

# Evaluating the method reproducibility of deep learning models in biodiversity research

Waqas Ahmed<sup>1,\*</sup>, Vamsi Krishna Kommineni<sup>1,2,3,\*</sup>,  
Birgitta König-Ries<sup>1,2,4</sup>, Jitendra Gaikwad<sup>1</sup>, Luiz Gadelha<sup>1</sup> and  
Sheeba Samuel<sup>1,4</sup>

<sup>1</sup> Heinz Nixdorf Chair of Distributed Information Systems, Friedrich-Schiller Universität Jena, Jena, Thuringia, Germany

<sup>2</sup> German Centre for Integrative Biodiversity Research (iDiv), Leipzig, Sachsen, Germany

<sup>3</sup> Department of Functional Biogeography, Max Planck Institute for Biogeochemistry, Jena, Thuringia, Germany

<sup>4</sup> Michael Stifel Center Jena, Jena, Thuringia, Germany

\* These authors contributed equally to this work.

## ABSTRACT

Artificial intelligence (AI) is revolutionizing biodiversity research by enabling advanced data analysis, species identification, and habitats monitoring, thereby enhancing conservation efforts. Ensuring reproducibility in AI-driven biodiversity research is crucial for fostering transparency, verifying results, and promoting the credibility of ecological findings. This study investigates the reproducibility of deep learning (DL) methods within the biodiversity research. We design a methodology for evaluating the reproducibility of biodiversity-related publications that employ DL techniques across three stages. We define ten variables essential for method reproducibility, divided into four categories: resource requirements, methodological information, uncontrolled randomness, and statistical considerations. These categories subsequently serve as the basis for defining different levels of reproducibility. We manually extract the availability of these variables from a curated dataset comprising 100 publications identified using the keywords provided by biodiversity experts. Our study shows that a dataset is shared in 50% of the publications; however, a significant number of the publications lack comprehensive information on deep learning methods, including details regarding randomness.

**Subjects** Biodiversity, Statistics, Data Mining and Machine Learning, Data Science

**Keywords** Reproducibility, Deep learning, Metadata, Biodiversity

## INTRODUCTION

Biodiversity research investigates the role of biological diversity, underlying mechanisms, and means to protect it (see e.g., <https://www.idiv.de/research/>). This interdisciplinary research is critical for developing effective conservation strategies and guiding policy decisions that mitigate ecosystem disruptions. In recent years, deep learning methods have been increasingly applied in many scientific fields, including biodiversity research.

These methods have the potential to process and analyze large amounts of biological data rapidly, leading to significant insights. For instance, [August et al. \(2020\)](#) demonstrated how artificial intelligence (AI) image classifiers can create new biodiversity datasets from

Submitted 11 July 2024

Accepted 28 November 2024

Published 5 February 2025

Corresponding author

Waqas Ahmed,  
ahmed.waqas@uni-jena.de

Academic editor

Dan Stowell

Additional Information and  
Declarations can be found on  
page 16

DOI 10.7717/peerj-cs.2618

© Copyright

2025 Ahmed et al.

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

social media imagery, highlighting the spatial and taxonomic biases that can influence ecological inference. Similarly, deep learning (DL) models have been utilized to analyze camera trap images for wildlife monitoring, enabling researchers to identify species and infer ecological patterns and processes with high accuracy in [Tabak et al. \(2019\)](#). However, there is a growing concern about the reproducibility, transparency, and trustworthiness of research findings produced using deep learning methods in this domain ([Feng et al., 2019](#); [GPAI, 2022](#)).

Reproducibility is essential in scientific research as it allows researchers to validate and advance methods and results, ensuring the reliability of scientific claims ([Samuel & König-Ries, 2021](#)). [Goodman, Fanelli & Ioannidis \(2016\)](#) define research reproducibility in three aspects: methods reproducibility, results reproducibility, and inferential reproducibility. In this article, we focus specifically on methods reproducibility within the deep learning analytical workflow, which involves replicating experimental and computational processes using the same data and tools to achieve consistent outcomes. While issues related to experimental design reproducibility, such as those in field studies, are important, they are outside the scope of this work. For instance, in natural studies like [Rovero et al. \(2013\)](#), which deploys camera traps for consistent data collection across different geographical locations, it is inherently challenging to achieve exact reproducibility due to factors such as variations in wildlife behavior or environmental conditions that affect the collected data. On the other hand, our focus remains on ensuring reproducibility within computational methods that are more controlled, such as deep learning pipelines. For example, [Norouzzadeh et al. \(2018\)](#) demonstrated the application of deep learning for automatically identifying and counting wildlife in camera trap images, which is critical for monitoring species populations and informing conservation strategies. However, failure to reproduce these methods accurately could lead to erroneous population estimates, thereby compromising conservation efforts and resource allocation. Furthermore, [Christin, Hervet & Lecomte \(2019\)](#) discuss how the reproducibility of deep learning models in ecological research is often challenged by the complexity and heterogeneity of biodiversity data, which includes interactions between variables, missing values, and non-linear patterns. Ensuring reproducibility is key to uncovering these complexities and improving the reliability of ecological models used for decision-making.

A deep learning pipeline is a structured sequence of processes in training and deploying a DL model ([El-Amir & Hamdy, 2020](#)). The pipeline typically begins with data collection and preprocessing, involving tasks such as data cleaning, normalization, and transformation. Following data preprocessing, the next stage consists of designing and selecting an appropriate deep learning architecture, considering factors like model complexity and the nature of the problem. Subsequently, model training takes place, where the chosen architecture is trained on the preprocessed data using optimization algorithms and specific hyperparameter configurations. After training, the model is evaluated and fine-tuned using various performance metrics to ensure its effectiveness in solving the targeted problem. The best-performing model is run on the test data to ensure an unbiased evaluation of the model's predictive performance. Finally, the trained model is deployed for real-world applications or further refinement.

To ensure the reproducibility of the deep learning pipeline, comprehensive documentation is crucial at each stage. This includes detailed records of the data collection steps (including providing a persistent identifier for each data point), data preprocessing steps, such as the specific data transformation techniques applied and any data augmentation strategies employed. Additionally, it is vital to document the specifics of the chosen deep learning architecture, including the exact configurations and versions of the neural network layers utilized. Detailed notes on the hyperparameter values selected during the model training phase are essential, as well as the training convergence criteria and the optimization algorithms employed. Proper documentation of the evaluation metrics used and the testing dataset ensures the reproducibility of the model's performance assessment. Finally, maintaining a record of the software libraries, hardware, frameworks, and versions utilized throughout the pipeline aids in replicating the experimental setup accurately.

In this article, we aim to shed light on the current state of reproducibility of deep learning methods applied to biodiversity research. We conducted a systematic literature review to identify publications that use deep learning techniques in biodiversity research using keywords provided by biodiversity experts ([Abdelmageed et al., 2022](#)). We define various reproducibility-related variables inspired by the current literature. We then curated a dataset of 100 publications and manually extracted reproducibility-related variables, such as the availability of datasets and code. We also analyzed advanced reproducibility variables, such as the specific deep learning methods employed, models, and hyperparameters. Our findings suggest that the reproducibility of deep learning methods in biodiversity research is generally low. However, there is a growing trend towards improving reproducibility, with more and more publications making their datasets and code available. Our study will contribute to the ongoing discourse on the reproducibility of deep learning methods in biodiversity research and help to improve the credibility and impact of these methods in this vital field.

In the following sections, we provide a detailed description of our study. We start with an overview of the state-of-the-art ("Related Work"). We provide the methodology of our study ("Methodology") We describe our results and discuss the implications of our work ("Discussion"). Finally, we summarize the key aspects of our study and provide future directions of our research ("Conclusion").

## RELATED WORK

Method reproducibility, as mentioned in [Goodman, Fanelli & Ioannidis \(2016\)](#), is an important aspect of progress in any scientific field. [Raff \(2019\)](#) provides some insight into reproducibility in machine learning (ML). The author studied the reproducibility of 255 ML articles published from 1984 to 2017 and tried to correlate the reproducibility of the articles with their characteristics. According to his observation, the main obstacle to reproducibility of results is insufficient explanation of the method and implementation details when the source code is not provided. However, some of the defined attributes were very subjective, such as the algorithmic difficulty of the work or the readability of the work.

While [Raff \(2019\)](#) provides insight into ML reproducibility in general, there is also research specifically related to the field of biodiversity ([Feng et al., 2019](#); [Schnitzer & Carson, 2016](#)). They pointed out that reproducibility is problematic because erroneous data is widespread and there is a lack of empirical studies to verify previous research findings. Several authors have pointed out the need for better data management and reporting practices to ensure the reproducibility of research results ([Stark, 2018](#); [Waide, Brunt & Servilla, 2017](#); [Samuel, Löffler & König-Ries, 2021](#)). In [Waide, Brunt & Servilla \(2017\)](#), the authors discuss the challenges associated with managing different data sets in ecological research, while [Leonelli \(2018\)](#) further argues that reproducibility should be a criterion for assessing research quality. However, setting standards for assessing reproducibility itself is an important issue for research. [Gundersen, Shamsaliei & Isdahl \(2022\)](#) in his latest study defined 22 variables that could be categorized according to different degrees of reproducibility. Three categories were mentioned: reproducible data, reproducible experiment and reproducible method. [Gundersen, Shamsaliei & Isdahl \(2022\)](#) used these variables to evaluate reproducibility support on 13 different open source machine learning platforms. Similarly, [Heil et al. \(2021\)](#) proposed three different standards for reproducibility in machine learning that have applications in the life sciences. Bronze, silver or gold standards are defined according to the availability of data, models and code. [Tatman, VanderPlas & Dane \(2018\)](#) went a step further and analyzed ML articles from International Conference on Machine Learning (ICML) and NeurIPS conferences and distinguished levels of reproducibility based on the resources provided with the article. To achieve the desired result, they recommended some practical steps such as providing code and data in an executable environment where all libraries and dependencies are linked. This allows code or experiments to run smoothly on a new machine. This recommendation is in line with the gold standard provided in [Heil et al. \(2021\)](#) that authors should produce reproducible results with the execution of a single command. In addition to the general recommendations, [Pineau et al. \(2021\)](#) proposed standard practices and activities to improve reproducibility in the AI community. Some of these are the reproducibility programme at the NeurIPS conference and the ML reproducibility checklist, which provides guidelines to authors before submitting to conferences. Inspired by [Pineau et al. \(2021\)](#), some of the major AI conferences (AAAI and IJCAI) have introduced similar reproducibility checklists. We have considered these guidelines and also previous work in [Gundersen, Shamsaliei & Isdahl \(2022\)](#) to develop 10 reproducibility variables. However, the work in [Gundersen, Shamsaliei & Isdahl \(2022\)](#) is designed to assess the reproducibility of different ML platforms, whereas we aim to assess the reproducibility of biodiversity research. Our variables also include the aspect of uncontrolled randomness and the statistical information necessary for reproducibility.

