Overview of LifeCLEF 2021: an evaluation of Machine-Learning based Species Identification and Species Distribution Prediction

Alexis Joly[1], Hervé Goëau[2], Stefan Kahl[6], Lukáš Picek[10], Titouan Lorieul[1], Elijah Cole[9], Benjamin Deneu[1], Maximilien Servajean[7], Andrew Durso[11], Isabelle Bolon[8], Hervé Glotin[3], Robert Planqué[4], Rafael Ruiz de Castañeda[8], Willem-Pier Vellinga[4], Holger Klinck[6], Tom Denton[12], Ivan Eggel[5], Pierre Bonnet[2], Henning Müller[5]

[1] Inria, LIRMM, Univ Montpellier, CNRS, Montpellier, France
[2] CIRAD, UMR AMAP, Montpellier, Occitanie, France
[3] Univ. Toulon, Aix Marseille Univ., CNRS, LIS, DYNI team, Marseille, France
[4] Xeno-canto Foundation, The Netherlands
[5] HES-SO, Sierre, Switzerland
[6] KLYCCB, Cornell Lab of Ornithology, Cornell University, USA
[7] LIRMM, AMI, Univ Paul Valéry Montpellier, Univ Montpellier, CNRS, France
[8] ISG, Dept of Community Health and Medicine, UNIGE, Switzerland
[9] Department of Computing and Mathematical Sciences, Caltech, USA
[10] Department of Cybernetics, FAV, University of West Bohemia, Czechia
[11] Department of Biological Sciences, Florida Gulf Coast University, USA
[12] Google LLC, San Francisco, USA

Abstract. Building accurate knowledge of the identity, the geographic distribution and the evolution of species is essential for the sustainable development of humanity, as well as for biodiversity conservation. However, the difficulty of identifying plants and animals is hindering the aggregation of new data and knowledge. Identifying and naming living plants or animals is almost impossible for the general public and is often difficult even for professionals and naturalists. Bridging this gap is a key step towards enabling effective biodiversity monitoring systems. The LifeCLEF campaign, presented in this paper, has been promoting and evaluating advances in this domain since 2011. The 2021 edition proposes four data-oriented challenges related to the identification and prediction of biodiversity: (i) PlantCLEF: cross-domain plant identification based on herbarium sheets, (ii) BirdCLEF: bird species recognition in audio soundscapes, (iii) GeoLifeCLEF: remote sensing based prediction of species, and (iv) SnakeCLEF: Automatic Snake Species Identification with Country-Level Focus.

1 LifeCLEF Lab Overview

Accurately identifying organisms observed in the wild is an essential step in ecological studies. Unfortunately, observing and identifying living organisms requires high levels of expertise. For instance, plants alone account for more than

400,000 different species and the distinctions between them can be quite subtle. Since the Rio Conference of 1992, this *taxonomic gap* has been recognized as one of the major obstacles to the global implementation of the Convention on Biological Diversity[1]. In 2004, Gaston and O'Neill [10] discussed the potential of automated approaches for species identification. They suggested that, if the scientific community were able to (i) produce large training datasets, (ii) precisely evaluate error rates, (iii) scale up automated approaches, and (iv) detect novel species, then it would be possible to develop a generic automated species identification system that would open up new vistas for research in biology and related fields.

Since the publication of [10], automated species identification has been studied in many contexts [3,12,22,35,41,50,51,59]. This area continues to expand rapidly, particularly due to advances in deep learning [2,11,36,42,52,54,55,56]. In order to measure progress in a sustainable and repeatable way, the LifeCLEF[2] research platform was created in 2014 as a continuation and extension of the plant identification task that had been run within the ImageCLEF lab[3] since 2011 [14,15,16]. Since 2014, LifeCLEF expanded the challenge by considering animals in addition to plants, and including audio and video content in addition to images [23,24,25,26,27,28,29]. Four challenges were evaluated in the context of LifeCLEF 2021 edition:

1. **PlantCLEF 2021**: Identifying plant pictures from herbarium sheets.
2. **BirdCLEF 2021**: Bird species recognition in audio soundscapes.
3. **GeoLifeCLEF 2021**: Species presence prediction at given locations based on occurrence, environmental and remote sensing data.
4. **SnakeCLEF 2021**: Automated snake species identification with Country-Level Focus.

The system used to run the challenges (registration, submission, leaderboard, etc.) was the AICrowd platform[4] for the PlantCLEF and ths SnakeCLEF challenge and the Kaggle platform[5] for GeoLifeCLEF and BirdCLEF challenges. In total, 834 teams/persons participated to LifeCLEF 2021 edition by submitting runs to at least one of the four challenges. In the following sections, we provide a synthesis of the methodology and main results of each of the four challenges. More details can be found in the overview reports of each challenge and the individual reports of the participants (references provided below).

2 PlantCLEF challenge: Identifying plant pictures from herbarium sheets

A detailed description of the task and a more complete discussion of the results can be found in the dedicated working note [13].

[1] `https://www.cbd.int/`
[2] `http://www.lifeclef.org/`
[3] `http://www.imageclef.org/`
[4] `https://www.aicrowd.com`
[5] `https://www.kaggle.com`

2.1 Objective

Automated identification of plants has recently improved considerably thanks to the progress of deep learning and the availability of training data with more and more photos in the field. In the context of LifeCLEF 2018, we measured a top-1 classification accuracy over 10K species up to 90 % and we showed that automated systems are not so far from human expertise [23]. However, this profusion of field images only concerns a few tens of thousands of species, mostly located in North America and Western Europe, with fewer images from the richest regions in terms of biodiversity such as tropical countries. On the other hand, for several centuries, botanists have collected, catalogued and systematically stored plant specimens in herbaria, particularly in tropical regions. Recent huge efforts by the biodiversity informatics community such as iDigBio[6] or e-ReColNat[7] made it possible to put millions of digitized collections online. Thus, the 2020 and 2021 editions of the PlantCLEF challenge were designed to evaluate to what extent automated plant species identification on tropical data deficient regions can be improved by the use of herbarium sheets. Herbarium collections potentially represent a large reservoir of data for training species prediction models. However, their visual appearance is very different from field photographs because the specimens are first dried and then crushed on a herbarium board before being digitized (see examples figure 1). This difference in appearance represents a very severe domain shift which makes the task of learning from one domain to the other very difficult. The main novelty of the 2021 edition over 2020 is that we provide new training data related to species *traits*, i.e attributes of the species such as their growth form, woodiness or habitat. Traits are a very valuable information that can potentially help improve the prediction of the models. Indeed, it can be assumed that species which share the same traits also share to some extent common visual appearances. This information can then potentially be used to guide the learning of a model through auxiliary loss functions for instance.

