possibility of missing detections (Hughes et al., 2021).

A solution to these challenges may lie in automated data processing that could mimic the behaviour of an expert analyst. Artificial intelligence has been the focus of intense methodological effort in ecology: it has been used to process various sources of data including imagery or passive and active acoustic data, and to detect, classify, localise, identify, estimate and predict at every biological scale, from individuals to ecosystems (Christin et al., 2019; Pichler & Hartig, 2022). Among artificial intelligence methods, deep learning has a wide and promising scope but often lacks approachable workflows for ecologists. Deep learning can be generally defined as a set of methods using "deep" (i.e. multi-layer) networks of artificial neurons to process and "learn" complex features from data: see LeCun et al. (2015). Among these, convolutional neural networks (CNN) have been initially developed for image content classification (Krizhevsky et al., 2017) but have also been used for classifying signals (Hinton et al., 2012) such as human activity classification (Mutegeki & Han, 2020), birds vocalisation classification (Kahl et al., 2021) or marine mammal detections (Shiu et al., 2020). Yet, CNN capacities remain unexplored in numerous fields such as RFID data processing.

Recent efforts have been made to automatically infer behavioural patterns from various types of biologgers through AI (Fannjiang et al., 2019; Wang, 2019): for instance, accelerometers have shown promising capacities to detect food-catching events (Brisson-Curadeau et al., 2021) or to classify activity (Jeantet et al., 2021; Sakamoto et al., 2009). RFIDeep builds upon these efforts to address the specific nature of RFID data. While active biologgers record rich, multidimensional data, their record time is limited because of the required trade-off between miniaturisation, storage capacity, power consumption and impact on wildlife (Bodey et al., 2018). In contrast, passive, battery-less RFID tags have no demonstrable impact on an animal's behaviour and function throughout the individual's lifetime but the trade-off is that although the tag is attached to the animal, detection only occurs

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä̈tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

at one or more fixed points (the active antennas), thereby offering a very narrow observation window, similar to the classical bird ringing observation approach (e.g. with mist nets). This creates a rather unique data structure with specific challenges for interpretation. RFID technology is also exposed to two major constraints because of the impossibility to detect multiple tags at the same time with a single antenna and the impossibility to instal several antennas at the same place, due to electromagnetic interference. By increasing probability of missing detections, the tag and reader collision problems create a trade-off between the number of deployed tags (the size of the dataset) and the probability of undetected individuals (the completeness of the dataset). This leads to challenges in inferring missing detections to correct the locations and movement patterns of individuals. Like in other automated data processes, such data imperfections need to be considered and if possible repaired with suitable algorithms.

Here, we demonstrate that non-explicit detection data from fixed observation points contain enough information to infer individual behaviour. Taking advantage of the recent developments in deep learning methods, we develop the "RFIDeep" workflow to automatically extract breeding status and outcome from detection data acquired by RFID antennas using CNN. We illustrate how deep learning methods detect biological features in RFID data with very high classification accuracy and demonstrate the use of a visualisation method not yet commonly implemented in ecology.

RFIDeep was initially developed for an "archetypal" real-life dataset with a ca. 25 years-long RFID detection time series collected on known-age/history king penguins *Aptenodytes patagonicus* (Gendner et al., 2005). Unlike flipper bands used until then, which are detrimental to the individuals (Gauthier-Clerc et al., 2004), the recording of every transit between the colony and the sea of RFID-tagged birds, throughout their life, allowed a more accurate and unbiased description of the reproductive patterns of the species (Descamps et al., 2002), and of the population's demographic parameters (Le Bohec et al., 2008). In these previous studies, all RFID detections were manually analysed by human experts and none of them used the entire dataset of RFID-tagged penguins. Since breeding king penguins exhibit highly stereotyped movement patterns (Descamps et al., 2002), they were good candidates for artificial intelligence classification. Based on direct field observations and molecular sexing data, we trained several CNN to infer RFID-tagged penguins' sex, breeding status and outcome (Breeding vs. Non-Breeding; Success vs. Failure), and breeding dates. We developed RFID-specific data augmentation steps to account for biological variance and data acquisition imperfections. We trained our classification process with field observation data and tested it with (i) annotated data to compare the performance of automatic classification with the human experts and (ii) independent field observation data.

We provide all source codes used in RFIDeep workflow that could be applicable for studies using RFID data acquisition and that could inspire ecologists to develop their deep learning process. Finally, a software named Sphenotron, developed to represent movements and locations (in or outside the breeding site) based on RFID detections, is provided with a sample dataset as an example of an RFID data visualisation method.

2 | MATERIALS AND METHODS

2.1 | Overall structure of RFIDeep workflow

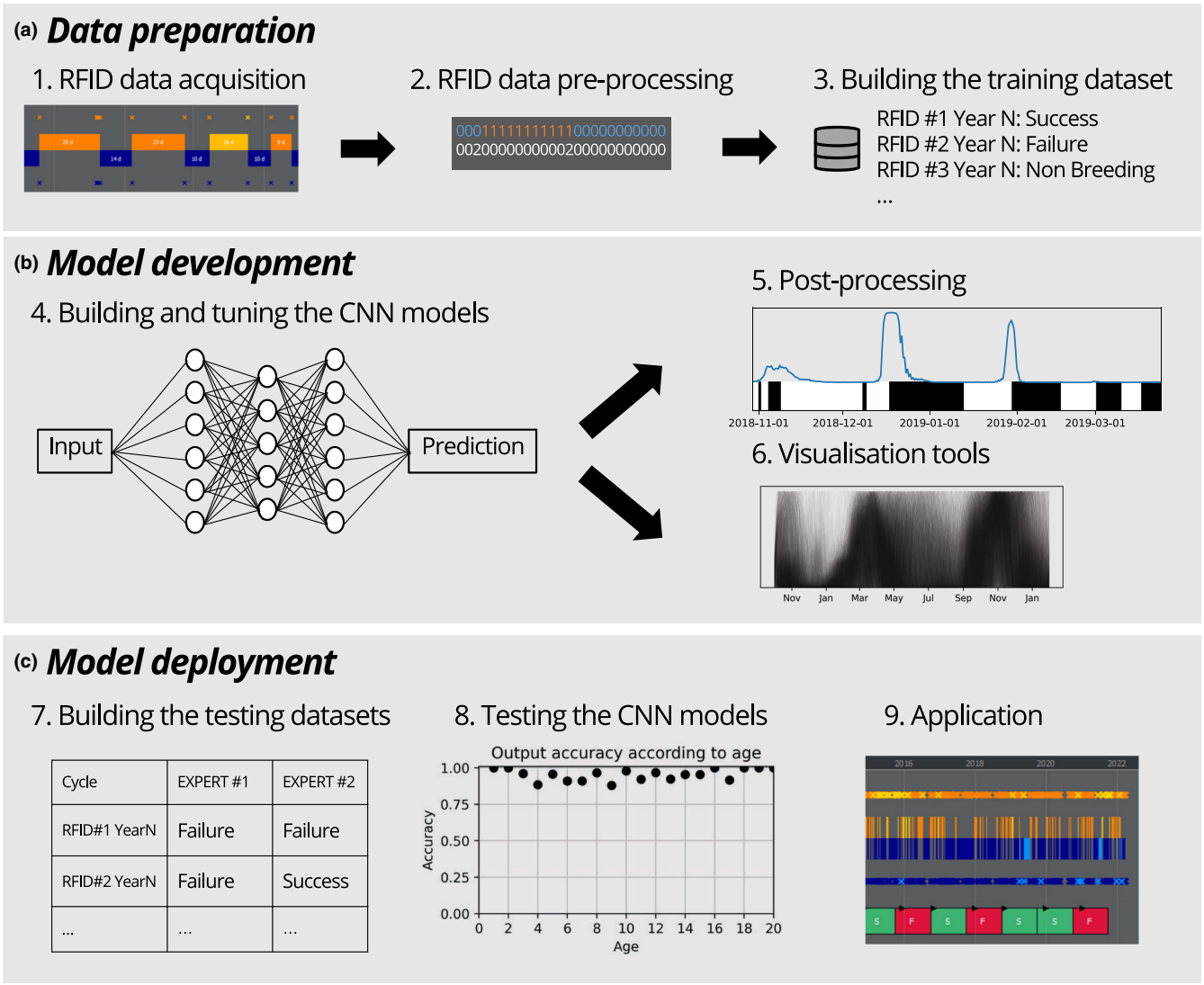
Figure 1 summarises the steps needed to classify RFID data within a deep learning framework and provides a comprehensive view of the use of the RFIDeep workflow. This workflow can be adapted to fit other acquisition systems and species, if we have access to (1) RFID detection data (timestamp and individual ID) collected in a repeatable way and (2) ground truth data for a subset of individuals. The size of this ground truth set depends heavily on the signal-noise ratio in the target system, but a key requirement is its diversity, which must cover all expected biological situations (Christin et al., 2019).