## METHODOLOGY

In this section, we will discuss the steps of our work analyzing the reproducibility of research articles in biodiversity research.

## Identification

To assess the reproducibility of methods in biodiversity research, we first tried to obtain unbiased and relevant publications. To do this, we adopted keywords from [Abdelmageed et al. \(2022\)](#), which were provided by biodiversity experts and were also used to develop a corpus in the field of biodiversity for large language models ([Abdelmageed et al., 2022](#)). These keywords are: “biodivers\*” OR “genetic diversity” OR “\*omic diversity” OR “phylogenetic diversity” OR “population diversity” OR “species diversity” OR “ecosystem diversity” OR “functional diversity” OR “microbial diversity” OR “soil diversity” AND “Deep Learning”.

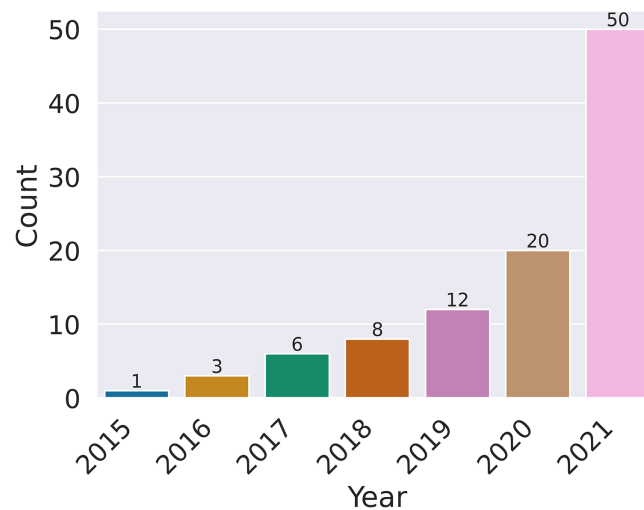
Using this search query, Google Scholar returned more than 8,000 articles from 2015 to 2021. Due to the impracticality of manually processing such a large dataset within the available time frame, we initially selected the first 100 articles and then randomly selected an additional 100 for further analysis. The first 100 articles were chosen with the belief that they would be statistically representative of the broader set. We compared the results of these first 100 articles with those of the randomly selected 100, which reinforced our belief, as the results were nearly identical. However, we acknowledge that this selection may not be entirely free from bias. Despite this, our sample covers a diverse range of publishers and publication years, ensuring a broad cross-section of the literature. Each selected publication was manually reviewed to extract information on ten key variables, with approximately 40 min dedicated to reviewing each article. This thorough manual curation aimed to enhance the reliability of our dataset. This manually curated dataset will serve as a benchmark for developing an automated system to extract variable information, enabling future analyses to encompass a larger number of articles while maintaining methodological rigor and reproducibility.

## Screening

Before analyzing the individual articles for reproducibility, we found that some articles could not be considered further for the following reasons: 1) duplication, 2) only an abstract was provided or the full article was not accessible, and 3) some articles were not empirical studies and therefore did not include experiments. After considering all these limitations, the number of articles was reduced to 100. [Figure 1](#) shows the distribution of articles over the period from 2015 to 2021, with most articles being from 2020 and 2021. [Figure 2](#) shows the publisher information for the considered articles, which are distributed quite randomly across more than ten publishers.

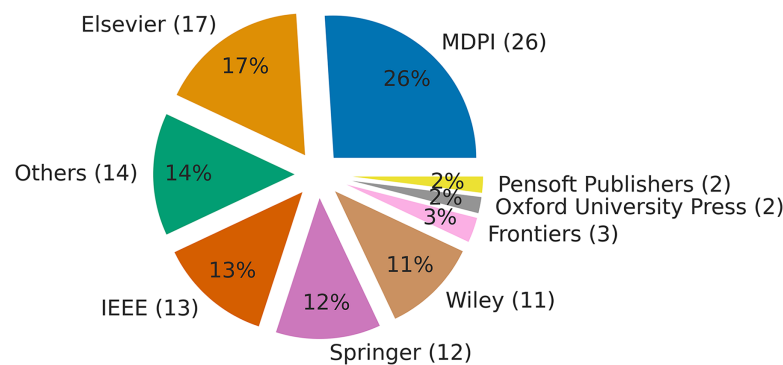
## Selection of reproducibility variables

There is no standard for assessing the reproducibility of DL biodiversity methods. A common practice for analyzing the reproducibility of research articles is to rerun the experiments using the same methodology as the original author. However, this requires a lot of time and computing resources. Also, sometimes it is not possible to work with the same resources such as hardware and software. Inspired by [Gundersen, Shamsaliei &](#)



**Figure 1** Number of publications analyzed for the presence of reproducibility variables, categorized by year.

Full-size DOI: [10.7717/peerj-cs.2618/fig-1](https://doi.org/10.7717/peerj-cs.2618/fig-1)



**Figure 2** Publisher information for the 100 publications selected for reproducibility analysis.

Full-size DOI: [10.7717/peerj-cs.2618/fig-2](https://doi.org/10.7717/peerj-cs.2618/fig-2)

*Isdahl (2022)*, we have formulated a set of 10 indicators that can be used as proxies for the probability that a result is reproducible. Instead of repeating each experiment, we look for these variables or factors that are considered important determinants of the reproducibility. Using the literature and the reproducibility checklists at conferences such as NeurIPS and AAAI (*Pineau et al., 2021*), we can divide these variables into four categories. Resource information (ReI) details the availability of datasets, source code, open-source tools, and proprietary model details crucial for reproduction efforts. Methodological information (MI) captures the specifics of software and hardware used, as well as a high-level overview of the deep learning methods employed. Randomness Information (RaI) addresses aspects of unpredictability in computational processes, ensuring they are documented for consistency. Statistical information (SI) focuses on the rigor of result analysis, advocating for multiple metrics and averaging techniques for



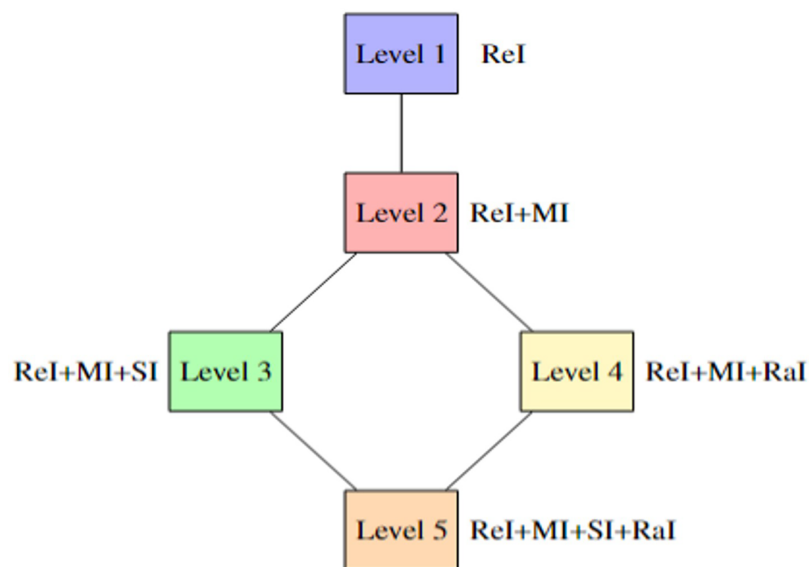
**Table 1** Definition of the various reproducibility variables used in this research article.

| Category                        | Variable                              | Variable short name | Variable description                                                                                                                                                                                     |
|---------------------------------|---------------------------------------|---------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Resource Information (ReI)      | Dataset                               | V1                  | Availability of dataset with persistent identifiers.                                                                                                                                                     |
|                                 | Source code                           | V2                  | Availability of source code to recreate the experiment (e.g., GitHub, GitLab, Zenodo).                                                                                                                   |
|                                 | Open source frameworks or environment | V3                  | Availability of open source software and hardware tools required to reproduce the work (e.g., docker container, virtual environments).                                                                   |
|                                 | Model architecture                    | V4                  | Availability of a deep learning model architecture or the accessibility of its internal working.                                                                                                         |
| Methodological Information (MI) | Software and Hardware Specification   | V5                  | Availability of information related to the type of hardware used (e.g., GPU), its specifications, and the version of software and libraries used.                                                        |
|                                 | Methods                               | V6                  | Availability of high-level information about the deep learning pipeline including the pre-processing and post-processing steps. The other parameters of the pipeline are assessed as separate variables. |
|                                 | Hyper-parameters                      | V7                  | Availability of hyperparameters used to train the model (e.g., number of epochs, learning rate, optimizer, etc.).                                                                                        |
| Randomness Information (RaI)    | Randomness                            | R                   | Availability of information about weight initialization, data shuffling, data augmentation, data train-test split.                                                                                       |
| Statistical Information (SI)    | Averaging result                      | S1                  | Availability of multiple model training and averaging them out rather than selecting only the highest value.                                                                                             |
|                                 | Evaluation metrics                    | S2                  | Availability of more than one evaluated metric (e.g., $R^2$ score along with Root Mean Square Error (RMSE) or Mean Absolute Error (MAE) or Loss of the model).                                           |

reliable evaluation. The comprehensive details that define these aspects of reproducibility are systematically itemized in [Table 1](#).