2.2 Dataset and Evaluation Protocol

The challenge is based on a dataset of 997 species mainly focused on the South America's Guiana Shield (figure 2), an area known to have one of the greatest diversity of plants in the world. It as evaluated as a cross-domain classification task where the training set consist of 321,270 herbarium sheets and 6,316 photos in the field to enable learning a mapping between the two domains. A valuable asset of this training set is that a set of 354 plant observations are provided with both herbarium sheets and field photos to potentially allow a more precise mapping between the two domains. In addition to the images, the training data includes the values of 5 traits for each 997 species. These trait data items were collected through the Encyclopedia of Life API [8] and were selected as the most

[6] `http://portal.idigbio.org/portal/search`
[7] `https://explore.recolnat.org/search/botanique/type=index`
[8] `https://eol.org/docs/what-is-eol/data-services`

Fig. 1: Field photos and herbarium sheets of the same specimen (*Tapirira guianensis* Aubl.). Despite the very different visual appearances between the two types of images, similar structures and shapes of flowers, fruits and leaves can be observed.

exhaustive ones, i.e.: "plant growth form", "habitat", "plant lifeform", "trophic guild" and "woodiness". Each of them was double-checked and completed by experts of the Guyanese flora, in order to ensure that each of the 1000 species have a validated value for each trait.

The test set relies on the data of two highly trusted experts and is composed of 3,186 photos in the fields related to 638 plant observations.

Participants were allowed to use complementary training data (e.g. for pre-training purposes) but on the condition that (i) the experiment is entirely reproducible, i.e. that the used external resource is clearly referenced and accessible to any other research group in the world, (ii) the use of external training data or not is mentioned for each run, and (iii) the additional resource does not contain any of the test observations. External training data was allowed but participants had to provide at least one submission that used only the data provided this year.

The main evaluation measure for the challenge is the Mean Reciprocal Rank (MRR), which is defined as

$$\frac{1}{Q}\sum_{q=1}^{Q}\frac{1}{\text{rank}_q}$$

where Q is the number of plant observations and rank_q is the predicted rank of the true label for the qth observation.

A second MRR score is computed on a subset of test set composed of the most difficult species, i.e. the ones that are the least frequently photographed in the field. They were selected based on the most comprehensive estimates of the available amount of field pictures from different data sources (IdigBio, GBIF, Encyclopedia of Life, Bing and Google Image search engines, previous datasets related to PlantCLEF and ExpertCLEF challenges). These difficult species are much more challenging in the sense that the discriminant features must necessary be learned from the herbarium data.

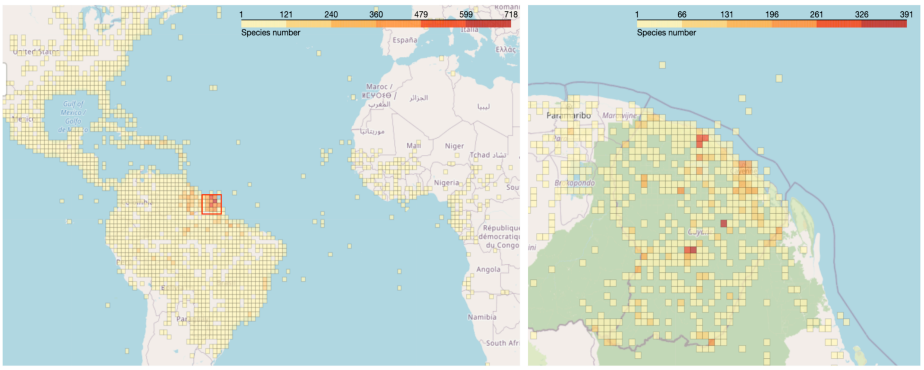


Fig. 2: Density grid maps of the number of species of geolocated plants in PlantCLEF2021. Many species have also been collected to a lesser extent in other regions outside French Guiana, such as the Americas and Africa.

2.3 Participants and Results

About 40 teams registered for the PlantCLEF challenge 2021 (PC21) and 4 of them finally submitted runs, *i.e.* files containing the predictions of the system(s) they ran. Details of the methods and systems used in the runs are synthesized in the overview working note paper of the task [13] and further developed in the individual working notes of participants (NeuonAI [5], Lehigh University [58]). Complementary runs based on the best performing approach during PlantCLEF2020 (a Few Shot Adversarial Domain Adaptation approach - FSADA - [53]) were also submitted by the organisers. In particular, we focused on assessing the impact of the trait information introduced this year. We report in Figure 3 the performance achieved by the 33 collected runs.

The main outcomes we can derive from that results are the following:

The most difficult PlantCLEF challenge ever. Traditional classification models based on CNNs perform very poorly on the task. Domain Adaptation methods (DA) based on CNNs perform much better but the task remains difficult even with these dedicated techniques. The best submitted run barely approaches a MRR of 0.2.

Genericity and stability. Regarding the difference between the two MRR metrics (whole test set vs. difficult species), the NeuonAI team demonstrated that it is possible to achieve equivalent and quite good performance for all species, even those that have few or no field photos at all in the training dataset. Rather than focusing on learning a common feature invariant domain as for the other team's submissions, the NeuonAI's approach focuses on a deep metric learning on features embeddings. Looking solely at the the second MRR score, this approach seems to be more effective in transferring knowledge to the least frequently photographed species (which is the most challenging objective). The

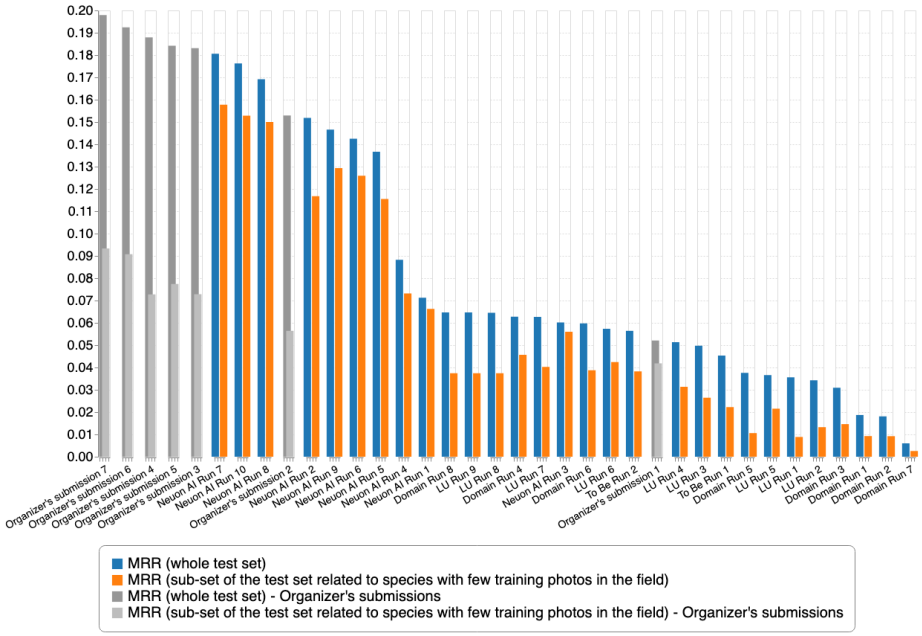


Fig. 3: PlantCLEF 2021 results

FSADA approach, on the other side, offers a better trade off considering all species together.