2.2 | Application on a seabird species long-term monitored by RFID

2.2.1 | RFID data acquisition

Here, we used data collected from the colony of king penguins (*A. patagonicus*) named 'La Grande Manchotière' and located at Possession Island, Crozet Archipelago (46°25 S, 51°45 E). This fieldwork was approved by the French ethics committee (last: APAFIS#29338-2020070210516365) and the French Polar Environmental Committee and permits handling animals and access breeding sites were delivered by the "Terres Australes et Antarctiques Françaises". A sub-area of the colony of ca. 10,000 breeding pairs has been electronically monitored since 1998 with RFID technology. As of 2022, four pathways between the sea and the colony (the only ways in or out of the colony) are equipped with permanent automatic identification systems (the detailed information of the field site and systems are described in Gendner et al., 2005). In short, these automatic systems are composed of paired antennas that record the direction of each commuting bird that has been implanted subcutaneously with an RFID tag. Patterns of presence and absence of ca. 15,000 RFID-tagged birds throughout their breeding seasons and life have thus been recorded since 1998. This has generated a large (and increasing) quantity of data, with, for instance, seven million individual detections as of 2022. To manage, visualise and use information in the field (e.g. select specific groups of birds of known age or history), we developed *Sphenotron*, a python software that displays the location (in or out of the colony) of the individuals during their life, based on the latest known location transition (entrance or exit) for each bird (see Supporting Information A). Thanks to well-known king penguin's stereotyped presence/absence patterns at the

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä¬tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



Cycle	EXPERT #1	EXPERT #2
RFID#1 YearN	Failure	Failure
RFID#2 YearN	Failure	Success
...	...	...

FIGURE 1 Overall structure of the RFIDeep workflow classifying Radio-Frequency IDentification (RFID) data with deep learning. The workflow is divided into three phases: data preparation, model development, and model deployment. (a) *Data preparation*. (1) RFID data acquisition: many individuals are equipped with RFID tags and antenna systems are installed at key locations to register the detections. A software called Sphenotron (Supporting Information A) has been developed to represent detections and transitions (in or out of a specific location, e.g. a colony or a nest) of RFID-tagged individuals. A colouring scheme is available to picture the ins (orange, above the line) and outs (blue, below the line) of tagged individuals. (2) RFID data pre-processing: a correction of missing detections is first applied. RFID data are then formatted (e.g. in or out of a specific location encoded as 1 and 0, and number of detections per time period) to have a unique and readable format for deep learning or for other analyses. (3) Building the training dataset: direct observations of RFID-tagged individuals are used to build a ground truth dataset of labelled vectors giving the true classification. (b) *Model development*. (4) Building and tuning the convolution neural network (CNN) models: the architecture of deep learning models and hyperparameters are tuned with the training dataset. Data augmentation is implemented to cover more biological and technical variance. An individual network is built for each classification problem (e.g. breeding status, sex). (5) Post-processing: classification networks are derived to extract other biological information requiring a post-processing step such as location of stereotyped patterns in RFID data (e.g. determination of the breeding dates with a probability curve (in blue) over presence/absence pattern in black and white, respectively). (6) Visualisation tools: models are validated and interpreted with visualisation tools (e.g. with black curves representing the focus of the model during the breeding season). (c) *Model deployment*. (7) Building the testing datasets: a testing step is used to remove biases induced during parameterisation with independent datasets (i.e. human expert classifications and independent ground truth dataset). (8) Testing the CNN models: model tests assess performance but also ensure that models consistently perform according to classes and individual characteristics (e.g. age, sex and life stage). (9) Application: classifications are applied to all detection data after pre-processing and formatting (i.e. after correction of missing detections and building of vectors), and results are represented in Sphenotron for each individual (successful breeding cycles "S" in green, failed breeding cycles "F" in red).

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitäts-Bibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

breeding site (see Figure S7), we can classify the breeding status of any RFID-tagged individual (Non-Breeding, Failed Breeding, Successful Breeding). Start of a breeding cycle (breeding date) can also be determined, it is defined as the beginning of the pattern characteristic of the courtship and incubation period, that is the first long sojourn at the colony following the annual moult and exceeding 10 days (Descamps et al., 2002). Additionally, the sex of an individual can also be derived from presence/absence patterns at the colony. An automatic sex determination has great potential application for many species where sex determination is challenging (i.e. monomorphic species like king penguins).

2.2.2 | RFID data pre-processing

Input data

To prepare the detection data in an appropriate format, we chose to represent absence and presence time-series for each breeding cycle with two vectors providing the location of the individual at the end of 12-h periods (states 0 and 1) and the number of detections occurring during the 12 h (Figure S8). For one individual and one given year *n*, we built vectors encompassing the breeding cycle. For the King penguin, vectors start 1 October of the year *Y* and end 31 January of the year *Y* + 2 to cover the entire >1-year breeding cycle of the species (Figure S7). We obtained two vectors of 974 elements for each individual and each year. This step can be tailored to match other RFID acquisition systems and species, for instance by dropping the first vector when no location ('in' or 'out') is defined but only simple detection are recorded, for example at a feeding site.

Missing detection correction

To tackle missing detections that can occur when individuals exit or enter their breeding site, we developed an algorithm to repair simple missing detections (i.e. detections on only one antenna of a pair, resulting in uncertainty in the individual's walking direction), as was similarly done by Austad et al. (2023). These corrections are usually trivial: for example, when an individual is detected only on the inside antenna, followed later by an entrance (i.e. outside-inside transition), an outside detection is inferred to restore a valid pattern in detections corresponding to the missed exit from the colony. We simply built the algorithm to detect all unrealistic successions of detections and to add the corresponding missing detections in all possible cases (see Supporting Information B).

2.2.3 | Building the training dataset

To build a training/ground truth dataset, we visually monitored 295 RFID-tagged individuals over 9 years (2011–2019), assessing their breeding status and behaviour directly through field observations. Birds were monitored from the beginning of the breeding season (November–January), thereby we were able to detect early breeding failures that may have been difficult to distinguish from non-breeding behaviour using RFID detections alone. Breeding outcomes (S: Success; F: Failure) from these study birds was determined according to the survival of their chicks until they fledged. The sex of individuals was determined with the observation of their first period in the colony, as females leave the breeding site right after egg laying, while males care for the egg (Descamps et al., 2002). A ground truth database with breeding status, timing of breeding, and sexing for 463 breeding cycles was then compiled over the years.