### Assessing the reproducibility of published articles

For all selected articles, we first manually checked the variables for each article. To reduce the degree of subjectivity, we (two of the authors) began by independently assessing the availability of each reproducibility variable. In the initial phase of our assessment, we encountered notable discrepancies due to ambiguous interpretations of the definitions of each variable, leading to inconsistent results from both authors. This was quantitatively evidenced by an average Cohen's Kappa value ([Cohen, 1960](#)) of 0.57, indicating a moderate level of agreement and highlighting the initial inconsistencies between annotators. To address these issues and improve the reliability of our analysis, we undertook a review and clarification of the variable definitions. In the end, we obtained the same responses for each article, as shown in [Table A1](#). These responses formed the basis for analysing reproducibility. We verified whether the article covered the functionality of each previously defined variable. After completing the assessment for each article, we quantified reproducibility across five distinct levels. These levels are designed to provide independent researchers with insight into the likelihood of successfully replicating the results. The higher the level of reproducibility, the more variables are accounted for, increasing the



**Figure 3** Number of levels along with the categories covered for respective levels.

Full-size DOI: 10.7717/peerj-cs.2618/fig-3

chances of reproducing the results with accuracy. The levels are structured to reflect the time and computational effort needed for reproducibility, especially when certain variables are missing. However, meeting the criteria for the basic level (Level 1) is essential for a study to be considered reproducible, as defined by [Goodman, Fanelli & Ioannidis \(2016\)](#). Levels beyond this build upon Level 1, offering increasingly rigorous criteria, but Level 1 remains the foundational requirement. [Figure 3](#) illustrates the different levels of reproducibility and the corresponding categories each covers.

- **Level 1** covers all variables defined in the ReI category, and a study is classified as *potentially reproducible* once this level is met.
- **Level 2** includes all variables from both the ReI and MI categories, classifying a study as *mostly reproducible* when these criteria are fulfilled.
- **Level 3** expands further, covering ReI, MI, and statistical information (SI).
- **Level 4** incorporates ReI, MI, and randomness information (RaI). These two levels do not necessarily build upon each other, offering additional insights without being required for basic reproducibility.
- **Level 5**, the highest level, encompasses all categories, ensuring the most comprehensive reproducibility criteria.

This tiered structure ensures flexibility and allows for varying degrees of reproducibility to be recognized. While Level 1 is the minimum threshold, Levels 2 and above provide additional layers of robustness, offering greater confidence in the study's replicability. We believe this approach balances thoroughness with practicality, while also accounting for



studies that may be classified as mostly reproducible without needing to meet every criterion at the higher levels.

## RESULTS

As discussed in “Discussion”, we divided the ten reproducibility variables we had defined into four categories: 1) resource information 2) methodological information 3) randomness information, and 4) statistical information. These four categories were further categorized into five different levels of reproducibility, allowing us to assess the reproducibility level of the publications more comprehensively.

### Variable level information

Figure 4 contains four plots, one for each category of information. Within each category, specific reproducibility variables are depicted with binary counts, illustrated by two side-by-side bars: one denoted by ‘Yes’ in a brick orange color and the other by ‘No’ in blue, along with the corresponding number of publications. In total, we made 1,000 individual judgements for the ten variables, 579 of these were positive *i.e* the respective information was provided, while this was not the case in 421 cases resulting in negative judgements. Detailed descriptions of all responses can be found at both the category and individual variable levels.

In the resources information category, 50 publications provided a dataset, 30 indicated the code repositories, 88 used open-source frameworks or environments, and 96 mentioned the model architectures (Fig. 4A).

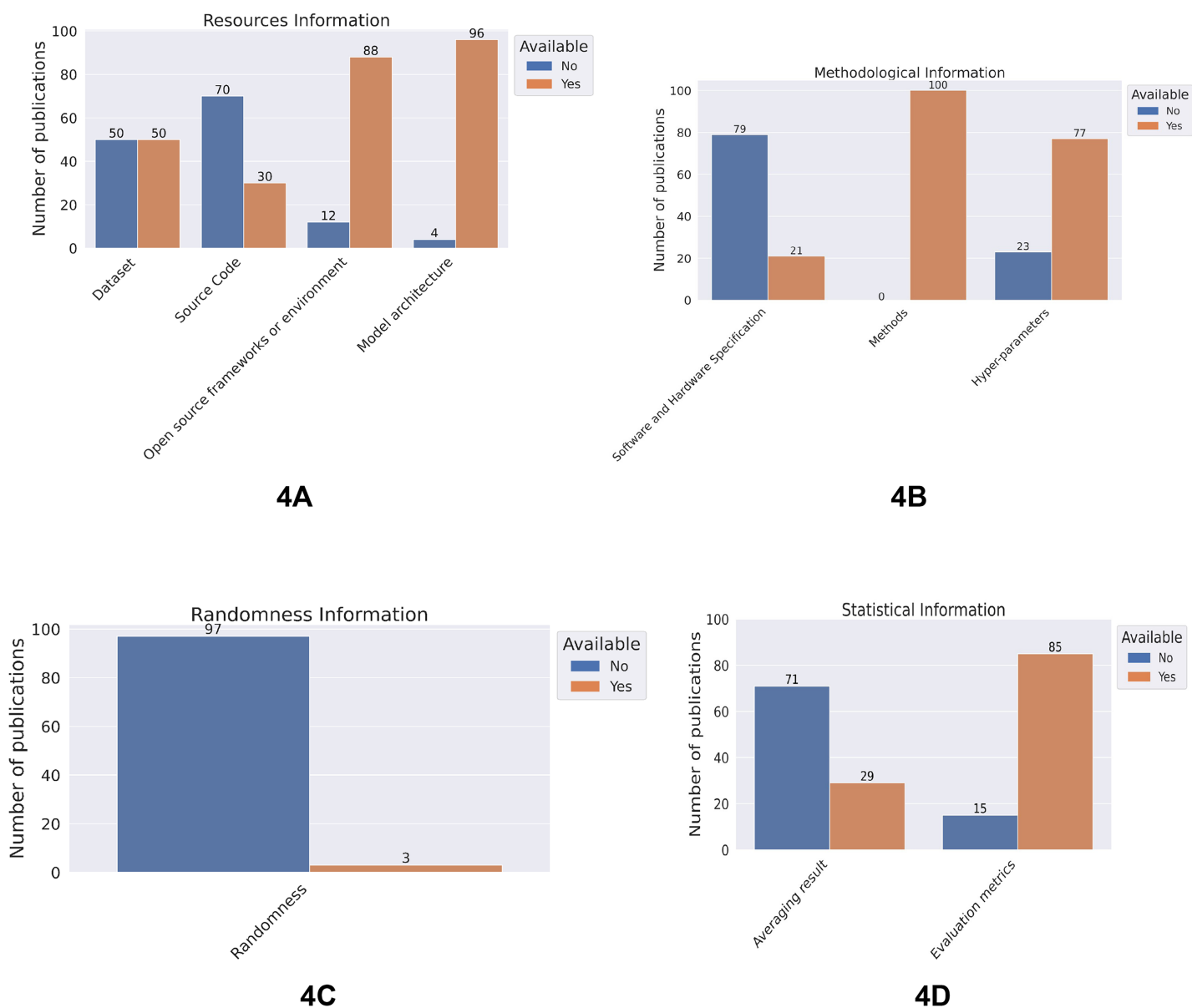
For the methodological information category, around a quarter of the publications included details about the hardware and the software (libraries) they used, all the publications explained methods that were used to build a machine learning pipeline and 77 publications provided the basic hyperparameters (Fig. 4B).

The number of publications that used a random seed in all possible ways (weight initialization, data shuffling, data augmentation, data train-test split and cuDNN GPU library) in their code is 3 (Fig. 4C). Regarding statistical considerations, 71 publications provided the average result of multiple model trainings, 85 publications evaluated their models with more than one evaluation metric. (Fig. 4D).

### Categorical level information

We have harmonized the presence of reproducibility variables for each categorical level in such a way that all the individual reproducibility variables in a category must be available (Yes) to mark that the specific categorical information is available (Yes). If one of the individual reproducibility variables is unavailable (No), the specific categorical information is unavailable (No). Mathematically,  $V1 \& V2 \& V3 \& V4 == \text{Yes}$  is used to denote the availability of resource information (Yes).

Table 2 shows the number of publications where all the variable information is available in a publication within the defined categories. The number of publications that satisfy the

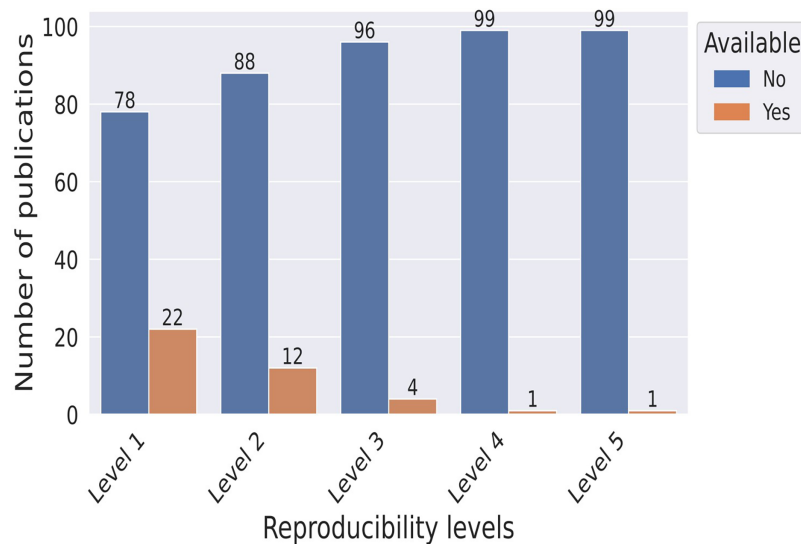


**Figure 4** Presence of the considered reproducible variables in four categories: (A) resources, (B) methodological information, (C) randomness, (D) statistical consideration for selected research publication.