The most informative species trait is the "plant growth form". Organizer's submissions 4, 5 and 6 demonstrate that adding an auxiliary task related based on species traits to the FSADA approach improve performance. As hypothesised, it seems to help gathering and discriminating wide groups of plant species sharing similar visual aspects (such as tendrils for climber plants, typical large leaves for tropical trees against smaller leaves for shrubs or long thin leaves and frequent flowers for herbs).

3 BirdCLEF challenge: Bird call identification in soundscape recordings

A detailed description of the task and a more complete discussion of the results can be found in the dedicated overview paper [31].

3.1 Objective

The *LifeCLEF Bird Recognition Challenge* (BirdCLEF) launched in 2014 and has since become the largest bird sound recognition challenge in terms of dataset size and species diversity with multiple tens of thousands of recordings covering

up to 1,500 species [17,30,32]. Birds are ideal indicators to identify early warning signs of habitat changes that are likely to affect many other species. They have been shown to respond to various environmental changes over many spatial scales. Large collections of (avian) audio data are an excellent resource to conduct research that can help to deal with environmental challenges of our time. The community platform Xeno-canto[9] launched in 2005 and hosts bird sounds from all continents and daily receives new recordings from some of the remotest places on Earth. The Xeno-canto archive currently consists of more than 635,000 focal recordings covering over 10,000 species of birds, making it one of the most comprehensive collections of bird sound recordings worldwide, and certainly the most comprehensive collection shared under Creative Commons licenses. Xeno-canto data was used for BirdCLEF in all past editions to provide researchers with large and diverse datasets for training and testing.

In recent years, research in the domain of bioacoustics shifted towards deep neural networks for sound event recognition [33,49]. In past editions, we have seen many attempts to utilize convolutional neural network (CNN) classifiers to identify bird calls based on visual representations of these sounds (i.e., spectrograms) [18,34,40]. Despite their success for bird sound recognition in focal recordings, the classification performance of CNN on continuous, omnidirectional soundscapes remained low. Passive acoustic monitoring can be a valuable sampling tool for habitat assessments and the observation of environmental niches which often are endangered. However, manual processing of large collections of soundscape data is not desirable and automated attempts can help to advance this process [57]. Yet, the lack of suitable validation and test data prevented the development of reliable techniques to solve this task. Bridging the acoustic gap between high-quality training recordings and soundscapes with high ambient noise levels is one of the most challenging tasks in the domain of audio event recognition.

The main goal of the 2021 edition of BirdCLEF was to open the field of bird song identification to a broader audience by providing both a challenging research task and a low barrier to entry. The competition was hosted on Kaggle[10] to attract machine learning experts from around the world to participate and submit. While the overall task was consistent with previous editions, the organization focused on providing entry-level resources to enable participants to achieve baseline results without the need for extensive dataset analysis and workflow implementation.

3.2 Dataset and Evaluation Protocol

Deploying a bird sound recognition system to a new recording and observation site requires classifiers that generalize well across different acoustic domains. Focal recordings of bird species from around the world form an excellent base to develop such a detection system. However, the lack of annotated soundscape data for a new deployment site poses a significant challenge. As in previous editions,

[9] `https://www.xeno-canto.org/`

[10] `https://www.kaggle.com/c/birdclef-2021`

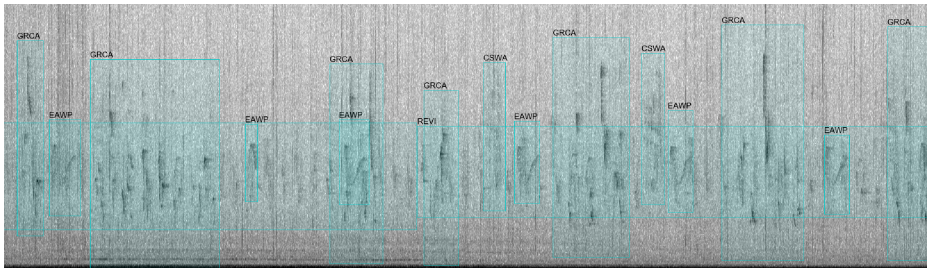


Fig. 4: Dawn chorus soundscapes often have an extremely high call density. The 2021 BirdCLEF dataset contained 100 fully annotated soundscapes recorded in South and North America.

training data was provided by the Xeno-canto community and consisted of more than 60,000 recordings covering 397 species from two continents (South and North America). Participants were allowed to use metadata to develop their systems. Most notably, we provided detailed location information on recording sites of focal and soundscape recordings, allowing participants to account for migration and spatial distribution of bird species. A validation dataset with 200 minutes of soundscape data was also provided.

The hidden test data contained 80 soundscape recordings of 10-minute duration covering four distinct recording locations. Validation data only contained soundscapes for two of the four locations. All audio data were collected with passive acoustic recorders from deployments in Colombia (COL), Costa Rica (COR), the Sierra Nevada (SNE) of California, USA and the Sapsucker Woods area (SSW) in Ithaca, New York, USA. Expert ornithologists provided annotations for a variety of quiet and extremely dense acoustic scenes (see Figure 4).

The goal of the task was to localize and identify all audible birds within the provided soundscape test set. Each soundscape was divided into segments of 5 seconds, and a list of audible species had to be returned for each segment. The used evaluation metric was the row-wise micro-averaged F1-score. In previous editions, ranking metrics were used to assess the overall classification performance. However, when applying bird call identification systems to real-world data, confidence thresholds have to be set in order to provide meaningful results. The F1-score as balanced metric between recall and precision appears to better reflect this circumstance. Precision and recall were determined based on the total number of true positives (TP), false positives (FP) and false negatives (FN) for each segment (i.e., row of the submission). More formally:

$$\textit{Micro-Precision} = \frac{TP_{sum}}{TP_{sum} + FP_{sum}}\,, \qquad \textit{Micro-Recall} = \frac{TP_{sum}}{TP_{sum} + FN_{sum}}$$

The micro F1-score as harmonic mean of the micro-precision and micro-recall for each segment is defined as:

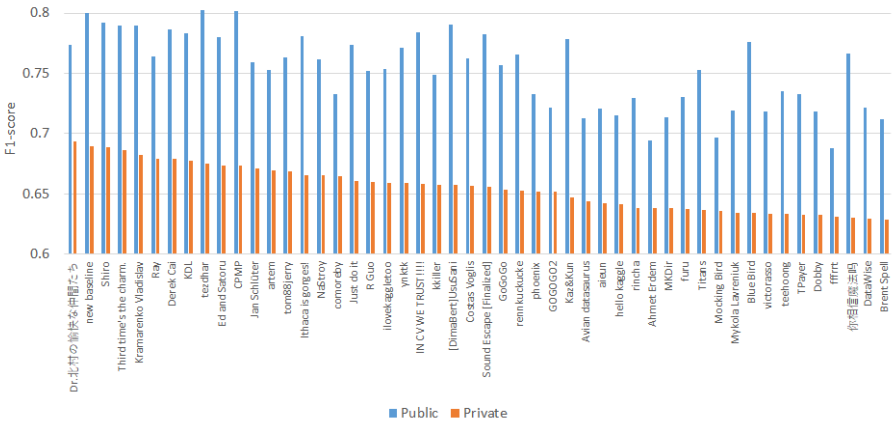


Fig. 5: Scores achieved by the best systems evaluated within the bird identification task of LifeCLEF 2021.

$$\textit{Micro-F1} = 2 \times \frac{\textit{Micro-Precision} \times \textit{Micro-Recall}}{\textit{Micro-Precision} + \textit{Micro-Recall}}$$

The average across all (segment-wise) F1-scores was used as the final metric. Segments that did not contain a bird vocalizations had to be marked with the "nocall" label, which acted as an additional class label for non-events. The micro-averaged F1-score reduces the impact of rare events, which only contribute slightly to the overall metric if misidentified. The classification performance on common classes (i.e., species with high vocal presence) is well reflected in the metric.