2.2.4 | Building and tuning the CNN models

Several models were built to describe breeding activities from regular movement patterns (see Figure S7) with a classification workflow (see Figure S9):

1. Two models to determine if an individual in a given year was a breeder (Breeding vs. Non-Breeding) and if the breeding cycle was successful (Success vs. Failure),
2. A model to distinguish the sex of an individual through sex classification of only breeding cycles and a prediction compiling all the sexes identified over the lifetime breeding seasons,
3. A model to determine the most likely breeding date of males and females separately, through post-processing of a CNN model.

CNN architecture of these models was chosen using a classical simple architecture (see LeCun et al., 2015) and through trial and error (see details in Figure S10). Each model was trained on a training set of 80% of the dataset, and the remaining 20% was used as a validation set to measure model performances and avoid overfitting (shown by low validation accuracy and high training accuracy), as suggested by Christin et al. (2019). Multiple training of the models with random splitting of the training/validation sets was performed to cross-validate the hyperparameters. Once the final hyperparameters were chosen, validation accuracy with the 20% validation set was recorded, and the final models were trained with 100% of the training datasets. When the models were applied to detection vectors to generate the classifications, the most probable class was chosen for the classification.

To extend the generalisation capacities of our models, we used a data augmentation process during the training of the models, in the sense of a distortion of the data (LeCun et al., 1998). In other words, during the training, the dataset is perturbed in a controlled way at each iteration: while the data quantity remains unchanged, its variability is increased, thus covering more possible situations, and avoiding overfitting. We used two types of augmentation: the first one consisted in shifting the breeding cycles by a random number of days, as usually done with imagery data to make the models translational invariant. At each iteration of the training, we shifted each training vector by a zero padding at the end or at the beginning of the vector, while trimming the same number of elements on the opposite side. We used a random offset between −30 and 30 days to cover large biological variability in the

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä=Tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

phenology of our species (but this can be easily changed in the code, and this step can be turned off). The second part of data augmentation focused on simulating missing RFID detections. In the actual dataset, the most frequent problem is the loss of a single detection, which is solved by our correction algorithm. Therefore, we chose to remove 10% of the single detections at each iteration, before applying our correction algorithm, allowing a complete recovery of the original detections for at least 50% of individuals (see Supporting Information C) and leaving uncorrected detections and erroneous locations to improve training generality. Models for determining the breeding status were trained with and without these data augmentation processes to assess the benefits of this step.

2.2.5 | Post-processing

Sex determination

With RFID detections of king penguins, a human expert can only distinguish males and females based on the length of the first periods in colony of the breeding cycles, therefore we thought it safer to assume that prediction over a single breeding season would be less reliable than prediction over the whole lifetime. We then averaged the classification probabilities for each sex, for each identified breeding attempt and we obtained the most probable sex over the lifetime of the individuals. We also registered the sex classification for each breeding cycle to measure the benefit of this pooling in classification performance. This step can be skipped for species where sex is readily identified but can be useful for species such as seabirds in which sex can be more easily determined using behavioural data.

Breeding date

We used CNN models to determine the breeding date by scanning all possible breeding cycles in a year and determining the most probable one (see Supporting Information D for details on the method). We obtained a certainty curve along the year, with the maximum corresponding to the most probable breeding date (as illustrated in Figure 2 with two true breeding cycles). In our King penguin study case, we trained two different models for males and females separately to account for the difference in patterns at the beginning of their breeding cycles.

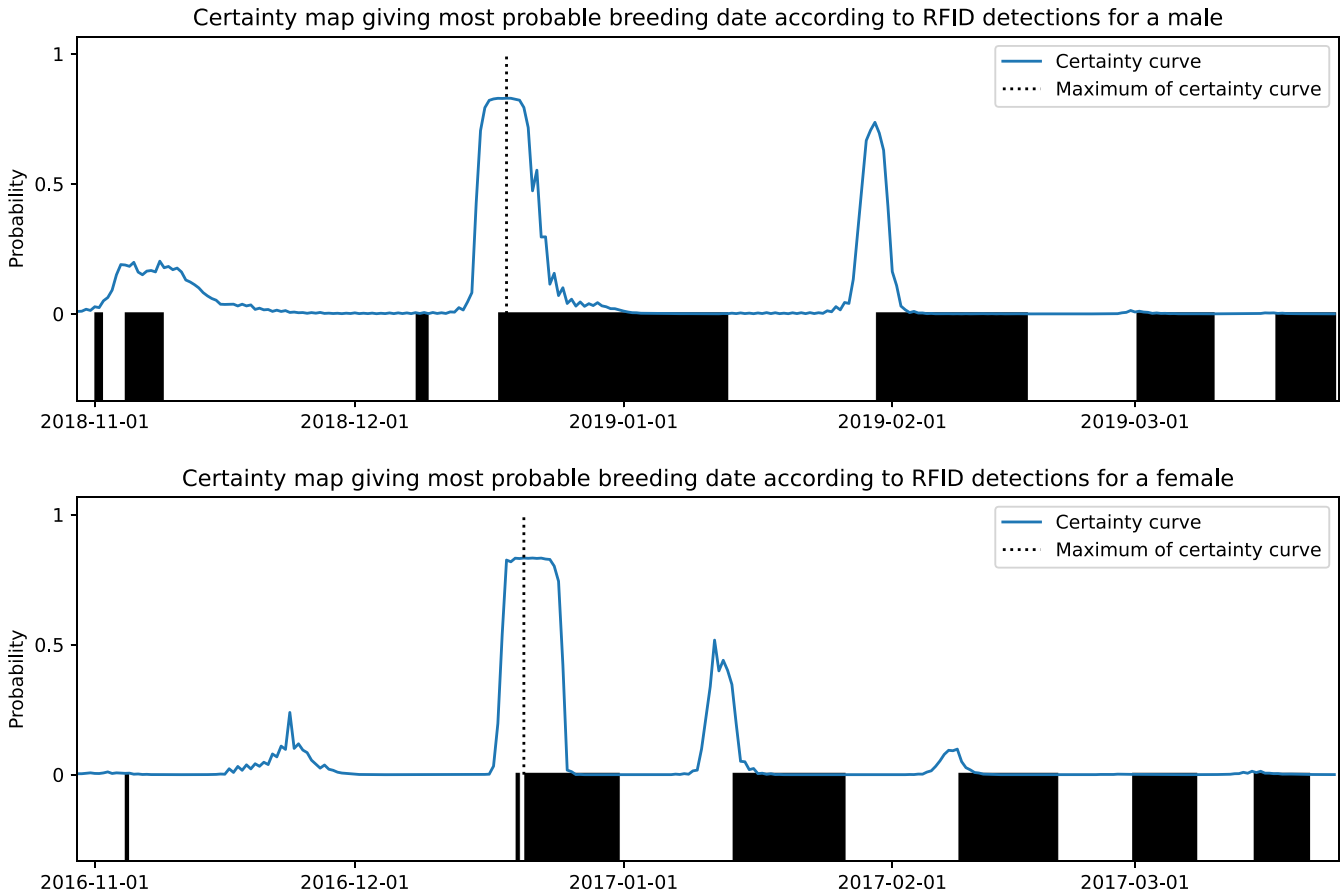


FIGURE 2 Examples of certainty maps produced by the scanning algorithm to detect the beginning of a stereotyped pattern. Here, the most probable breeding date of a male and a female was determined. The blue curve represents the probability (between 0 and 1) that the breeding cycle starts on a relevant date. The black and white bars in the lower part of the figures represent the location of the Radio-Frequency IDentification (RFID)-tagged individual (inside and outside the colony, respectively). The most probable breeding date corresponds to the maximum of the blue curve (dashed line).