Full-size DOI: 10.7717/peerj-cs.2618/fig-4

**Table 2** Number of selected research publications that contain all the variables information in a specific category along four categories.

| Category                   | Publication count |
|----------------------------|-------------------|
| Resource information       | 22                |
| Methodological information | 20                |
| Randomness information     | 3                 |
| Statistical information    | 26                |



**Figure 5** Bar plot indicating the number of publications satisfying the different levels of reproducibility. [Full-size DOI: 10.7717/peerj-cs.2618/fig-5](https://doi.org/10.7717/peerj-cs.2618/fig-5)

Resources, Methodological, Randomness, and Statistical information categories are 22, 20, 3, and 26, respectively.

### Reproducibility levels of publications

As per the definitions of the reproducibility levels described in “Methodology”, Levels 1 and 5 are the lowest and highest, respectively. According to Fig. 5, Only one publication meets the highest level of reproducibility, while 22 publications fulfill the basic criteria for being classified as “potentially reproducible.” Additionally, 12 publications meet the Level 2 criteria, covering both ReI and MI categories, and are classified as “mostly reproducible”.

### Reproducibility status of publications by year

The number of publications that meet the defined reproducibility criteria follows a linear trend with respect to year (Fig. 6).

## DISCUSSION

In biodiversity research, deep learning methods are becoming part of many studies that run large-scale experiments. These gave us the opportunity to orchestrate and extract the 10 reproducibility variables from 100 publications (Table 1 and Table A1). As a result, we recorded 1,000 total responses: 579 were positive, and 421 were negative. All the positive responses were dominated by four variables (230 responses): 1) Open source frameworks or environment, 2) Model architecture, 3) Methods, and 4) Evaluation metrics and the negative responses were dominated by only one variable: Randomness.

Most of the publications that employed deep learning models use open-source frameworks or environments like Tensorflow and PyTorch with the programming language Python/R, and they also provide model architectures either as a figure/table in the



**Figure 6** Number of articles meeting criteria for the four categories: (A) resources, (B) methodological information, (C) randomness, (D) statistical consideration for selected research publications by year.

Full-size DOI: [10.7717/peerj-cs.2618/fig-6](https://doi.org/10.7717/peerj-cs.2618/fig-6)

publication or described in the text with respective citations. Some of the publications used licensed programming languages like Matlab, comparatively, it is negligible. We looked for high-level information on the whole deep-learning pipeline in methods, all the publications provided compact information.

Most of the publications use more than one metric to evaluate their models, for example, when it comes to regression tasks, they use  $R^2$  score along with root mean square error (RMSE) or mean absolute error (MAE) or loss of the model, *etc.* The information from the publications is in compliance with reproducible workflow guidelines for the positively dominated variables.

We found datasets only in 50 publications; in the other 50 publications, there was no tangential information about the dataset. There were some cases where the authors linked to some data-providing websites to find the data that was used in the publication, but those

websites will change over time and finding the exact data that was used in the publication is not possible without providing persistent identifiers of the respective data points.

Source code is one of the fundamental variables of reproducibility. However, 30% of the publications provided their source code.

In 21 publications, authors have provided specifications about both software and hardware. Without specific information about the hardware and software, the reproducibility results will change because the random generators work differently with different hardware and software changes with each version.

Hyperparameters are the values that are chosen to control the model learning process. This means with each set of hyperparameters, the model will provide a varied range of results. However, information about basic hyperparameters (epochs, learning rate, optimizer and loss function) was missing from 23 publications. Reproducing the results of these 21 publications is difficult due to the missing essential hyperparameter information.

Averaging results is also an important aspect of reproducibility, after each training process, results will change slightly because of the certain random initializations through the deep learning pipeline. However, in our study, 71 publications did not report the multiple training results.

We opted for manually updating the variables by going through each publication and extracting the required information for the presence of reproducibility variables because we did not find a system or technique that could automatically extract this information from a publication. Since we are extracting the information about variables from publications manually, it is only possible to work with a small dataset, which is also the limitation of this study.

Due to recent developments with Large Language Models (LLMs), we are considering extracting the reproducible variables information from publications using LLMs from the year 2022 onwards (*Ahmed et al., 2023*; *Kommineni, König-Ries & Samuel, 2024*). This will allow us to implement our analysis on large-scale publications.

## CONCLUSION

In this article, we presented our pipeline for assessing the reproducibility of deep learning methods in biodiversity research. Inspired by the current state of the art, we established a comprehensive set of ten variables, categorized into four distinct groups, to effectively quantify the reproducibility of DL empirical research. Based on the defined categories, we documented the availability of each variable across 100 selected publications over the period from 2015 to 2021. From the total 1,000 responses for the 10 variables, 57.9% show the availability of the variables in the publication, while the remaining 42.1% are primarily characterized by the absence of randomness-related information. The highest and lowest reproducibility levels are satisfied by only one and ten publications, respectively. Given the use of deep learning to advance biodiversity research, improving reproducibility of the DL methods is crucial. Considering the limitations of the manual approach and the relatively small dataset analyzed until 2021, our future endeavors will focus on implementing a semi-automatic approach that leverages Large Language models for extracting information on reproducible variables from publications.

## APPENDICES

**Table A1** presence of different reproducibility variables. ‘y’ denotes the presence of variable information, while ‘n’ signifies the absence of variable information.

| Publication                                    | V1 | V2 | V3 | V4 | V5 | V6 | V7 | R | S1 | S2 |
|------------------------------------------------|----|----|----|----|----|----|----|---|----|----|
| <i>Klein, McKown &amp; Tershy, 2015</i>        | n  | n  | y  | n  | n  | y  | n  | n | n  | n  |
| <i>Khalighifar et al. (2021)</i>               | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Choe, Chi &amp; Thorne, 2021</i>            | n  | n  | y  | y  | y  | y  | y  | n | y  | y  |
| <i>Mahmood et al. (2016)</i>                   | y  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Younis et al. (2020)</i>                    | y  | y  | y  | y  | y  | y  | y  | n | y  | y  |
| <i>Schwartz &amp; Alfaro (2021)</i>            | y  | y  | y  | y  | y  | y  | y  | n | y  | y  |
| <i>Chen et al. (2020)</i>                      | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Fujisawa et al. (2023)</i>                  | y  | y  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Weinstein (2018)</i>                        | n  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Chalmers et al. (2021)</i>                  | n  | n  | y  | y  | y  | y  | n  | n | n  | y  |
| <i>Guirado et al. (2019)</i>                   | y  | n  | y  | y  | y  | y  | y  | n | y  | y  |
| <i>Zuolkernan et al. (2020)</i>                | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Villon et al. (2018)</i>                    | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Fairbrass et al. (2019)</i>                 | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Weinstein et al. (2019)</i>                 | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Hu et al. (2020)</i>                        | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Alshahrani et al. (2021)</i>                | y  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Villon et al. (2020)</i>                    | n  | n  | y  | y  | n  | y  | n  | n | y  | y  |
| <i>Botella et al. (2018)</i>                   | y  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Salamon et al. (2017)</i>                   | y  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Mac Aodha et al. (2018)</i>                 | y  | y  | y  | y  | y  | y  | y  | n | y  | y  |
| <i>Schindler &amp; Steinhage (2021)</i>        | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Bjerger et al. (2021)</i>                   | n  | y  | y  | y  | n  | y  | y  | y | y  | y  |
| <i>Guirado et al. (2017)</i>                   | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Hussein et al. (2021a)</i>                  | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Rammer &amp; Seidl (2019)</i>               | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Anand et al. (2021)</i>                     | n  | n  | n  | y  | n  | y  | n  | n | n  | y  |
| <i>Zizka et al. (2021)</i>                     | y  | y  | y  | y  | n  | y  | y  | y | n  | n  |
| <i>Demertzis, Iliadis &amp; Anezakis, 2018</i> | n  | n  | y  | y  | n  | y  | n  | n | y  | y  |
| <i>Malerba, Wright &amp; Macreadie, 2021</i>   | y  | n  | y  | y  | n  | y  | y  | n | n  | n  |
| <i>Huang &amp; Basanta (2021)</i>              | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Campos-Taberner et al. (2020)</i>           | y  | n  | n  | y  | n  | y  | n  | n | n  | y  |
| <i>Rousset et al. (2021)</i>                   | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>López-Jiménez et al. (2019)</i>             | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Schuettpelz et al. (2017)</i>               | n  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Schiller et al. (2021)</i>                  | y  | y  | y  | y  | y  | y  | y  | y | y  | y  |
| <i>Heredia (2017)</i>                          | n  | y  | y  | y  | n  | y  | y  | n | y  | y  |