3.3 Participants and Results

1,004 participants from 70 countries on 816 teams entered the BirdCLEF 2021 competition and submitted a total of 9,307 runs. Details of the best methods and systems used are synthesized in the overview working notes paper of the task [31] and further developed in the individual working notes of participants. In Figure 5 we report the performance achieved by the top 50 collected runs. The private leaderboard score is the primary metric and was revealed to participants after the submission deadline to avoid probing the hidden test data. Public leaderboard scores were visible to participants over the course of the entire challenge.

The baseline F1-score in this year's edition was 0.4799 (public 0.5467) with all segments marked as non-events, and 686 teams managed to score above this threshold. The best submission achieved a F1-score of 0.6932 (public 0.7736) and the top 10 best performing systems were within only 2% difference in score. The vast majority of approaches was based on convolutional neural network

ensembles and mostly differed in pre- and post-processing and neural network backbone. Interestingly, the choice of CNN backbone does not seem to have significant impact on the overall score. Off-the-shelve architectures like MobileNet, EfficientNet, or DenseNet all seem to perform well on this task. Participants mostly used mel scale spectrograms as model inputs and the most commonly used augmentation method was mix-up (i.e., overlapping samples to emulate simultaneously vocalizing birds). Post-processing in the form of bagging and thresholding scores, location based filtering, or even decision trees as separate stage to combine scores and metadata appeared to be the most important measure to achieve high scores.

4 GeoLifeCLEF challenge: species prediction based on occurrence data, environmental data and remote sensing data

A detailed description of the task and a more complete discussion of the results can be found in the dedicated working note [37].

4.1 Objective

Automatic prediction of the list of species most likely to be present at a given location is useful for many scenarios related to biodiversity management and conservation. First, it can improve species identification tools (whether automatic, semi-automatic or based on traditional field guides) by reducing the list of candidate species observable at a given site.

Moreover, it can facilitate decision making related to land use and land management with regard to biodiversity conservation obligations (e.g. to determine new buildable areas or new natural areas to be protected).

Last but not least, it can be used in the context of educational and citizen science initiatives, e.g. to determine regions of interest with a high species richness or vulnerable habitats to be monitored carefully.

4.2 Data Set and Evaluation Protocol

Data collection. The data for this year's challenge is the same as last year reorganized in a more easy-to-use and compact format. A detailed description of the GeoLifeCLEF 2020 dataset is provided in [6]. In a nutshell, it consists of over 1.9 million observations covering $31,435$ plant and animal species distributed across US and France (as shown in Figure 7). Each species observation is paired with high-resolution covariates (RGB-IR imagery, land cover and altitude) as illustrated in Figure 6. These high-resolution covariates are resampled to a spatial resolution of 1 meter per pixel and provided as 256×256 images covering a 256m $\times$ 256m square centered on each observation. RGB-IR imagery come from the 2009-2011 cycle of the National Agriculture Imagery Program (NAIP) for the U.S.[11], and from the BD-ORTHO® 2.0 and ORTHO-HR® 1.0

[11] `https://www.fsa.usda.gov`

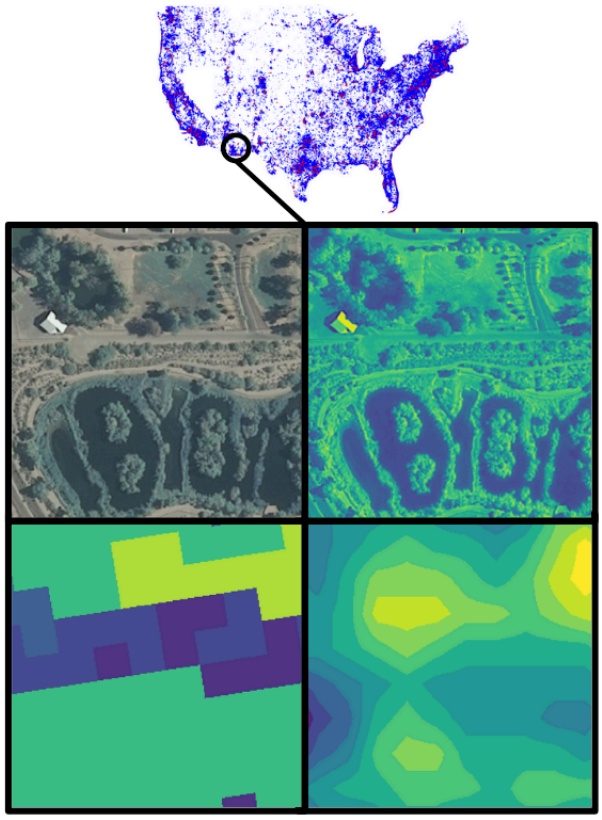
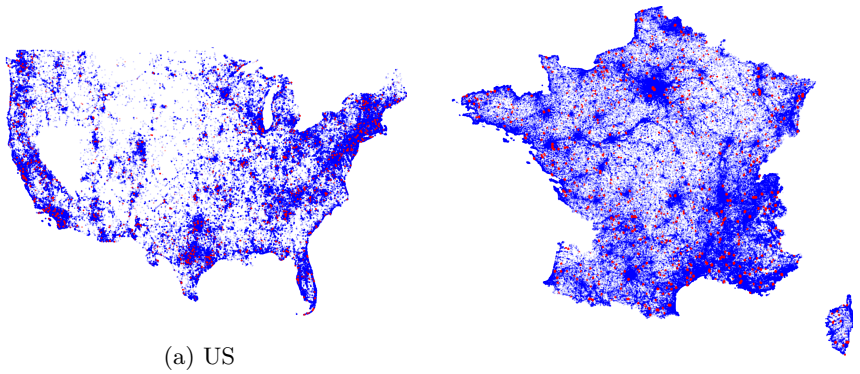


Fig. 6: In the GeoLifeCLEF dataset, each species observation is paired with high-resolution covariates (clockwise from top left: RGB imagery, IR imagery, altitude, land cover).

databases from the IGN for France[12]. Land cover data originates from the National Land Cover Database (NLCD) [21] for the U.S. and from CESBIO[13] for France. All elevation data comes from the NASA Shuttle Radar Topography Mission (SRTM)[14]. In addition, the dataset also includes traditional coarser resolution covariates: bio-climatic rasters (1km^2/pixel, from WorldClim [20]) and pedologic rasters (250m^2/pixel, from SoilGrids [19]).