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitäts-Bibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2.6 | Visualisation tools

We used visual explanation techniques to show parts of the input data that are identified by the convolutional layers and used to perform the classification. We leveraged techniques recently developed to produce heat maps on images classified by a 2-D CNN algorithm to show which pixels contribute most to the classification (saliency maps, Zeiler & Fergus, 2014 and class activation mapping, Zhou et al., 2016). We produced this type of visualisation on breeding cycles with the GRADient-weighted Class Activation Mapping (Grad-CAM) algorithm (Selvaraju et al., 2017) that was directly applicable to the 1-D CNN layers. In short, the Grad-CAM uses the gradients of the final convolutional layer to produce a coarse localization map from an input image (or vector) by searching for pixels whose intensity should be increased to increase the probability of a given class. We ran this algorithm on all breeding cycles in our dataset and obtained a graph of importance value for each element of the vector (each 12-h period in our example) for a particular class of interest. We computed these activation plots and compared to the raw input detection data for (1) the Breeding versus Non-Breeding model, which restricts the dataset to breeding cycles classified as breeding in order to identify the features used by the algorithm to detect a breeding cycle, and (2) the Success versus Failure model, which includes only breeding cycles classified as successful in order to identify the regions of the breeding cycle that were indicative of a success.

2.2.7 | Testing the CNN models

To test the overall classification performance, we used a global accuracy metric of the different models given by $ACC = N_{correctpredictions} / N_{predictions}$ (Powers, 2020). Since our ground truth datasets were well balanced across classes (168 Non-Breeding; 131 Failure; 164 Success), the global accuracy metric did not approach its limits through class unbalance, and it provided a simple and effective metric of overall classification performance for training. To provide a measure of classification accuracy for all possible classification thresholds, we also used the AUC-ROC score (Area Under the Receiver Operating Characteristic Curve; Fawcett, 2006). To assess the accuracy of breeding date determination, we used a threshold of 5 days between the true breeding date and the predicted date to define whether or not a breeding date was correctly predicted (see Supporting Information E). Two fully independent datasets that were never used in model training were used to quantify an unbiased estimate of model performance (Kuhn & Johnson, 2013): (i) an additional ground truth dataset for years 2021 and 2022 containing 302 field observations of breeding outcomes only (Successful or Failed breeding seasons) and (ii) a blind-labelled testing dataset encompassing 917 breeding cycles of penguin individuals: breeding status and breeding date were not determined through field observations but by human experts who examined the RFID detections of individuals using our custom-designed Sphenotron software (see Supporting Information A). Human experts, with a strong knowledge and experience of the species in the field, were trained using the ground truth dataset, blindly examining detection data to infer breeding cycles, and cross-checking previously assigned breeding cycles. Two human experts were chosen to label the same dataset, and we tested our models with both classifications. We also computed the global accuracy metric between the datasets labelled by the two human experts to assess human variability in classification. The performance of the lifetime sexing method was compared to a molecular sexing dataset of 6196 birds (molecular sexing method adapted from Griffiths et al., 1998 showing 100% accuracy in typical cases, Purwaningrum et al., 2019). Because sex was estimated with a variable number of breeding cycles between individuals (we used all available breeding cycles for each bird), we also tested whether the accuracy of pooled sexing increased when including additional breeding cycles. Finally, we computed the accuracy of the models for each age class and for males and females separately to test whether the performance of our models was consistent over the whole dataset.

3 | RESULTS

3.1 | Model training

We chose 200 epochs (i.e. training iterations) for training of each model, which yielded the best results for validating model accuracy while avoiding overfitting. Each model took approximately 1 h to train using a laptop computer with a CPU Intel Core i7-10750H (2.60 GHz) and a GPU Nvidia GeForce GTX 1660 Ti, a non-prohibitive technology as of 2023. The performance of models, according to the validation datasets used to select the CNN architecture and hyperparameters, reached near perfection for the three models, that is Breeding versus Non-Breeding, Success versus Failure, and Male versus Female, with global accuracy of 99.1%, 99.7% and 100%, respectively. As expected, the three models without a data augmentation step achieved lower performances with global accuracy of 94.6%, 91.5% and 96.6% for Breeding versus Non-Breeding, Success versus Failure, and Male versus Female, respectively.

3.2 | Model visualisation

Activation maps $L^{c}_{(grad-CAM)}$ (Figure 3) showed that for the vectors classified in the Breeding class (i.e. individuals that attempt to breed; Figure 3a,c), the model focused on the beginning of breeding, when long periods in the colony occur for breeders but not for non-breeding birds that do not have long fasting periods on the breeding site. For the Success class (Figure 3b,d), the model focused the pre-winter period and the post-winter feeding period. As expected, these are the parts of the breeding cycle that can be missing if the breeding fails during incubation, brooding or even during the winter fasting period. Unsurprisingly, the visualisation maps relied on the

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä=tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