**Table A1 (continued)**

| Publication                                            | V1 | V2 | V3 | V4 | V5 | V6 | V7 | R | S1 | S2 |
|--------------------------------------------------------|----|----|----|----|----|----|----|---|----|----|
| <i>Martins et al. (2021)</i>                           | y  | n  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Browning et al. (2018)</i>                          | y  | n  | y  | y  | n  | y  | n  | n | y  | y  |
| <i>Guirado et al. (2020)</i>                           | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Ayhan et al. (2020)</i>                             | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Loddo, Loddo &amp; Di Ruberto, 2021</i>             | n  | n  | y  | y  | n  | y  | n  | n | y  | y  |
| <i>Jamil, Rahman &amp; Haider, 2021</i>                | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Neves et al. (2021)</i>                             | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Jin et al. (2021)</i>                               | y  | n  | y  | n  | n  | y  | n  | n | n  | y  |
| <i>Mohanty, Hughes &amp; Salathé, 2016</i>             | y  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Xie et al. (2019)</i>                               | y  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Tian et al. (2020)</i>                              | n  | n  | n  | n  | n  | y  | n  | n | y  | y  |
| <i>Gimenez et al. (2022)</i>                           | n  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Ortega Adarme et al. (2020)</i>                     | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Dunker et al. (2021)</i>                            | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Villon et al. (2016)</i>                            | n  | n  | y  | y  | n  | y  | n  | n | n  | y  |
| <i>Dyrmann et al. (2021)</i>                           | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Arruda et al. (2021)</i>                            | y  | n  | y  | n  | n  | y  | y  | n | n  | n  |
| <i>Kislov &amp; Korznikov (2020)</i>                   | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Hussein et al. (2021b)</i>                          | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Becker et al. (2023)</i>                            | n  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Padubidri et al. (2021)</i>                         | y  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Carranza-Rojas et al. (2017)</i>                    | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Hamdi, Brandmeier &amp; Straub, 2019</i>            | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Ali et al. (2021)</i>                               | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Norouzzadeh et al. (2020)</i>                       | y  | y  | n  | y  | n  | y  | y  | n | n  | y  |
| <i>Sun et al. (2019)</i>                               | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Carranza-Rojas, Mata-Montero &amp; Goeau (2018)</i> | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Torney et al. (2019)</i>                            | y  | y  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Rowe et al. (2021)</i>                              | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Cota et al. (2021)</i>                              | y  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Hussain et al. (2021)</i>                           | y  | n  | y  | y  | n  | y  | y  | n | n  | n  |
| <i>Alkhelaiwi et al. (2021)</i>                        | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Norouzzadeh et al. (2018)</i>                       | n  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Zhou et al. (2021)</i>                              | y  | n  | y  | y  | y  | y  | y  | n | y  | n  |
| <i>Karar et al. (2021)</i>                             | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Ferreira et al. (2020)</i>                          | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Monedero et al. (2021)</i>                          | n  | n  | y  | y  | n  | y  | y  | n | n  | n  |
| <i>Gerovichev et al. (2021)</i>                        | y  | y  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Pu, Xv &amp; Deng, 2021</i>                         | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Tang, Zhang &amp; Zhao, 2021</i>                    | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |
| <i>Morgan &amp; Braasch (2021)</i>                     | n  | n  | y  | y  | n  | y  | y  | n | y  | y  |

(Continued)

**Table A1 (continued)**

| Publication                              | V1 | V2 | V3 | V4 | V5 | V6 | V7 | R | S1 | S2 |
|------------------------------------------|----|----|----|----|----|----|----|---|----|----|
| <i>Milosević et al. (2020)</i>           | n  | n  | y  | y  | n  | y  | n  | n | n  | n  |
| <i>August et al. (2020)</i>              | y  | y  | y  | y  | n  | y  | n  | n | n  | n  |
| <i>Wagner et al. (2020)</i>              | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Hass &amp; Jokar Arsanjani (2020)</i> | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>DeLancey et al. (2019)</i>            | y  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Spiesman et al. (2021)</i>            | n  | n  | n  | y  | n  | y  | y  | n | n  | y  |
| <i>Jiang et al. (2023)</i>               | y  | y  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Charles et al. (2021)</i>             | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Pouliot et al. (2021)</i>             | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Marre et al. (2020)</i>               | n  | n  | y  | y  | y  | y  | y  | n | n  | y  |
| <i>Radig et al. (2021)</i>               | n  | n  | n  | y  | n  | y  | n  | n | n  | y  |
| <i>Yubo et al. (2020)</i>                | y  | n  | n  | y  | n  | y  | y  | n | y  | y  |
| <i>Liang et al. (2020)</i>               | n  | n  | n  | y  | n  | y  | n  | n | n  | n  |
| <i>Banzi &amp; Abayo (2021)</i>          | y  | n  | y  | y  | n  | y  | y  | n | n  | n  |
| <i>Yang et al. (2021)</i>                | n  | n  | n  | y  | n  | y  | y  | n | n  | y  |
| <i>Villon et al. (2021)</i>              | n  | n  | n  | y  | n  | y  | n  | n | n  | n  |
| <i>Malik, Faisal &amp; Hussein, 2021</i> | y  | n  | n  | y  | n  | y  | y  | n | n  | y  |
| <i>Khalighifar et al. (2019)</i>         | y  | y  | y  | y  | n  | y  | n  | n | y  | n  |
| <i>Loulidi et al. (2022)</i>             | n  | n  | n  | y  | n  | y  | y  | n | n  | y  |
| <i>Xi et al. (2019)</i>                  | n  | n  | y  | y  | n  | y  | y  | n | y  | n  |
| <i>Kankane &amp; Kang (2021)</i>         | n  | n  | y  | y  | n  | y  | y  | n | n  | y  |
| <i>Lee et al. (2017)</i>                 | y  | y  | y  | y  | n  | y  | y  | n | n  | n  |

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This work was supported by the thurAI project funded by the Thuringia State and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, funded by the German Research Foundation (FZT 118, 202548816). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

Thuringia State and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig.

German Research Foundation (FZT 118, 202548816).

### Competing Interests

Luiz Gadelha is an Academic Editor for PeerJ Computer Science.

## Author Contributions

- Waqas Ahmed conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Vamsi Krishna Kommineni conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Birgitta König-Ries conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Jitendra Gaikwad conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Luiz Gadelha conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Sheeba Samuel conceived and designed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.

## Data Availability

The following information was supplied regarding data availability:

The code and data are available at GitHub and Zenodo:

- <https://github.com/fusion-jena/Reproduce-DLmethods-Biodiv>

- Kommineni, V. K., Ahmed, W., König-Ries, B., Gaikwad, J., Gadelha, L., & Samuel, S. (2024). Evaluating the method reproducibility of deep learning models in the biodiversity research [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13987177>.

## Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj-cs.2618#supplemental-information>.

## REFERENCES

- Abdelmageed N, Löffler F, Feddoul L, Algergawy A, Samuel S, Gaikwad J, Kazem A, König-Ries B. 2022. Biodivnere: Gold standard corpora for named entity recognition and relation extraction in the biodiversity domain. *Biodiversity Data Journal* 10: DOI 10.3897/BDJ.10.e89481.
- Ahmed W, Kommineni VK, Koenig-ries B, Samuel S. 2023. How reproducible are the results gained with the help of deep learning methods in biodiversity research? *Biodiversity Information Science and Standards* 7: DOI 10.3897/biss.7.112698.
- Ali A, Fathalla A, Salah A, Bekhit M, Eldesouky E. 2021. Marine data prediction: an evaluation of machine learning, deep learning, and statistical predictive models. *Computational Intelligence and Neuroscience* 2021(1):8551167 DOI 10.1155/2021/8551167.
- Alkhelaiwi M, Boulila W, Ahmad J, Koubaa A, Driss M. 2021. An efficient approach based on privacy-preserving deep learning for satellite image classification. *Remote Sensing (Basel)* 13(11):2221 DOI 10.3390/rs13112221.
- Alshahrani HM, Al-Wesabi FN, Al Duhayyim M, Nemri N, Kadry S, Alqaralleh BAY. 2021. An automated deep learning based satellite imagery analysis for ecology management. *Ecological Informatics* 66(6):101452 DOI 10.1016/j.ecoinf.2021.101452.

- Anand A, Pandey MK, Srivastava PK, Gupta A, Khan ML. 2021.** Integrating multi-sensors data for species distribution mapping using deep learning and envelope models. *Remote Sensing (Basel)* **13**(16):3284 DOI [10.3390/rs13163284](https://doi.org/10.3390/rs13163284).
- Arruda VLS, Piontekowski VJ, Alencar A, Pereira RS, Matricardi EAT. 2021.** An alternative approach for mapping burn scars using landsat imagery, Google Earth Engine, and deep learning in the Brazilian savanna. *Remote Sensing Applications Society and Environment* **22**(100472):100472 DOI [10.1016/j.rsase.2021.100472](https://doi.org/10.1016/j.rsase.2021.100472).
- August TA, Pescott OL, Joly A, Bonnet P. 2020.** AI naturalists might hold the key to unlocking biodiversity data in social media imagery. *Patterns* **1**(7):100116 DOI [10.1016/j.patter.2020.100116](https://doi.org/10.1016/j.patter.2020.100116).
- Ayhan B, Kwan C, Budavari B, Kwan L, Lu Y, Perez D, Li J, Skarlatos D, Vlachos M. 2020.** Vegetation detection using deep learning and conventional methods. *Remote Sensing (Basel)* **12**(15):2502 DOI [10.3390/rs12152502](https://doi.org/10.3390/rs12152502).
- Banzi J, Abayo T. 2021.** Plant species identification from leaf images using deep learning models (CNN-LSTM architecture). *Tanzania Journal of Forestry and Nature Conservation* **90**(3):93–103.
- Becker A, Russo S, Puliti S, Lang N, Schindler K, Wegner JD. 2023.** Country-wide retrieval of forest structure from optical and SAR satellite imagery with deep ensembles. *ISPRS Journal of Photogrammetry and Remote Sensing* **195**(11):269–286 DOI [10.1016/j.isprsjprs.2022.11.011](https://doi.org/10.1016/j.isprsjprs.2022.11.011).
- Bjerger K, Nielsen JB, Sepstrup MV, Helsing-Nielsen F, Høye TT. 2021.** An automated light trap to monitor moths (*Lepidoptera*) using computer vision-based tracking and deep learning. *Sensors (Basel)* **21**(2):343 DOI [10.3390/s21020343](https://doi.org/10.3390/s21020343).
- Botella C, Joly A, Bonnet P, Monestiez P, Munoz F. 2018.** A deep learning approach to species distribution modelling. In: *Multimedia Tools and Applications for Environmental & Biodiversity Informatics*. Cham: Springer International Publishing, 169–199.
- Browning E, Bolton M, Owen E, Shoji A, Guilford T, Freeman R. 2018.** Predicting animal behaviour using deep learning: GPS data alone accurately predict diving in seabirds. *Methods in Ecology and Evolution* **9**(3):681–692 DOI [10.1111/2041-210X.12926](https://doi.org/10.1111/2041-210X.12926).
- Campos-Taberner M, García-Haro FJ, Martínez B, Izquierdo-Verdiguier E, Atzberger C, Camps-Valls G, Gilabert MA. 2020.** Understanding deep learning in land use classification based on sentinel-2 time series. *Scientific Reports* **10**(1):17188 DOI [10.1038/s41598-020-74215-5](https://doi.org/10.1038/s41598-020-74215-5).
- Carranza-Rojas J, Goeau H, Bonnet P, Mata-Montero E, Joly A. 2017.** Going deeper in the automated identification of herbarium specimens. *BMC Evolutionary Biology* **17**(1):181 DOI [10.1186/s12862-017-1014-z](https://doi.org/10.1186/s12862-017-1014-z).
- Carranza-Rojas J, Mata-Montero E, Goeau H. 2018.** Hidden biases in automated image-based plant identification. In: *2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI)*. Piscataway: IEEE.
- Chalmers C, Fergus P, Wich S, Longmore SN. 2021.** Modelling animal biodiversity using acoustic monitoring and deep learning. In: *2021 International Joint Conference on Neural Networks (IJCNN)*. Piscataway: IEEE.
- Charles CP, Kim PHC, de Almeida AG, Do Nascimento EV, Da Rocha LGS, Vivaldini KCT. 2021.** Detection of invasive vegetation through UAV and deep learning. In: *2021 Latin American Robotics Symposium (LARS), 2021 Brazilian Symposium on Robotics (SBR), and 2021 Workshop on Robotics in Education (WRE)*. Piscataway: IEEE.