Train-test split. The full set of occurrences was split in a training and testing set using a spatial block holdout procedure as illustrated in Figure 7. This limits the effect of *spatial auto-correlation* in the data [46]. Using this splitting procedure, a model cannot achieve a high performance by simply interpolating

[12] `https://geoservices.ign.fr`
[13] `http://osr-cesbio.ups-tlse.fr/~oso/posts/2017-03-30-carte-s2-2016/`
[14] `https://lpdaac.usgs.gov/products/srtmgl1v003/`



(a) US

(b) France

Fig. 7: Occurrences distribution over the US and France in GeoLifeCLEF 2021. Blue dots represent training data, red dots represent test data.

between training samples. The split was based on a global grid of 5km × 5km quadrats. 2.5% of these quadrats were randomly sampled and the observations falling in those formed the test set. 10% of those observations were used for the public leaderboard on Kaggle while the remaining 90% allowed to compute the private leaderboard providing the final results of the challenge. Similarly, another 2.5% of the quadrats were randomly sampled to provide an official validation set. The remaining quadrats and their associated observations were assigned to the training set.

Evaluation metric. For each occurrence in the test set, the goal of the task was to return a candidate set of species likely to be present at that location. To measure the precision of the predicted sets, top-30 error rate was chosen as the main evaluation criterion. Each observation i is associated with a single ground-truth label y_i corresponding to the observed species. For each observation, the submissions provided 30 candidate labels $\hat{y}_{i,1}, \hat{y}_{i,2}, \ldots, \hat{y}_{i,30}$. The top-30 error rate is then computed using

$$\text{Top-30 error rate} = \frac{1}{N}\sum_{i=1}^{N} e_i \quad \text{where} \quad e_i = \begin{cases} 1 & \text{if } \forall k \in \{1,\ldots,30\},\ \hat{y}_{i,k} \neq y_i \\ 0 & \text{otherwise} \end{cases}$$

Note that this evaluation metric does not try to correct the sampling bias inherent to present-only observation data (linked to the density of population, etc.). The absolute value of the resulting figures should thus be taken with care. Nevertheless, this metric does allow to compare the different approaches and to determine which type of input data and of models are useful for the species presence detection task.

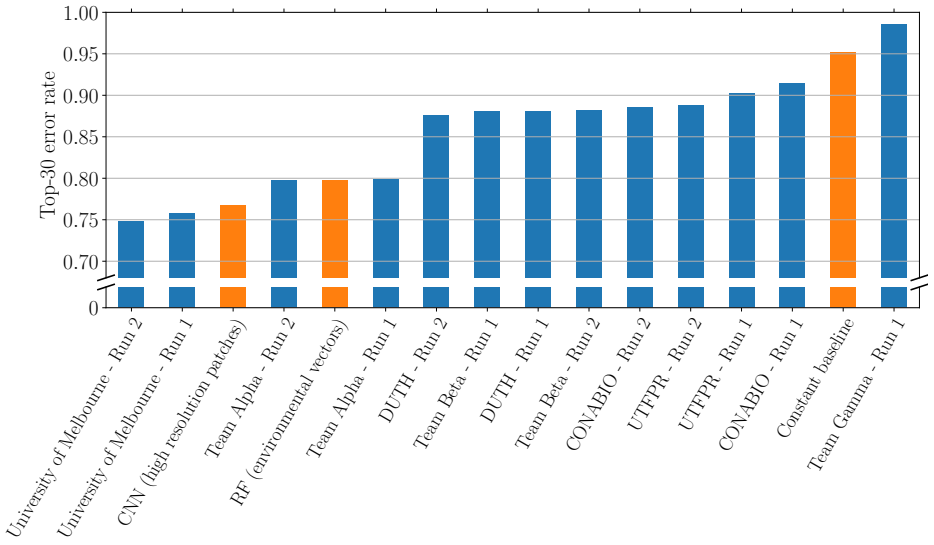


Fig. 8: Results of the GeoLifeCLEF 2021 task. The top-30 error rates of the submissions of each participant are shown in blue. The provided baselines are shown in orange.

4.3 Participants and Results

Seven teams participated to the GeoLifeCLEF 2021 challenge (hosted on Kaggle[15]) and submitted a total of 26 submissions: *University of Melbourne*, *DUTH* (Democritus University of Thrace), *CONABIO* (Comisión Nacional para el Conocimiento y Uso de la Biodiversidad), *UTFPR* (Federal University of Technology – Paraná) as well as three participants for which we could not identify the affiliation and which we denote here, respectively, as *Team Alpha*, *Team Beta* and *Team Gamma*. Details of the methods used in the submitted runs are synthesized in the overview working note paper for this task [37]. Runs of the winning team are further developed in the individual working note [48].

In Figure 8, we report the performance achieved by the collected runs. The main outcome of the challenge is that a method based on a convolutional neural network (CNN) trained solely on RGB imagery (*University of Melbourne - Run 1*) easily beats a classical model used for species distribution modelling [9] consisting of a random forest using punctual environmental variables (*RF (environmental vectors)*). This might come as a surprise as it did not make use of any bioclimatic or soil type variable which are often considered as the most informative in the ecological literature.

Generally speaking, CNN-based models trained on high resolution patches used in runs by *University of Melbourne* and *Team Alpha* as well as in the baseline *CNN (high resolution patches)* are very competitive and efficient compared to the

[15] `https://www.kaggle.com/c/geolifeclef-2021/`

traditional model (*RF (environmental vectors)*). This observation tends to show that (i) important information explaining the species composition is contained in the high-resolution patches, and, (ii) convolutional neural networks are able to capture and exploit this information.

One question raised by the challenge is how to properly aggregate the different variables provided as input. Adding altitude data to the model (*University of Melbourne - Run 2*) provides an improvement in prediction accuracy backing the intuition that this variable is informative of the species distribution. However, aggregating all the variables does not mechanically lead to higher performance: *CNN (high resolution patches)* makes use of the additional land cover data but its performance is not as good as the two runs from *University of Melbourne*. It seems that it is important not to aggregate the features representation of those variables too early in the architectures of the networks: concatenation of higher-level features (*University of Melbourne - Run 2*) is more efficient than early aggregation (*CNN (high resolution patches)*). Furthermore, it is unclear for now whether the information contained in the high-resolution patches is complementary or redundant to the one captured from the bioclimatic and soil variables and whether they should be used together or not. Finally, there remains considerable room for improvement on this challenge as the winning solution does not make use of all the different patches provided and its top-30 error rate is still high, near 75% error rate.

5 SnakeCLEF challenge: Automated snake species identification with Country-Level Focus.

A detailed description of the task and a more complete discussion of the results can be found in the dedicated overview paper [44].

5.1 Objective

To build an automatic and robust image-based system for snake species identification is an important goal for biodiversity, conservation, and global health. With over half a million victims of death and disability from venomous snakebite annually, such a system could significantly improve eco-epidemiological data and treatment outcomes (e.g. based on the specific use of antivenoms) [1,4]. This applies especially in remote geographic areas, where snake species identification assistance has a bigger potential to save lives.