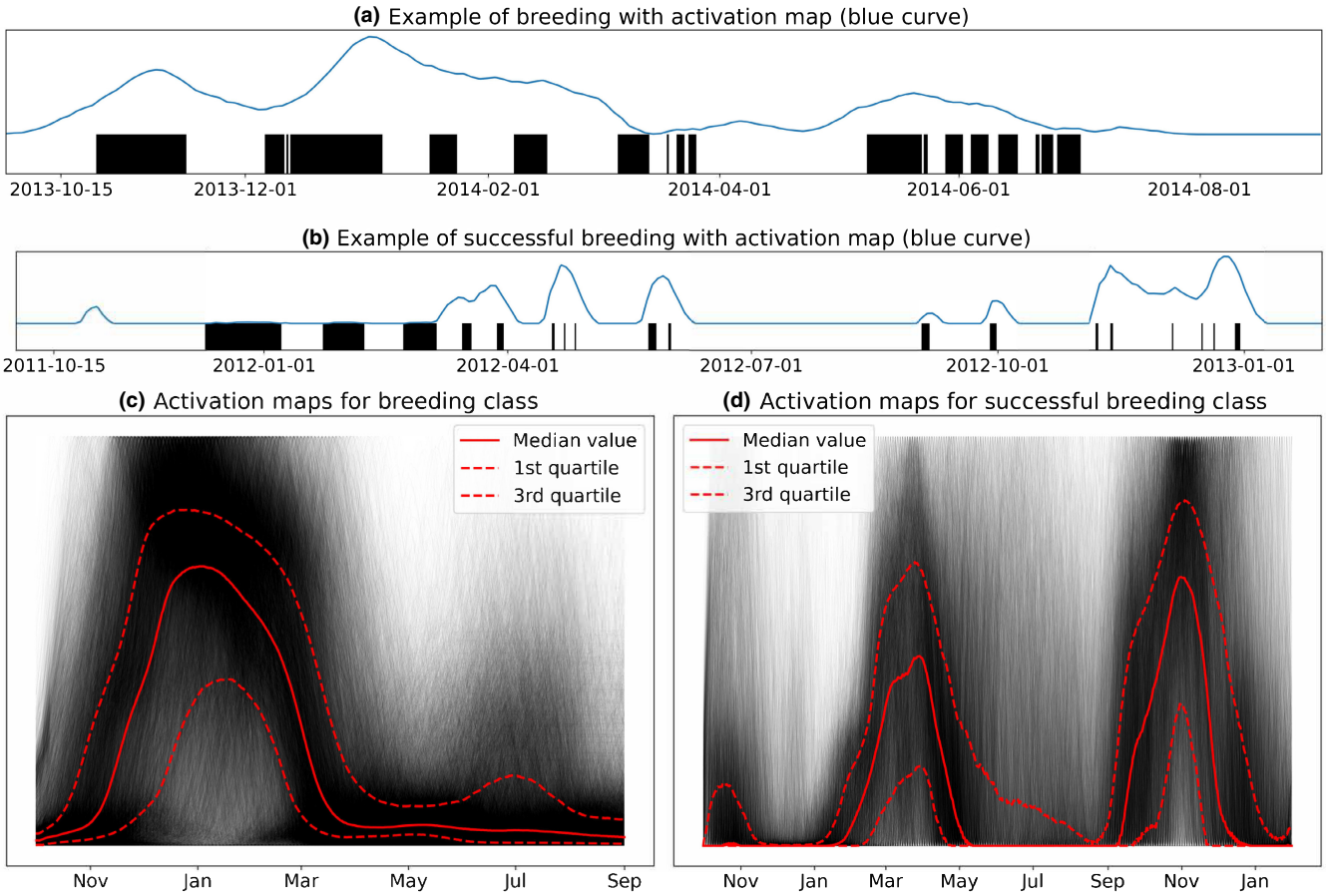


FIGURE 3 Activation map (blue curve) showing regions of the cycle used by the convolution neural network procedure to produce the classifications and simplified presence/absence pattern (black: inside the colony; white: outside the colony) for two true breeding cycles (a, b), and median maps (c, d) illustrating all maps with the median curve, the first quartile and the third quartile. (a, c) Correspond to the Breeding class (all breeding cycles classified as Breeding) and (b, d) to the Success class (all breeding cycles classified as Success), respectively.

same regions that human experts reported using as criteria for determining whether individuals actually attempted to breed and succeeded in breeding.

3.3 | Model deployment

The trained models were used to predict the breeding status and dates of all RFID-tagged individuals since 1998 (i.e. 85,524 breeding cycles from 14,795 different individuals). On the laptop computer used here, prediction (from raw RFID data to classification) of breeding status and sex of all birds required 140 s, but it took 1.1 h for the determination of the breeding date due to the number of predictions needed (320 for each breeding cycle). In comparison, it took a human about 1 min to make the same decision as RFIDeep for one bird and one breeding cycle, which would correspond to 1320 h or 165 workdays (8 h per day) to classify all breeding cycles. Tests of model's classifications against two human expert classification datasets resulted in high global accuracy metric (Table S1, e.g. $ACC_{Bvs.NB} = 0.963$ and $ACC_{Svs.F} = 0.967$ for prediction vs. dataset 1). Similarities between the expert-labelled datasets were globally equivalent to the accuracy of our CNN models, indicating the high performance of the automatic classification procedure. The AUC scores of Breeding versus Non-Breeding and Success versus Failure computed with the human expert classification were even higher (e.g. $AUC_{Bvs.NB} = 0.993$ and $AUC_{Svs.F} = 0.992$ for prediction versus dataset 1, see Figure S11). As expected, models with data augmentation consistently performed better than models without any transformation of the input data. Furthermore, tests on the independent ground truth dataset confirmed the high performance: all cycles were indeed well classified as breeding season and Success versus Failure classification reached a 96% accuracy and an AUC-ROC score of 0.995. The lifetime classification sexing procedure yielded an 88.6% accuracy compared to the molecular data ($AUC_{sex} = 0.930$ see Figure S12). Before pooling lifetime sex probabilities, the global accuracy of sexing was only 81.7%. As expected, we also found that sex classification accuracy from the pooling of lifetime sex probabilities increased with the number of breeding cycles used to determine the sex of an individual (see Figure S12). According to the expert-labelled dataset 1, predictions were slightly better for males than for females for the breeding status ($ACC_{males} = 0.945$ vs. $ACC_{females} = 0.932$) and inversely for the breeding dates ($ACC_{males} = 0.852$ vs. $ACC_{females} = 0.889$). The

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitat Stabibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

breeding dates also appeared to be less predictable for young individuals (see Figure S13).

4 | DISCUSSION

In this study, we developed and tested a complete workflow based on CNN models to automatically infer behavioural and fitness traits from RFID-tagged animal detection data. Based on a train-test split approach, we showcased the potential of deep learning to adequately replace human expertise in RFID data processing in a much shorter time span. Remarkably, humanlike performance to translate patterns from detection data into meaningful biological parameters was reached with a rather simple CNN architecture and standard desktop computing capacity. To improve results, we used time-shift data augmentation to mimic the variability that could occur due to biological mechanisms (e.g. a shift in breeding dates) and simulated data dropouts to mimic technical constraints (e.g. missing detections). We also developed a post-processing step to extract dates of breeding, and, with a visualisation technique, we identified the regions of the dataset used by the models to classify the breeding cycles. Such a framework can be used beyond our example dataset and help to quickly classify the breeding activities of many individuals, even more so for long-term projects for which pre-processing analysis is very time- and labor-consuming (in our example, we worked on ca. 15,000 individuals over ca. 25 years). Certainly, we do anticipate a wide application for other colonial birds with their many ins and outs of their colonies and with their complex patterns of phenology—however, we acknowledge that RFID monitoring may be more challenging in some species (Bonter & Bridge, 2011). For non-colonial species, we anticipate a good performance of our model with likely simpler RFID data (e.g. simple detection at a feeding site), if detection rate is high enough to capture regular patterns. While it is still challenging to successfully transfer pre-trained deep learning from a study case to another (Marcus, 2018), RFIDeep workflow is tailored for any study classifying behaviours based on RFID-tagged animal detections. RFIDeep was successfully tested and used on another species, the Adélie penguin (*Pygoscelis adeliae*), for which breeding is markedly different from our first dataset example, yet monitored with a similar automatic RFID setup. Given that our model performed well for these contrasting datasets (global accuracy of 95.2% for breeding outcome determination and 93.5% for sex, see Supporting Information F for details), we argue that any RFID-monitored species with stereotyped movements during a given life stage could certainly benefit from the RFIDeep workflow, such as bumblebees (Molet et al., 2008), Leach's Storm-petrels (Zangmeister et al., 2009), hummingbirds (Bandivadekar et al., 2018), as well as other penguin species (Chiaradia & Kerry, 1999; Horswill et al., 2014). Other configurations of acquisition system are compatible with this workflow, as for example (1) RFID antennas at the entrance of cave for Yelkouan Shearwater (*Puffinus yelkouan*; Austad et al., 2023) or for southern bent-winged bat (*Miniopterus orianae bassanii*; van Harten et al., 2019); (2) RFID-equipped nesting-box allowing continuous record of presence/absence in the nesting box of equipped individuals for Tengmalm's owl (*Aegolius funereus*; Zárybnická et al., 2016) that could certainly allow the classification of breeding status for multiple species with nest attendance patterns (Bambini et al., 2019); (3) feeders equipped with antennas that detect RFID-equipped birds coming to search for food and that record frequency of visits and time-spent (Bandivadekar et al., 2018); and (4) safe passages equipped with RFID antennas such as eco-passages across/over or under roads that can be frequently used by some monitored individuals (Dexter et al., 2016) and be informative on their biological status. We are confident that the RFIDeep workflow will help biologists to adopt deep learning applications more easily, either by using the codes directly or by adapting it for their requirements.

Furthermore, the missing detection correction and data augmentation algorithms implemented in RFIDeep have great potential to tackle uncompleted and/or low-quality datasets, such as those produced by mobile RFID antennas temporarily deployed (Cristofari et al., 2018). It is important to note that our correction algorithm does not generate new data, but explicitly fills in information (here, direction of the bird) that is already contained in other detections in order to help classification. This is in contrast with generative AI, which is capable of generating new data. However, classification results could be used to recover fully missing events as a post-processing step: for example, we may infer a fully missing detection date given breeding date and outcome, and sex.