- Chen X, Zhao J, Chen Y-H, Zhou W, Hughes AC. 2020. Automatic standardized processing and identification of tropical bat calls using deep learning approaches. *Biological Conservation* 241(108269):108269 DOI 10.1016/j.biocon.2019.108269.
- Choe H, Chi J, Thorne JH. 2021. Mapping potential plant species richness over large areas with deep learning, MODIS, and species distribution models. *Remote Sensing (Basel)* 13(13):2490 DOI 10.3390/rs13132490.
- Christin S, Hervet É, Lecomte N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution* 10(10):1632–1644 DOI 10.1111/2041-210X.13256.
- Cohen J. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20(1):37–46 DOI 10.1177/001316446002000104.
- Cota G, Sagan V, Maimaitijiang M, Freeman K. 2021. Forest conservation with deep learning: a deeper understanding of human geography around the betampona nature reserve, Madagascar. *Remote Sensing (Basel)* 13(17):3495 DOI 10.3390/rs13173495.
- DeLancey ER, Simms JF, Mahdianpari M, Brisco B, Mahoney C, Kariyeva J. 2019. Comparing deep learning and shallow learning for large-scale wetland classification in alberta, canada. *Remote Sensing (Basel)* 12(1):2 DOI 10.3390/rs12010002.
- Demertzis K, Iliadis LS, Anezakis V-D. 2018. Extreme deep learning in biosecurity: the case of machine hearing for marine species identification. *Journal of Information and Telecommunication* 2(4):492–510 DOI 10.1080/24751839.2018.1501542.
- Dunker S, Motivans E, Rakosy D, Boho D, Mäder P, Hornick T, Knight TM. 2021. Pollen analysis using multispectral imaging flow cytometry and deep learning. *New Phytologist* 229(1):593–606 DOI 10.1111/nph.16882.
- Dyrmann M, Mortensen AK, Linneberg L, Høye TT, Bjerger K. 2021. Camera assisted roadside monitoring for invasive alien plant species using deep learning. *Sensors (Basel)* 21(18):6126 DOI 10.3390/s21186126.
- El-Amir H, Hamdy M. 2020. *Deep learning pipeline*. Berkeley: Apress.
- Fairbrass AJ, Firman M, Williams C, Brostow GJ, Titheridge H, Jones KE. 2019. CityNet—deep learning tools for urban ecoacoustic assessment. *Methods in Ecology and Evolution* 10(2):186–197 DOI 10.1111/2041-210X.13114.
- Feng X, Park DS, Walker C, Peterson AT, Merow C, Papeş M. 2019. A checklist for maximizing reproducibility of ecological niche models. *Nature Ecology & Evolution* 3(10):1382–1395 DOI 10.1038/s41559-019-0972-5.
- Ferreira AC, Silva LR, Renna F, Brandl HB, Renoult JP, Farine DR, Covas R, Doutrelant C. 2020. Deep learning-based methods for individual recognition in small birds. *Methods in Ecology and Evolution* 11(9):1072–1085 DOI 10.1111/2041-210X.13436.
- Fujisawa T, Noguerales V, Meramveliotakis E, Papadopoulou A, Vogler AP. 2023. Image-based taxonomic classification of bulk insect biodiversity samples using deep learning and domain adaptation. *Systematic Entomology*. 48(3):387–401 DOI 10.1111/syen.12583.
- Gerovichev A, Sadeh A, Winter V, Bar-Massada A, Keasar T, Keasar C. 2021. High throughput data acquisition and deep learning for insect ecoinformatics. *Frontiers in Ecology and Evolution* 9:1 DOI 10.3389/fevo.2021.600931.
- Gimenez O, Kervellec M, Fanjul J-B, Chaine A, Marescot L, Bollet Y, Duchamp C. 2022. Trade-off between deep learning for species identification and inference about predator-prey co-occurrence. *Computo*. Available at <https://arxiv.org/pdf/2108.11509>.
- Goodman SN, Fanelli D, Ioannidis JPA. 2016. What does research reproducibility mean? *Science Translational Medicine* 8(341):341ps12 DOI 10.1126/scitranslmed.aaf5027.

- GPAl. 2022. Biodiversity & Artificial Intelligence, Opportunities and Recommendations, Report, November 2022, Global Partnership on AI.
- Guirado E, Alcaraz-Segura D, Cabello J, Puertas-Ruiz S, Herrera F, Tabik S. 2020. Tree cover estimation in global drylands from space using deep learning. *Remote Sensing (Basel)* 12(3):343 DOI 10.3390/rs12030343.
- Guirado E, Tabik S, Alcaraz-Segura D, Cabello J, Herrera F. 2017. Deep-learning versus obia for scattered shrub detection with Google Earth imagery: ziziphus lotus as case study. *Remote Sensing* 9(12):1220 DOI 10.3390/rs9121220.
- Guirado E, Tabik S, Rivas ML, Alcaraz-Segura D, Herrera F. 2019. Whale counting in satellite and aerial images with deep learning. *Scientific Reports* 9(1):14259 DOI 10.1038/s41598-019-50795-9.
- Gundersen OE, Shamsaliei S, Isdahl RJ. 2022. Do machine learning platforms provide out-of-the-box reproducibility? *Future Generation Computer Systems* 126(7604):34–47 DOI 10.1016/j.future.2021.06.014.
- Hamdi ZM, Brandmeier M, Straub C. 2019. Forest damage assessment using deep learning on high resolution remote sensing data. *Remote Sensing (Basel)* 11(17):1976 DOI 10.3390/rs11171976.
- Hass FS, Jokar Arsanjani J. 2020. Deep learning for detecting and classifying ocean objects: application of YoloV3 for iceberg–ship discrimination. *ISPRS International Journal of Geo-Information* 9(12):758 DOI 10.3390/ijgi9120758.
- Heil BJ, Hoffman MM, Markowitz F, Lee S-I, Greene CS, Hicks SC. 2021. Reproducibility standards for machine learning in the life sciences. *Nature Methods* 18(10):1132–1135 DOI 10.1038/s41592-021-01256-7.
- Heredia I. 2017. Large-scale plant classification with deep neural networks. In: *Proceedings of the Computing Frontiers Conference*. New York, NY, USA: ACM.
- Hu J, Huang W, Su Y, Liu Y, Xiao P. 2020. BatNet++: a robust deep learning-based predicting models for calls recognition. In: *2020 5th International Conference on Smart Grid and Electrical Automation (ICSGEA)*. Piscataway: IEEE.
- Huang Y-P, Basanta H. 2021. Recognition of endemic bird species using deep learning models. *IEEE Access* 9:102975–102984 DOI 10.1109/ACCESS.2021.3098532.
- Hussain A, Barua B, Osman A, Abozariba R, Asyhari AT. 2021. Performance of MobileNetV3 transfer learning on handheld device-based real-time tree species identification. In: *2021 26th International Conference on Automation and Computing (ICAC)*. Piscataway: IEEE.
- Hussein BR, Malik OA, Ong W-H, Slik JWF. 2021a. Automated extraction of phenotypic leaf traits of individual intact herbarium leaves from herbarium specimen images using deep learning based semantic segmentation. *Sensors (Basel)* 21(13):4549 DOI 10.3390/s21134549.
- Hussein BR, Malik OA, Ong W-H, Slik JWF. 2021b. Reconstruction of damaged herbarium leaves using deep learning techniques for improving classification accuracy. *Ecological Informatics* 61(101243):101243 DOI 10.1016/j.ecoinf.2021.101243.
- Jamil S, Rahman M, Haider A. 2021. Bag of features (BoF) based deep learning framework for bleached corals detection. *Big Data and Cognitive Computing* 5(4):53 DOI 10.3390/bdcc5040053.
- Jiang Y, Balaban M, Zhu Q, Mirarab S. 2023. DEPP: deep learning enables extending species trees using single genes. *Systematic Biology* 72(1):17–34 DOI 10.1093/sysbio/syac031.
- Jin L, Yu J, Yuan X, Du X. 2021. Fish classification using DNA barcode sequences through deep learning method. *Symmetry (Basel)* 13(9):1599 DOI 10.3390/sym13091599.