Snake species identification difficulty lies in the high intra-class and low inter-class variance in appearance, which may depend on geographic location, color morph, sex, or age (Figure 10 and Figure 9). At the same time, many species are visually similar to other species (e.g. mimicry). Our knowledge of which snake species occur in which countries is incomplete, and it is common that most or all images of a given snake species might originate from a small handful of countries or even a single country. Furthermore, many snake species resemble species found on other continents, with which they are entirely allopatric. Knowing the geographic origin of an unidentified snake can narrow down the possible correct identifications considerably. In no location on Earth do more than 125 of the approximate 3,900 snake species co-occur [47]. Thus, regularization to all countries is a critical component of an automated snake identification system.

5.2 Dataset and Evaluation Protocol

Dataset Overview: For this year's challenge, we have prepared a dataset consisting of 386,006 images belonging to 772 snake species from 188 countries and all continents. The dataset has a heavy long-tailed class distribution, where the most frequent species (*Thamnophis sirtalis*) is represented by 22,163 images and the least frequent by just 10 (*Achalinus formosanus*).

Such a distribution with small inter-class variance and high intra-class variance creates a challenging task. We provide a simple train/val (90% / 10%) split to validate preliminary results while ensuring the same species distributions. The test set data consist of 23,673 images submitted to the iNaturalist platform within the first four months of 2021. Unlike in previous years, where the final testing set remained undisclosed, we provided the test data without labels to the participants.

Metadata: Besides images, we provided 3 level hierarchical taxonomic labels (family, genus, species) and location context (continent, country). The geographical information was included for approximately 85% of the development images

Fig. 9: *Naja nigricincta* from northern Namibia (left) and South Africa (right), demonstrating geographical variation within a species. *© Di Franklin - iNaturalist, and © bryanmaritz - iNaturalist*

Fig. 10: Variation in *Vipera berus* (European Adder) color and pattern. Examples from Germany, Switzerland and Poland. © *Thorsten Stegmann - iNaturalist,* © *jandetka - iNaturalist,* © *jandetka - iNaturalist, and* © *chorthippus - iNaturalist.*

and all test images. Additionally, we provide a mapping matrix (MM) describing species-country presence to allow better worldwide regularization.

$$\mathrm{MM}_{cs} = \begin{cases} 1 & \text{if species } S \in \text{country } C \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The vast majority (77%) of all images came from the United States and Canada, with 9% from Latin American and the Caribbean, 5.7% from Europe, 4.5% from Asia, 1.8% from Africa, and 1.5% from Australia/Oceania. Bias at smaller spatial scales undoubtedly exists as well [38], largely due to where participants in citizen science projects are concentrated. Nevertheless, snake species from nearly every country were represented, with 46/215 (21%) of countries having all of their snake species represented, mostly in Europe. Nearly half of all countries (106/215; 49%) had more than 50% of their snake species represented (Figure 11). Priority areas for improvement of the training dataset in future rounds are countries with high diversity and low citizen science participation, especially Indonesia, Papua New Guinea, Madagascar, and several central African and Caribbean countries (Figure 12).

Evaluation: The main goal of this challenge was to build a system that is capable of recognizing 772 snake species based on the given unseen image and

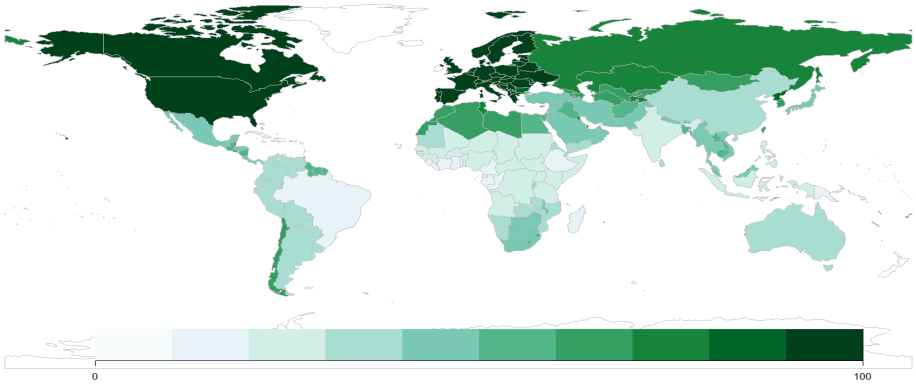


Fig. 11: Percentage of snake species per country included in SnakeCLEF2021. The countries with biggest coverage are in Europe, Oceania, and North America.

relevant geographical location, with a focus on worldwide performance. To assure that, we defined the macro F1 country performance Macro F_{1_c} as the main metric. We calculate it as the mean of country F1 scores:

$$\text{Macro } F_{1_c} = \frac{1}{N}\sum_{c=0}^{N} F_{1_c}\,, \quad F_{1_c} = \frac{1}{\sum_{s=1}^{k} MM_{cs}} \times \sum_{s=0}^{N} F_{1_s} MM_{cs} \tag{2}$$

where c is country index, s is species index, and country performance (F_{1_c}). To get the $F1_S$ we use following formula for each species:

$$F_{1_s} = 2 \times \frac{P_s \times R_s}{P_s + R_s} \tag{3}$$

5.3 Participants and Results

A total of 7 teams participated in the SnakeCLEF 2021 challenge and submitted a total of 46 runs. We have seen a vast increase in interest related to automatic snake recognition from the last year [8]. Interestingly, three participating teams are originated from India – the country with the most snakebites worldwide [39]. Details of the best methods and systems used are synthesized in the overview working notes paper of the task [44] and further developed in the individual working notes. In Figure 13, we report the performance achieved by all collected runs. The best performing model achieved an impressive Macro F_{1_c} of 0.903.

The main outcomes we can derive from that results are the following:

Object detection improves classification: Utilization of the detection network for a better region of interest selection showed a significant performance

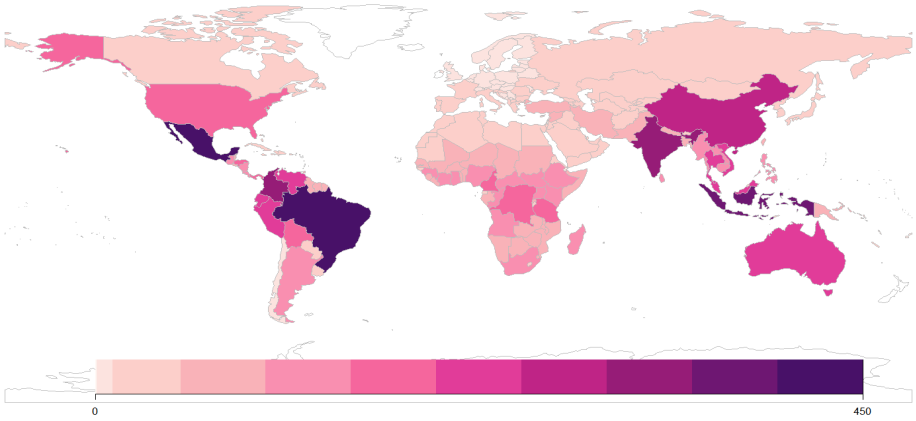


Fig. 12: Worldwide snake species distribution, showing the number of species that are found in each country. Large countries in the tropics (Brazil, Mexico, Colombia, India, Indonesia) have more than 300 species.

gain in the case of the winning team. However, such an approach requires additional labelling procedure and the build of two neural network models. Furthermore, a two-stage solution might be too heavy for deployment on edge devices; thus, its usage is probably impossible.