Both validation (during training) and testing steps (afterward) showed the high performances of the RFIDeep models, on the one hand, in reference to the ground truth data (96% accuracy for Success vs. Failure) and, on the other hand, with a human-machine comparison (>96% accuracy for Success vs. Failure and >93% accuracy for Breeding vs. Non-Breeding). Even though we developed a software specifically to display detections and locations (inside or outside the colony) of RFID-tagged individuals during their life (Sphenotron, see Supporting Information A), the distinction between specific breeding status can be challenging, if not impossible. For example, a confounding situation in our case study happens between non-breeding and failed breeding when the failure occurs very early in the season. By using automatic classification, we standardised the bias among all breeding classifications throughout the years of monitoring through the removal of variability related to potential differences in human expert interpretation. This allowed for remarkably fast extraction of life history parameters of the monitored individuals, necessary to estimate population vital rates (e.g. survival, breeding success) and viability, in addition to other breeding and phenological traits. For example, breeding success inferred with a very good classification accuracy (97.5% of accuracy in the classification of successful vs. non-breeding or failed breeding) can then be used to estimate fecundity rates of the monitored population with high confidence for all monitored years. Further analysis on classification scores given by the models also showed that in situations where the classification is wrong, the score is more uncertain (average score for well classified success: 0.995 vs. average score for misclassified success: 0.785; see

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä=Tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

distribution of scores in Figure S14). It leads us to consider an 'uncertain' class to avoid misclassification of a large part of cycles that cannot be classified with RFID detections only.

Our analysis highlights the benefits of data augmentation to cope with more biological variance than contained in our ground truth data. Data augmentation is commonly used to improve deep learning applications (Taylor & Nitschke, 2018) and sometimes developed in the application of deep learning in ecology (e.g. with image data, Kälin et al., 2019 or audio data, Kahl et al., 2021), and has significantly enhanced our classification process. While data augmentation is usually done by adding random noise to the dataset (e.g. in pictures for 2-D CNN classification with image rotations for instance, Pawara et al., 2017), here we aimed to mimic biological variance and technical limitations of the RFID data acquisition systems. Doing so, we covered a large variance in breeding dates, enabling us to anticipate breeding seasons that could begin earlier or later than those existing in our ground truth data, because of environmental shifts already observed or expected in the coming years/decades (Visser et al., 2021). This applies not only for the Sphenisciformes species used in this study but likely also for other species with a high variance in breeding phenology (De Villemereuil et al., 2020). Another interesting aspect of the automatic classification of breeding cycles is the independence among predictions. Indeed, each breeding cycle was analysed without supplement information about the year (e.g. average breeding success, phenological data), the individual (e.g. age, body condition), and previous and future breeding cycles. The breeding classification of lifetime datasets by human experts can induce bias for quantifying the inter-individual and intra-individual heterogeneity in breeding cycles since they are usually not classified independently. However, while there may be an advantage to having independent classifications, the lifetime information may also be beneficial, for instance to better determine the breeding date of the very first breeding seasons that tend to be less predictable for numerous species (see Figure S13) and to consider intra-individual repeatability of phenological and behavioural traits. It would also be useful to train CNN models with mixed data (e.g. RFID detections and automated weighting at the detection point) to increase the classification accuracy and/or complexity, and to refine further some of the analyses (e.g. the stage of breeding failure), as it has also been done in other fields (Ahsan et al., 2020). With visualisation techniques, we showed which parts of the datasets are mostly used to perform classification. They provided a peek into the deep learning 'black box', making the process more transparent for the user, a shortcoming that often prevents its use by ecologists (Borowiec et al., 2022). We argue that such a step can help expand the potential of deep learning to describe and analyse ecological big data. In our example, while activation maps are primarily used by CNN for classification, their visualisation allows the detection of the specific breeding activities or features, such as seasonal phenology. In our application on king penguins, the CNN models showed that the presence or absence of pre- and post-winter chick feeding patterns were the most important criteria for predicting breeding outcome. Although these regions can be used for distinguishing between failure and success, it reinforces our interest in using these visualisation techniques not only to understand how our deep learning models work, but also to detect regions of interest in our datasets. It also highlights the use of CNN models that are not frequently found in ecological studies but have great potential, for instance, to detect hidden patterns in large datasets. Moreover, to cope with the recent explosion of big data acquisition due to increasingly sophisticated, miniaturised, autonomous, and powerful data collection instruments (Williams et al., 2020), visualisation tools are becoming increasingly important in detecting similar patterns in given classes or differences between similar classes. For instance, identifying parts of the vocalisation essential to distinguish between species or even individuals is key in bioacoustic studies (Kobayashi et al., 2021; Stowell et al., 2016). Visualisation techniques have also been used to select the most informative variables to infer animal behaviours from multi-sensor data (in green turtles, *Chelonia mydas*, Jeantet et al., 2021) or to select the most relevant morphological characters to identify species (among midges, Milošević et al., 2020 and mosquitoes, Park et al., 2020). By developing tools to help users unleash the vast potential of machine learning in ecology and to increase numerous benefits of RFID technology, we also aim with RFIDeep to foster low-impact monitoring of sensitive species by reducing human presence and intervention in wild habitats (Hughes et al., 2021; Rafiq et al., 2021). We are convinced that combining automatic data collection and real-time data analysis and storage will help secure key ecological information over time necessary to continuously monitor the health of wild populations and their ecosystems.

AUTHOR CONTRIBUTIONS

Gaël Bardon, Robin Cristofari, Nicolas Chatelain, Michel Gauthier-Clerc, Jean-Paul Gendner, Yves Handrich, Charles-Edouard Salmon, Daniel P. Zitterbart and Céline Le Bohec conceived the ideas and designed methodology. Gaël Bardon, Robin Cristofari, Téo Barracho, Marine Benoiste, Nicolas Chatelain, Flávia A. N. Fernandes, Michel Gauthier-Clerc, Aymeric Houstin, Adélie Krellenstein, Nicolas Lecomte, Charles-Edouard Salmon, Emiliano Trucchi, Benoit Vallas and Céline Le Bohec collected the data. Gaël Bardon, Robin Cristofari, Téo Barracho, Marine Benoiste, Claire Ceresa, Flávia A. N. Fernandes, Michel Gauthier-Clerc, Aymeric Houstin, Charles-Edouard Salmon, Benoit Vallas and Céline Le Bohec cured and/or analysed the data. Gaël Bardon, Robin Cristofari, Alexander Winterl, Claire Ceresa, Julien Courtecuisse, Aymeric Houstin, Charles-Edouard Salmon, Emily M. Wong, Daniel P. Zitterbart and Céline Le Bohec wrote codes for RFIDeep and software. Robin Cristofari, Nicolas Chatelain, Julien Courtecuisse, Michel Gauthier-Clerc, Jean-Paul Gendner, Yves Handrich, Daniel P. Zitterbart and Céline Le Bohec developed the hardware. Project administration and supervision was done by Céline Le Bohec. Gaël Bardon and Céline Le Bohec led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

This study was supported by the Institut Polaire Français Paul-Emile Victor (IPEV) within the framework of the Project 137-ANTAVIA,

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitäts-Bibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

by the Centre Scientifique de Monaco with additional support from the LIA-647 and RTPI-NUTRESS (CSM/CNRS-UNISTRA), by the Centre National de la Recherche Scientifique (CNRS) through the Programme Zone Atelier Antarctique et Terres Australes (ZATA), by the Deutsche Forschungsgemeinschaft (DFG) grants ZI1525/3-1 and ZI1525/7-1, and by the Academy of Finland grant #331320. We are deeply grateful to all the wintering and summering members of Project 137, Denis Allemand, Benjamin Friess, Yvon Le Maho, Victor Planas-Bielsa, Claire Saraux and all the other colleagues and students within the team who have contributed over the last 20 years to the improvement of hardware, software and databases. We also sincerely thank the IPEV logistics teams for their important and continued support in the field. We are grateful to two reviewers for their constructive comments.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14187.