- Kankane A, Kang D. 2021.** Detection of seashore debris with fixed camera images using computer vision and deep learning. In: *2021 6th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. Piscataway: IEEE.
- Karar ME, Alsunaydi F, Albusaymi S, Alotaibi S. 2021.** A new mobile application of agricultural pests recognition using deep learning in cloud computing system. *Alexandria Engineering Journal* **60**(5):4423–4432 DOI [10.1016/j.aej.2021.03.009](https://doi.org/10.1016/j.aej.2021.03.009).
- Khalighifar A, Brown RM, Goyes Vallejos J, Peterson AT. 2021.** Deep learning improves acoustic biodiversity monitoring and new candidate forest frog species identification (genus platymantis) in the Philippines. *Biodiversity and Conservation* **30**(3):643–657 DOI [10.1007/s10531-020-02107-1](https://doi.org/10.1007/s10531-020-02107-1).
- Khalighifar A, Komp E, Ramsey JM, Gurgel-Gonçalves R, Peterson AT. 2019.** Deep learning algorithms improve automated identification of chagas disease vectors. *Journal of Medical Entomology* **56**(5):1404–1410 DOI [10.1093/jme/tjz065](https://doi.org/10.1093/jme/tjz065).
- Kislov DE, Korznikov KA. 2020.** Automatic windthrow detection using very-high-resolution satellite imagery and deep learning. *Remote Sensing (Basel)* **12**(7):1145 DOI [10.3390/rs12071145](https://doi.org/10.3390/rs12071145).
- Klein DJ, McKown MW, Tershy BR. 2015.** Deep learning for large scale biodiversity monitoring. In: *Bloomberg Data for Good Exchange Conference*.
- Kommineni VK, König-Ries B, Samuel S. 2024.** From human experts to machines: An llm supported approach to ontology and knowledge graph construction. ArXiv DOI [10.48550/arXiv.2403.08345](https://doi.org/10.48550/arXiv.2403.08345).
- Lee SH, Chan CS, Mayo SJ, Remagnino P. 2017.** How deep learning extracts and learns leaf features for plant classification. *Pattern Recognition* **71**(4):1–13 DOI [10.1016/j.patcog.2017.05.015](https://doi.org/10.1016/j.patcog.2017.05.015).
- Leonelli S. 2018.** Rethinking reproducibility as a criterion for research quality. In: *Including a Symposium on Mary Morgan: Curiosity, Imagination, and Surprise*. Vol. 36. Leeds, England: Emerald Publishing Limited, 129–146.
- Liang J, Li P, Zhao H, Han L, Qu M. 2020.** Forest species classification of UAV hyperspectral image using deep learning. In: *2020 Chinese Automation Congress (CAC)*. Piscataway: IEEE.
- Loddo A, Loddo M, Di Ruberto C. 2021.** A novel deep learning based approach for seed image classification and retrieval. *Computers and Electronics in Agriculture* **187**(106269):106269 DOI [10.1016/j.compag.2021.106269](https://doi.org/10.1016/j.compag.2021.106269).
- Loulidi A, Houssa R, Buhl-Mortensen L, Zidane H, Rhinane H. 2022.** Automatic fish detection from different marine environments video using deep learning. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**(W3-2021):191–198 DOI [10.5194/isprs-archives-XLVI-4-W3-2021-191-2022](https://doi.org/10.5194/isprs-archives-XLVI-4-W3-2021-191-2022).
- López-Jiménez E, Vasquez-Gomez JI, Sanchez-Acevedo MA, Herrera-Lozada JC, Uriarte-Arcia AV. 2019.** Columnar cactus recognition in aerial images using a deep learning approach. *Ecological Informatics* **52**(3):131–138 DOI [10.1016/j.ecoinf.2019.05.005](https://doi.org/10.1016/j.ecoinf.2019.05.005).
- Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, Harder B, Kinsey L, Mead GR, Newson SE, Pandourski I, Parsons S, Russ J, Szodoray-Paradi A, Szodoray-Paradi F, Tilova E, Girolami M, Brostow G, Jones KE. 2018.** Bat detective—deep learning tools for bat acoustic signal detection. *PLOS Computational Biology* **14**(3):e1005995 DOI [10.1371/journal.pcbi.1005995](https://doi.org/10.1371/journal.pcbi.1005995).
- Mahmood A, Bennamoun M, An S, Sohel F, Boussaid F, Hovey R, Kendrick G, Fisher RB. 2016.** Automatic annotation of coral reefs using deep learning. In: *OCEANS, 2016 MTS/IEEE Monterey*. Piscataway: IEEE.



- Malerba ME, Wright N, Macreadie PI. 2021.** A continental-scale assessment of density, size, distribution and historical trends of farm dams using deep learning convolutional neural networks. *Remote Sensing (Basel)* **13**(2):319 DOI [10.3390/rs13020319](https://doi.org/10.3390/rs13020319).
- Malik OA, Faisal M, Hussein BR. 2021.** Ensemble deep learning models for fine-grained plant species identification. In: *2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. Piscataway: IEEE.
- Marre G, De Almeida Braga C, Ienco D, Luque S, Holon F, Deter J. 2020.** Deep convolutional neural networks to monitor coralligenous reefs: operationalizing biodiversity and ecological assessment. *Ecological Informatics* **59**(101110):101110 DOI [10.1016/j.ecoinf.2020.101110](https://doi.org/10.1016/j.ecoinf.2020.101110).
- Martins JAC, Nogueira K, Osco LP, Gomes FDG, Furuya DEG, Gonçalves WN, Sant'Ana DA, Ramos APM, Liesenberg V, dos Santos JA, de Oliveira PTS, Junior JM. 2021.** Semantic segmentation of tree-canopy in urban environment with pixel-wise deep learning. *Remote Sensing (Basel)* **13**(16):3054 DOI [10.3390/rs13163054](https://doi.org/10.3390/rs13163054).
- Milosević D, Milosavljević A, Predić B, Medeiros AS, Savić-Zdravković D, Stojković Piperac M, 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):135160 DOI [10.1016/j.scitotenv.2019.135160](https://doi.org/10.1016/j.scitotenv.2019.135160).
- Mohanty SP, Hughes DP, Salathé M. 2016.** Using deep learning for image-based plant disease detection. *Frontiers in Plant Science* **7**:1419 DOI [10.3389/fpls.2016.01419](https://doi.org/10.3389/fpls.2016.01419).
- Monedero Í, Barbancho J, Márquez R, Beltrán JF. 2021.** Cyber-physical system for environmental monitoring based on deep learning. *Sensors (Basel)* **21**(11):3655 DOI [10.3390/s21113655](https://doi.org/10.3390/s21113655).
- Morgan MM, Braasch J. 2021.** Long-term deep learning-facilitated environmental acoustic monitoring in the capital region of new york state. *Ecological Informatics* **61**(101242):101242 DOI [10.1016/j.ecoinf.2021.101242](https://doi.org/10.1016/j.ecoinf.2021.101242).
- Neves AK, Körting TS, Fonseca LMG, Soares AR, Girolamo-Neto CD, Heipke C. 2021.** Hierarchical mapping of Brazilian savanna (cerrado) physiognomies based on deep learning. *Journal of Applied Remote Sensing* **15**(4):044504 DOI [10.1117/1.JRS.15.044504](https://doi.org/10.1117/1.JRS.15.044504).
- Norouzzadeh MS, Morris D, Beery S, Joshi N, Jovic N, Clune J. 2020.** A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution* **12**(1):150–161 DOI [10.1111/2041-210X.13504](https://doi.org/10.1111/2041-210X.13504).
- Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J. 2018.** Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America* **115**(25):E5716–E5725 DOI [10.1073/pnas.1719367115](https://doi.org/10.1073/pnas.1719367115).
- Ortega Adarme M, Queiroz Feitosa R, Nigri Happ P, Aparecido De Almeida C, Rodrigues Gomes A. 2020.** Evaluation of deep learning techniques for deforestation detection in the brazilian amazon and cerrado biomes from remote sensing imagery. *Remote Sensing (Basel)* **12**(6):910 DOI [10.3390/rs12060910](https://doi.org/10.3390/rs12060910).
- Padubidri C, Kamilaris A, Karatsiolis S, Kamminga J. 2021.** Counting sea lions and elephants from aerial photography using deep learning with density maps. *Animal Biotelemetry* **9**(1):6506 DOI [10.1186/s40317-021-00247-x](https://doi.org/10.1186/s40317-021-00247-x).
- Pineau J, Vincent-Lamarre P, Sinha K, Larivière V, Beygelzimer A, d'Alché Buc F, Fox E, Larochelle H. 2021.** Improving reproducibility in machine learning research (a report from the neurips 2019 reproducibility program). *The Journal of Machine Learning Research* **22**(1):7459–7478.