CNN outperforms ViT in Snake Recognition: Similarly to last year challenge [43], all participants featured deep convolutional neural networks. Besides CNNs, Vision Transformers (ViT) [7] were utilized by two teams. Interestingly, the performance of the ViT was slightly worst, which is contradictory to their performance in fungi recognition [45], thus showing that ViT might not be the best option for all fine-grained tasks.

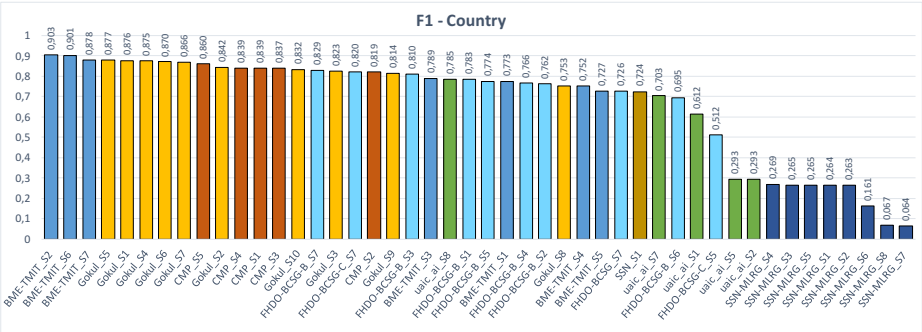


Fig. 13: Official Macro F_{1_c} scores achieved by all runs to the SnakeCLEF 2021 competition.

6 Conclusions and Perspectives

The main outcome of this collaborative evaluation is a new snapshot of the performance of state-of-the-art computer vision, bio-acoustic and machine learning techniques towards building real-world biodiversity monitoring systems. This study shows that recent deep learning techniques still allow some consistent progress for most of the evaluated tasks. One of the main new outcomes of this edition of LifeCLEF is the appearance of Visual Transformers among the best models of the SnakeCLEF task, which is the most straightforward task of LifeCLEF to experiment this new type of models. Even if their performance is still slightly inferior to that of convolutional neural networks, there is no doubt that they are now an alternative to be considered in the future. On the contrary, the 50 best methods of the BirdCLEF sound recognition task are solely based on convolutional neural networks ensembles. Interestingly, the choice of the CNN backbone does not seem to be the most determining factor of the better performance. The devil is in the detail, typically in the pre-processing and post-processing methodologies. The geolifeclef task also confirms the power of convolutional neural networks for this type of task, revealing their ability to recognise species habitats even when they are only trained on remote sensing images only (i.e. without any additional environmental data as input). Regarding the cross-domain plant identification task, the main outcome was that the performance of state-of-the-art domain adaptation methods such as FSDA can be improved by bringing additional information to the adversarial discriminator such as species traits or species taxonomy.

Acknowledgements This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No° 863463 (Cos4Cloud project), and the support of #DigitAG.

References

1. Bolon, I., Durso, A.M., Botero Mesa, S., Ray, N., Alcoba, G., Chappuis, F., Ruiz de Castañeda, R.: Identifying the snake: First scoping review on practices of communities and healthcare providers confronted with snakebite across the world. PLoS one **15**(3), e0229989 (2020)
2. Bonnet, P., Goëau, H., Hang, S.T., Lasseck, M., Šulc, M., Malécot, V., Jauzein, P., Melet, J.C., You, C., Joly, A.: Plant identification: experts vs. machines in the era of deep learning. In: Multimedia Tools and Applications for Environmental & Biodiversity Informatics, pp. 131–149. Springer (2018)
3. Cai, J., Ee, D., Pham, B., Roe, P., Zhang, J.: Sensor network for the monitoring of ecosystem: Bird species recognition. In: Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on (2007). https://doi.org/10.1109/ISSNIP.2007.4496859
4. de Castañeda, R.R., Durso, A.M., Ray, N., Fernández, J.L., Williams, D.J., Alcoba, G., Chappuis, F., Salathé, M., Bolon, I.: Snakebite and snake identification: empowering neglected communities and health-care providers with ai. The Lancet Digital Health **1**(5), e202–e203 (2019)

5. Chulif, S., Chang, Y.L.: Improved herbarium-field triplet network for cross-domain plant identification: Neuon submission to lifeclef 2021 plant. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
6. Cole, E., Deneu, B., Lorieul, T., Servajean, M., Botella, C., Morris, D., Jojic, N., Bonnet, P., Joly, A.: The GeoLifeCLEF 2020 dataset. arXiv preprint arXiv:2004.04192 (2020)
7. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
8. Durso, A.M., Moorthy, G.K., Mohanty, S.P., Bolon, I., Salathé, M., Ruiz De Castañeda, R.: Supervised learning computer vision benchmark for snake species identification from photographs: Implications for herpetology and global health. Frontiers in Artificial Intelligence **4**, 17 (2021)
9. Evans, J.S., Murphy, M.A., Holden, Z.A., Cushman, S.A.: Modeling species distribution and change using random forest. In: Predictive species and habitat modeling in landscape ecology, pp. 139–159. Springer (2011)
10. Gaston, K.J., O'Neill, M.A.: Automated species identification: why not? Philosophical Transactions of the Royal Society of London B: Biological Sciences **359**(1444), 655–667 (2004)
11. Ghazi, M.M., Yanikoglu, B., Aptoula, E.: Plant identification using deep neural networks via optimization of transfer learning parameters. Neurocomputing **235**, 228–235 (2017)
12. Glotin, H., Clark, C., LeCun, Y., Dugan, P., Halkias, X., Sueur, J.: Proc. 1st workshop on Machine Learning for Bioacoustics - ICML4B. ICML, Atlanta USA (2013), `http://sabiod.org/ICML4B2013_book.pdf`
13. Goëau, H., Bonnet, P., Joly, A.: Overview of PlantCLEF 2021: cross-domain plant identification. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
14. Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef 2013 plant identification task. In: CLEF task overview 2013, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2013, Valencia, Spain. Valencia (2013)
15. Goëau, H., Bonnet, P., Joly, A., Boujemaa, N., Barthélémy, D., Molino, J.F., Birnbaum, P., Mouysset, E., Picard, M.: The imageclef 2011 plant images classification task. In: CLEF task overview 2011, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2011, Amsterdam, Netherlands. (2011)
16. Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Barthélémy, D., Boujemaa, N., Molino, J.F.: Imageclef2012 plant images identification task. In: CLEF task overview 2012, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2012, Rome, Italy. Rome (2012)
17. Goëau, H., Glotin, H., Planqué, R., Vellinga, W.P., Stefan, K., Joly, A.: Overview of birdclef 2018: monophone vs. soundscape bird identification. In: CLEF task overview 2018, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2018, Avignon, France. (2018)
18. Grill, T., Schlüter, J.: Two convolutional neural networks for bird detection in audio signals. In: 2017 25th European Signal Processing Conference (EUSIPCO). pp. 1764–1768 (Aug 2017). https://doi.org/10.23919/EUSIPCO.2017.8081512
19. Hengl, T., de Jesus, J.M., Heuvelink, G.B., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B.,