DATA AVAILABILITY STATEMENT

RFIDeep and Sphenotron codes are accessible from Zenodo repository: https://doi.org/10.5281/zenodo.7986367 (Bardon & Le Bohec, 2023).

ORCID

Gaël Bardon https://orcid.org/0000-0002-0243-9178
Robin Cristofari https://orcid.org/0000-0001-8430-7453
Alexander Winterl https://orcid.org/0000-0003-0688-9317
Téo Barracho https://orcid.org/0000-0002-8742-4062
Nicolas Chatelain https://orcid.org/0000-0002-1346-3595
Julien Courtecuisse https://orcid.org/0000-0002-0179-379X
Flávia A. N. Fernandes https://orcid.org/0000-0001-7897-6976
Yves Handrich https://orcid.org/0000-0002-4478-5562
Aymeric Houstin https://orcid.org/0000-0001-8676-819X
Nicolas Lecomte https://orcid.org/0000-0002-8473-5375
Emiliano Trucchi https://orcid.org/0000-0002-1270-5273
Emily M. Wong https://orcid.org/0000-0003-1445-1614
Daniel P. Zitterbart https://orcid.org/0000-0001-9429-4350
Céline Le Bohec https://orcid.org/0000-0003-0149-6477

REFERENCES

Afanasyev, V., Buldyrev, S. V., Dunn, M. J., Robst, J., Preston, M., Bremner, S. F., Briggs, D. R., Brown, R., Adlard, S., & Peat, H. J. (2015). Increasing accuracy: A new design and algorithm for automatically measuring weights, travel direction and radio frequency identification (RFID) of penguins. *PLoS One*, *10*, e0126292.

Ahsan, M. M., Alam, E. T., Trafalis, T., & Huebner, P. (2020). Deep MLP-CNN model using mixed-data to distinguish between COVID-19 and non-COVID-19 patients. *Symmetry*, *12*, 1526.

Austad, M., Oppel, S., Crymble, J., Greetham, H., Sahin, D., Lago, P., Metzger, B., & Quillfeldt, P. (2023). The effects of temporally distinct light pollution from ships on nocturnal colony attendance in a threatened seabird. *Journal of Ornithology*, *164*, 527–536.

Bambini, G., Schlicht, E., & Kempenaers, B. (2019). Patterns of female nest attendance and male feeding throughout the incubation period in blue tits *Cyanistes caeruleus*. *Ibis*, *161*, 50–65.

Bandivadekar, R. R., Pandit, P. S., Sollmann, R., Thomas, M. J., Logan, S. M., Brown, J. C., Klimley, A. P., & Tell, L. A. (2018). Use of rfid technology to characterize feeder visitations and contact network of hummingbirds in urban habitats. *PLoS One*, *13*, e0208057.

Bardon, G., & Le Bohec, C. (2023). Codes for: "RFIDeep: Unfolding the potential of deep learning for radio-frequency identification". *Zenodo*.

Bodey, T. W., Cleasby, I. R., Bell, F., Parr, N., Schultz, A., Votier, S. C., & Bearhop, S. (2018). A phylogenetically controlled meta-analysis of biologging device effects on birds: Deleterious effects and a call for more standardized reporting of study data. *Methods in Ecology and Evolution*, *9*, 946–955.

Bonter, D. N., & Bridge, E. S. (2011). Applications of radio frequency identification (RFID) in ornithological research: A review. *Journal of Field Ornithology*, *82*, 1–10.

Borowiec, M. L., Dikow, R. B., Frandsen, P. B., McKeeken, A., Valentini, G., & White, A. E. (2022). Deep learning as a tool for ecology and evolution. *Methods in Ecology and Evolution*, *13*, 1640–1660.

Brisson-Curadeau, É., Handrich, Y., Elliott, K. H., & Bost, C. A. (2021). Accelerometry predicts prey-capture rates in the deep-diving king penguin aptenodytes patagonicus. *Marine Biology*, *168*, 1–10.

Chiaradia, A. F., & Kerry, K. R. (1999). Daily nest attendance and breeding performance in the little penguin *Eudyptula minor* at Phillip Island, Australia. *Marine Ornithology*, *27*, 13–20.

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, *10*, 1632–1644.

Cristofari, R., Liu, X., Bonadonna, F., Cherel, Y., Pistorius, P., Le Maho, Y., Raybaud, V., Stenseth, N. C., Le Bohec, C., & Trucchi, E. (2018). Climate-driven range shifts of the king penguin in a fragmented ecosystem. *Nature Climate Change*, *8*, 245–251.

De Villemereuil, P., Charmantier, A., Arlt, D., Bize, P., Brekke, P., Brouwer, L., Cockburn, A., Côté, S. D., Dobson, F. S., Evans, S. R., Festa-Bianchet, M., Gamelon, M., Hamel, S., Hegelbach, J., Jerstad, K., Kempenaers, B., Kruuk, L., Kumpula, J., Kvalnes, T., … Chevin, L.-M. (2020). Fluctuating optimum and temporally variable selection on breeding date in birds and mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 31969–31978.

Descamps, S., Gauthier-Clerc, M., Gendner, J. P., & Le Maho, Y. (2002). The annual breeding cycle of unbanded king penguins aptenodytes patagonicus on possession Island (crozet). *Avian Science*, *2*, 87–98.

Dexter, C., Appleby, R., Edgar, J., Scott, J., & Jones, D. (2016). Using complementary remote detection methods for retrofitted ecopassages: A case study for monitoring individual koalas in south-East Queensland. *Wildlife Research*, *43*, 369.

Fagerstone, K. A., & Johns, B. E. (1987). Transponders as permanent identification markers for domestic ferrets, black-footed ferrets, and other wildlife. *The Journal of Wildlife Management*, *51*, 294–297.

Fannjiang, C., Mooney, T. A., Cones, S., Mann, D., Shorter, K. A., & Katija, K. (2019). Augmenting biologging with supervised machine learning to study in situ behavior of the medusa chrysaora fuscescens. *Journal of Experimental Biology*, *222*, jeb207654.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, *27*, 861–874.

Gauthier-Clerc, M., Gendner, J. P., Ribic, C., Fraser, W. R., Woehler, E. J., Descamps, S., Gilly, C., Le Bohec, C., & Le Maho, Y. (2004). Long-term effects of flipper bands on penguins. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*, S423–S426.

Gendner, J. P., Gauthier-Clerc, M., Le Bohec, C., Descamps, S., & Le Maho, Y. (2005). A new application for transponders in studying penguins. *Journal of Field Ornithology*, *76*, 138–142.

Gibbons, W. J., & Andrews, K. M. (2004). Pit tagging: Simple technology at its best. *Bioscience*, *54*, 447–454.

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä̈tsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Griffiths, R., Double, M. C., Orr, K., & Dawson, R. J. (1998). A dna test to sex most birds. *Molecular Ecology*, *7*, 1071–1075.

Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*, 82–97.

Horswill, C., Matthiopoulos, J., Green, J. A., Meredith, M. P., Forcada, J., Peat, H., Preston, M., Trathan, P. N., & Ratcliffe, N. (2014). Survival in macaroni penguins and the relative importance of different drivers: Individual traits, predation pressure and environmental variability. *Journal of Animal Ecology*, *83*, 1057–1067.