- Pouliot D, Alavi N, Wilson S, Duffe J, Pasher J, Davidson A, Daneshfar B, Lindsay E. 2021.** Assessment of landsat based deep-learning membership analysis for development of from-to change time series in the prairie region of Canada from 1984 to 2018. *Remote Sensing (Basel)* **13**(4):634 DOI [10.3390/rs13040634](https://doi.org/10.3390/rs13040634).
- Pu L, Xv J, Deng F. 2021.** An automatic method for tree species point cloud segmentation based on deep learning. *Journal of the Indian Society of Remote Sensing* **49**(9):2163–2172 DOI [10.1007/s12524-021-01358-x](https://doi.org/10.1007/s12524-021-01358-x).
- Radig B, Bodesheim P, Korsch D, Denzler J, Haucke T, Klasen M, Steinhage V. 2021.** Automated visual large scale monitoring of faunal biodiversity. *Pattern Recognition and Image Analysis* **31**(3):477–488 DOI [10.1134/S1054661821030214](https://doi.org/10.1134/S1054661821030214).
- Raff E. 2019.** A step toward quantifying independently reproducible machine learning research. *Advances in Neural Information Processing Systems* **32**.
- Rammer W, Seidl R. 2019.** Harnessing deep learning in ecology: an example predicting bark beetle outbreaks. *Frontiers in Plant Science* **10**:1327 DOI [10.3389/fpls.2019.01327](https://doi.org/10.3389/fpls.2019.01327).
- Rousset G, Despinoy M, Schindler K, Mangeas M. 2021.** Assessment of deep learning techniques for land use land cover classification in southern new caledonia. *Remote Sensing (Basel)* **13**(12):2257 DOI [10.3390/rs13122257](https://doi.org/10.3390/rs13122257).
- Rovero F, Zimmermann F, Berzi D, Meek P. 2013.** “Which camera trap type and how many do i need?” a review of camera features and study designs for a range of wildlife research applications. *Hystrix* **24**(2) DOI [10.4404/hystrix-24.2-8789](https://doi.org/10.4404/hystrix-24.2-8789).
- Rowe B, Eichinski P, Zhang J, Roe P. 2021.** Acoustic auto-encoders for biodiversity assessment. *Ecological Informatics* **62**(101237):101237 DOI [10.1016/j.ecoinf.2021.101237](https://doi.org/10.1016/j.ecoinf.2021.101237).
- Salamon J, Bello JP, Farnsworth A, Kelling S. 2017.** Fusing shallow and deep learning for bioacoustic bird species classification. In: *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Piscataway: IEEE.
- Samuel S, König-Ries B. 2021.** Understanding experiments and research practices for reproducibility: an exploratory study. *PeerJ* **9**(2):e11140 DOI [10.7717/peerj.11140](https://doi.org/10.7717/peerj.11140).
- Samuel S, Löffler F, König-Ries B. 2021.** Machine learning pipelines: provenance, reproducibility and FAIR data principles. In: Glavic B, Braganholo V, Koop D, eds. *Provenance and Annotation of Data and Processes—8th and 9th International Provenance and Annotation Workshop, IPAW 2020 + IPAW 2021, Virtual Event, July 19–22, 2021, Proceedings, volume 12839 of Lecture Notes in Computer Science*. Cham: Springer, 226–230.
- Schiller C, Schmidtlein S, Boonman C, Moreno-Martínez A, Kattenborn T. 2021.** Deep learning and citizen science enable automated plant trait predictions from photographs. *Scientific Reports* **11**(1):16395 DOI [10.1038/s41598-021-95616-0](https://doi.org/10.1038/s41598-021-95616-0).
- Schindler F, Steinhage V. 2021.** Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. *Ecological Informatics* **61**(101215):101215 DOI [10.1016/j.ecoinf.2021.101215](https://doi.org/10.1016/j.ecoinf.2021.101215).
- Schnitzer SA, Carson WP. 2016.** Would ecology fail the repeatability test? *BioScience* **66**(2):98–99 DOI [10.1093/biosci/biv176](https://doi.org/10.1093/biosci/biv176).
- Schuettpelz E, Frandsen P, Dikow R, Brown A, Orli S, Peters M, Metallo A, Funk V, Dorr L. 2017.** Applications of deep convolutional neural networks to digitized natural history collections. *Biodiversity Data Journal* **5**:e21139 DOI [10.3897/BDJ.5.e21139](https://doi.org/10.3897/BDJ.5.e21139).
- Schwartz ST, Alfaro ME. 2021.** Sashimi: a toolkit for facilitating high-throughput organismal image segmentation using deep learning. *Methods in Ecology and Evolution* **12**(12):2341–2354 DOI [10.1111/2041-210X.13712](https://doi.org/10.1111/2041-210X.13712).

- Spiesman BJ, Gratton C, Hatfield RG, Hsu WH, Jepsen S, McCornack B, Patel K, Wang G. 2021. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports* 11(1):7580 DOI 10.1038/s41598-021-87210-1.
- Stark PB. 2018. Before reproducibility must come preproducibility. *Nature* 557(7706):613–614 DOI 10.1038/d41586-018-05256-0.
- Sun Y, Huang J, Ao Z, Lao D, Xin Q. 2019. Deep learning approaches for the mapping of tree species diversity in a tropical wetland using airborne lidar and high-spatial-resolution remote sensing images. *Forests* 10(11):1047 DOI 10.3390/f10111047.
- Tabak MA, Norouzzadeh MS, Wolfson DW, Sweeney SJ, VerCauteren KC, Snow NP, Halseth JM, Di Salvo PA, Lewis JS, White MD, Teton B, Beasley JC, Schlichting PE, Boughton RK, Wight B, Newkirk ES, Ivan JS, Odell EA, Brook RK, Lukacs PM, Moeller AK, Mandeville EG, Clune J, Miller RS. 2019. Machine learning to classify animal species in camera trap images: applications in ecology. *Methods in Ecology and Evolution* 10(4):585–590 DOI 10.1111/2041-210X.13120.
- Tang F, Zhang D, Zhao X. 2021. Efficiently deep learning for monitoring *ipomoea cairica* (L.) sweets in the wild. *Mathematical Biosciences and Engineering* 18(2):1121–1135 DOI 10.3934/mbe.2021060.
- Tatman R, VanderPlas J, Dane S. 2018. A practical taxonomy of reproducibility for machine learning research.
- Tian J, Wang L, Yin D, Li X, Diao C, Gong H, Shi C, Menenti M, Ge Y, Nie S, Ou Y, Song X, Liu X. 2020. Development of spectral-phenological features for deep learning to understand spartina alterniflora invasion. *Remote Sensing of Environment* 242(111745):111745 DOI 10.1016/j.rse.2020.111745.
- Torney CJ, Lloyd-Jones DJ, Chevallier M, Moyer DC, Maliti HT, Mwita M, Kohi EM, Hopcraft GC. 2019. A comparison of deep learning and citizen science techniques for counting wildlife in aerial survey images. *Methods in Ecology and Evolution* 10(6):779–787 DOI 10.1111/2041-210X.13165.
- Villon S, Chaumont M, Subsol G, Villéger S, Claverie T, Mouillot D. 2016. Coral reef fish detection and recognition in underwater videos by supervised machine learning: comparison between deep learning and HOG+SVM methods. In: *Advanced Concepts for Intelligent Vision Systems, Lecture Notes in Computer Science*. Cham: Springer International Publishing, 160–171.
- Villon S, Iovan C, Mangeas M, Claverie T, Mouillot D, Villéger S, Vigliola L. 2021. Automatic underwater fish species classification with limited data using few-shot learning. *Ecological Informatics* 63(101320):101320 DOI 10.1016/j.ecoinf.2021.101320.
- Villon S, Mouillot D, Chaumont M, Darling ES, Subsol G, Claverie T, Villéger S. 2018. A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological Informatics* 48(10):238–244 DOI 10.1016/j.ecoinf.2018.09.007.
- Villon S, Mouillot D, Chaumont M, Subsol G, Claverie T, Villéger S. 2020. A new method to control error rates in automated species identification with deep learning algorithms. *Scientific Reports* 10(1):10972 DOI 10.1038/s41598-020-67573-7.
- Wagner FH, Dalagnol R, Tagle Casapia X, Streher AS, Phillips OL, Gloor E, Aragão LEOC. 2020. Regional mapping and spatial distribution analysis of canopy palms in an Amazon forest using deep learning and VHR images. *Remote Sensing (Basel)* 12(14):2225 DOI 10.3390/rs12142225.
- Waide RB, Brunt JW, Servilla MS. 2017. Demystifying the landscape of ecological data repositories in the United States. *BioScience* 67(12):1044–1051 DOI 10.1093/biosci/bix117.

- Weinstein BG. 2018.** Scene-specific convolutional neural networks for video-based biodiversity detection. *Methods in Ecology and Evolution* 9(6):1435–1441 DOI 10.1111/2041-210X.13011.
- Weinstein BG, Marconi S, Bohlman S, Zare A, White E. 2019.** Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing (Basel)* 11(11):1309 DOI 10.3390/rs11111309.
- Xi T, Wang J, Han Y, Wang T, Ji L. 2019.** The effect of background on a deep learning model in identifying images of butterfly species. *Electrical and Electronics Engineering: An International Journal* 8(1):01–08 DOI 10.14810/elelij.2019.8101.
- Xie J, Hu K, Zhu M, Yu J, Zhu Q. 2019.** Investigation of different CNN-based models for improved bird sound classification. *IEEE Access* 7:175353–175361 DOI 10.1109/ACCESS.2019.2957572.
- Yang D-Q, Ren G-P, Tan K, Huang Z-P, Li D-P, Li X-W, Wang J-M, Chen B-H, Xiao W. 2021.** An adaptive automatic approach to filtering empty images from camera traps using a deep learning model. *Wildlife Society Bulletin* 45(2):230–236 DOI 10.1002/wsb.1176.
- Younis S, Schmidt M, Weiland C, Dressler S, Seeger B, Hickler T. 2020.** Detection and annotation of plant organs from digitised herbarium scans using deep learning. *Biodiversity Data Journal* 8(10):e57090 DOI 10.3897/BDJ.8.e57090.
- Yubo Z, Zhuoran Y, Jiuchun Y, Yuanyuan Y, Dongyan W, Yucong Z, Fengqin Y, Lingxue Y, Liping C, Shuwen Z. 2020.** A novel model integrating deep learning for land use/cover change reconstruction: a case study of zhenlai county, northeast china. *Remote Sensing (Basel)* 12(20):3314 DOI 10.3390/rs12203314.
- Zhou C-L, Ge L-M, Guo Y-B, Zhou D-M, Cun Y-P. 2021.** A comprehensive comparison on current deep learning approaches for plant image classification. *Journal of Physics Conference Series* 1873(1):012002 DOI 10.1088/1742-6596/1873/1/012002.
- Zizka A, Silvestro D, Vitt P, Knight TM. 2021.** Automated conservation assessment of the orchid family with deep learning. *Conservation Biology* 35(3):897–908 DOI 10.1111/cobi.13616.
- Zuolkernan IA, Dhou S, Judas J, Sajun AR, Gomez BR, Hussain LA, Sakhnini D. 2020.** Towards an IoT-based deep learning architecture for camera trap image classification. In: *2020 IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*. Piscataway: IEEE.