Hughes, E. J., Mady, R. P., & Bonter, D. N. (2021). Evaluating the accuracy and biological meaning of visits to rfid-enabled bird feeders using video. *Ecology and Evolution*, *11*, 17132–17141.

Iserbyt, A., Griffioen, M., Borremans, B., Eens, M., & Müller, W. (2018). How to quantify animal activity from radio-frequency identification (RFID) recordings. *Ecology and Evolution*, *8*, 10166–10174.

Jeantet, L., Vigon, V., Geiger, S., & Chevallier, D. (2021). Fully convolutional neural network: A solution to infer animal behaviours from multi-sensor data. *Ecological Modelling*, *450*, 109555.

Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236.

Kälin, U., Lang, N., Hug, C., Gessler, A., & Wegner, J. D. (2019). Defoliation estimation of forest trees from ground-level images. *Remote Sensing of Environment*, *223*, 143–153.

Kobayashi, K., Masuda, K., Haga, C., Matsui, T., Fukui, D., & Machimura, T. (2021). Development of a species identification system of japanese bats from echolocation calls using convolutional neural networks. *Ecological Informatics*, *62*, 101253.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, *60*, 84–90.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling* (Vol. *26*). Springer.

Le Bohec, C., Durant, J. M., Gauthier-Clerc, M., Stenseth, N. C., Park, Y. H., Pradel, R., Gremillet, D., Gendner, J. P., & Le Maho, Y. (2008). King penguin population threatened by southern ocean warming. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 2493–2497.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*, 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*, 2278–2324.

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv preprint arXiv:1801.00631*.

Milošević, D., Milosavljević, A., Predić, B., Medeiros, A. S., Savić-Zdravković, D., Piperac, M. S., Kostić, T., Spasić, F., & Leese, F. (2020). Application of deep learning in aquatic bioassessment: Towards automated identification of non-biting midges. *Science of the Total Environment*, *711*, 135160.

Molet, M., Chittka, L., Stelzer, R. J., Streit, S., & Raine, N. E. (2008). Colony nutritional status modulates worker responses to foraging recruitment pheromone in the bumblebee bombus terrestris. *Behavioral Ecology and Sociobiology*, *62*, 1919–1926.

Mutegeki, R., & Han, D. S. (2020). A cnn-lstm approach to human activity recognition. In: *2020 International conference on artificial intelligence in information and communication (ICAIIC)* (pp. 362–366). IEEE.

Park, J., Kim, D. I., Choi, B., Kang, W., & Kwon, H. W. (2020). Classification and morphological analysis of vector mosquitoes using deep convolutional neural networks. *Scientific Reports*, *10*, 1012.

Pawara, P., Okafor, E., Schomaker, L., & Wiering, M. (2017). Data augmentation for plant classification. In: *Advanced concepts for intelligent vision systems: 18th international conference, ACIVS 2017, Antwerp, Belgium, September 18–21, 2017, Proceedings 18* (pp. 615–626). Springer.

Pichler, M., & Hartig, F. (2022). Machine learning and deep learning—A review for ecologists. *Methods in Ecology and Evolution*, *14*, 994–1016.

Powers, D. M. (2020). Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*.

Purwaningrum, M., Nugroho, H. A., Asvan, M., Karyanti, K., Alviyanto, B., Kusuma, R., & Haryanto, A. (2019). Molecular techniques for sex identification of captive birds. *Veterinary World*, *12*, 1506–1513.

Rafiq, K., Appleby, R. G., Edgar, J. P., Radford, C., Smith, B. P., Jordan, N. R., Dexter, C. E., Jones, D. N., Blacker, A. R., & Cochrane, M. (2021). Wildwid: An open-source active rfid system for wildlife research. *Methods in Ecology and Evolution*, *12*, 1580–1587.

Sakamoto, K. Q., Sato, K., Ishizuka, M., Watanuki, Y., Takahashi, A., Daunt, F., & Wanless, S. (2009). Can ethograms be automatically generated using body acceleration data from free-ranging birds? *PLoS One*, *4*, e5379.

Schooley, R. L., Van Horne, B., & Burnham, K. P. (1993). Passive integrated transponders for marking free-ranging townsend's ground squirrels. *Journal of Mammalogy*, *74*, 480–484.

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).

Shiu, Y., Palmer, K., Roch, M. A., Fleishman, E., Liu, X., Nosal, E. M., Helble, T., Cholewiak, D., Gillespie, D., & Klinck, H. (2020). Deep neural networks for automated detection of marine mammal species. *Scientific Reports*, *10*, 607.

Stowell, D., Morfi, V., & Gill, L. F. (2016). Individual identity in songbirds: Signal representations and metric learning for locating the information in complex corvid calls. *arXiv preprint arXiv:1603.07236*.

Taylor, L., & Nitschke, G. (2018). Improving deep learning with generic data augmentation. In: *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 1542–1547). IEEE.

van Harten, E., Reardon, T., Lumsden, L., Meyers, N., Prowse, T., Weyland, J., & Lawrence, R. (2019). High detectability with low impact: Optimizing large pit tracking systems for cave-dwelling bats. *Ecology and Evolution*, *9*, 10916–10928.

Visser, M. E., Lindner, M., Gienapp, P., Long, M. C., & Jenouvrier, S. (2021). Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (*Parus major*). *Proceedings of the Royal Society B*, *288*, 20211337.

Wang, G. (2019). Machine learning for inferring animal behavior from location and movement data. *Ecological Informatics*, *49*, 69–76.

Williams, H. J., Taylor, L. A., Benhamou, S., Bijleveld, A. I., Clay, T. A., de Grissac, S., Demšar, U., English, H. M., Franconi, N., Gómez-Laich, A., Griffiths, R. C., Kay, W. P., Morales, J. M., Potts, J. R., Rogerson, K. F., Rutz, C., Spelt, A., Trevail, A. M., Wilson, R. P., & Börger, L. (2020). Optimizing the use of biologgers for movement ecology research. *Journal of Animal Ecology*, *89*, 186–206.

Zangmeister, J. L., Haussmann, M. F., Cerchiara, J., & Mauck, R. A. (2009). Incubation failure and nest abandonment by leach's storm-petrels detected using pit tags and temperature loggers. *Journal of Field Ornithology*, *80*, 373–379.

Zárybnická, M., Kubizňák, P., Šindelář, J., & Hlaváč, V. (2016). Smart nest box: A tool and methodology for monitoring of cavity-dwelling animals. *Methods in Ecology and Evolution*, *7*, 483–492.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In: *Computer vision–ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part I 13* (pp. 818–833). Springer.

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitaetsbibliothek Der Fau, Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2921–2929).

Zydlewski, G. B., Horton, G., Dubreuil, T., Letcher, B., Casey, S., & Zydlewski, J. (2006). Remote monitoring of fish in small streams. *Fisheries*, *31*, 492–502.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Supplementary Materials.

How to cite this article: Bardon, G., Cristofari, R., Winterl, A., Barracho, T., Benoiste, M., Ceresa, C., Chatelain, N., Courtecuisse, J., Fernandes, F. A. N., Gauthier-Clerc, M., Gendner, J.-P., Handrich, Y., Houstin, A., Krellenstein, A., Lecomte, N., Salmon, C.-E., Trucchi, E., Vallas, B., Wong, E. M. … Le Bohec, C. (2023). RFIDeep: Unfolding the potential of deep learning for radio-frequency identification. *Methods in Ecology and Evolution*, *14*, 2814–2826. https://doi.org/10.1111/2041-210X.14187

2041210x, 2023, 11, Downloaded from https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.14187 by Universitä-tsbibliothek Der Fau , Wiley Online Library on [11/01/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License