et al.: Soilgrids250m: Global gridded soil information based on machine learning. PLoS one **12**(2) (2017)
20. Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G., Jarvis, A.: Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology: A Journal of the Royal Meteorological Society **25**(15), 1965–1978 (2005)
21. Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N., Wickham, J., Megown, K.: Completion of the 2011 national land cover database for the conterminous united states – representing a decade of land cover change information. Photogrammetric Engineering & Remote Sensing **81**(5), 345–354 (2015)
22. Joly, A., Goëau, H., Bonnet, P., Bakić, V., Barbe, J., Selmi, S., Yahiaoui, I., Carré, J., Mouysset, E., Molino, J.F., et al.: Interactive plant identification based on social image data. Ecological Informatics **23**, 22–34 (2014)
23. Joly, A., Goëau, H., Botella, C., Glotin, H., Bonnet, P., Vellinga, W.P., Müller, H.: Overview of LifeCLEF 2018: a large-scale evaluation of species identification and recommendation algorithms in the era of ai. In: Jones, G.J., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) CLEF: Cross-Language Evaluation Forum for European Languages. Experimental IR Meets Multilinguality, Multimodality, and Interaction, vol. LNCS. Springer, Avigon, France (Sep 2018)
24. Joly, A., Goëau, H., Botella, C., Kahl, S., Servajean, M., Glotin, H., Bonnet, P., Planqué, R., Stöter, F.R., Vellinga, W.P., Müller, H.: Overview of LifeCLEF 2019: Identification of Amazonian Plants, South & North American Birds, and Niche Prediction. In: Crestani, F., Brascher, M., Savoy, J., Rauber, A., Müller, H., Losada, D.E., Bürki, G.H., Bürki, G.H., Cappellato, L., Ferro, N. (eds.) CLEF 2019 - Conference and Labs of the Evaluation Forum. Experimental IR Meets Multilinguality, Multimodality, and Interaction, vol. LNCS, pp. 387–401. Lugano, Switzerland (Sep 2019). https://doi.org/10.1007/978-3-030-28577-7_29, `https://hal.umontpellier.fr/hal-02281455`
25. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Champ, J., Planqué, R., Palazzo, S., Müller, H.: LifeCLEF 2016: Multimedia Life Species Identification Challenges. In: Fuhr, N., Quaresma, P., Gonçalves, T., Larsen, B., Balog, K., Macdonald, C., Cappellato, L., Ferro, N. (eds.) CLEF: Cross-Language Evaluation Forum. Experimental IR Meets Multilinguality, Multimodality, and Interaction, vol. LNCS, pp. 286–310. Springer, Évora, Portugal (Sep 2016). https://doi.org/10.1007/978-3-319-44564-9_26, `https://hal.archives-ouvertes.fr/hal-01373781`
26. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Lombardo, J.C., Planque, R., Palazzo, S., Müller, H.: LifeCLEF 2017 Lab Overview: Multimedia Species Identification Challenges. In: Jones, G.J., Lawless, S., Gonzalo, J., Kelly, L., Goeuriot, L., Mandl, T., Cappellato, L., Ferro, N. (eds.) CLEF: Cross-Language Evaluation Forum. Experimental IR Meets Multilinguality, Multimodality, and Interaction, vol. LNCS, pp. 255–274. Springer, Dublin, Ireland (Sep 2017). https://doi.org/10.1007/978-3-319-65813-1_24, `https://hal.archives-ouvertes.fr/hal-01629191`
27. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planque, R., Rauber, A., Fisher, B., Müller, H.: LifeCLEF 2014: Multimedia Life Species Identification Challenges. In: CLEF: Cross-Language Evaluation Forum. Information Access Evaluation. Multilinguality, Multimodality, and Interac-

tion, vol. LNCS, pp. 229–249. Springer International Publishing, Sheffield, United Kingdom (Sep 2014). https://doi.org/10.1007/978-3-319-11382-1_20, https://hal.inria.fr/hal-01075770
28. Joly, A., Goëau, H., Glotin, H., Spampinato, C., Bonnet, P., Vellinga, W.P., Planqué, R., Rauber, A., Palazzo, S., Fisher, B., et al.: Lifeclef 2015: multimedia life species identification challenges. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction, pp. 462–483. Springer (2015)
29. Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., De Castaneda, R.R., Bolon, I., Durso, A., et al.: Overview of lifeclef 2020: a system-oriented evaluation of automated species identification and species distribution prediction. In: International Conference of the Cross-Language Evaluation Forum for European Languages. pp. 342–363. Springer (2020)
30. Kahl, S., Clapp, M., Hopping, A., Goëau, H., Glotin, H., Planqué, R., Vellinga, W.P., Joly, A.: Overview of birdclef 2020: Bird sound recognition in complex acoustic environments. In: CLEF task overview 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
31. Kahl, S., Denton, T., Klinck, H., Glotin, H., Goëau, H., Vellinga, W.P., Planqué, R., Joly, A.: Overview of BirdCLEF 2021: Bird call identification in soundscape recordings. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
32. Kahl, S., Stöter, F.R., Glotin, H., Planqué, R., Vellinga, W.P., Joly, A.: Overview of birdclef 2019: Large-scale bird recognition in soundscapes. In: CLEF task overview 2019, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2019, Lugano, Switzerland. (2019)
33. Kahl, S., Wood, C.M., Eibl, M., Klinck, H.: Birdnet: A deep learning solution for avian diversity monitoring. Ecological Informatics **61**, 101236 (2021)
34. Lasseck, M.: Audio-based bird species identification with deep convolutional neural networks. In: CLEF working notes 2018, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2018, Avignon, France. (2018)
35. Lee, D.J., Schoenberger, R.B., Shiozawa, D., Xu, X., Zhan, P.: Contour matching for a fish recognition and migration-monitoring system. In: Optics East. pp. 37–48. International Society for Optics and Photonics (2004)
36. Lee, S.H., Chan, C.S., Remagnino, P.: Multi-organ plant classification based on convolutional and recurrent neural networks. IEEE Transactions on Image Processing **27**(9), 4287–4301 (2018)
37. Lorieul, T., Cole, E., Deneu, B., Servajean, M., Joly, A.: Overview of GeoLifeCLEF 2021: Predicting species distribution from 2 million remote sensing images. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
38. Millar, E.E., Hazell, E.C., Melles, S.: The 'cottage effect'in citizen science? spatial bias in aquatic monitoring programs. International Journal of Geographical Information Science **33**(8), 1612–1632 (2019)
39. Mohapatra, B., Warrell, D.A., Suraweera, W., Bhatia, P., Dhingra, N., Jotkar, R.M., Rodriguez, P.S., Mishra, K., Whitaker, R., Jha, P., et al.: Snakebite mortality in india: a nationally representative mortality survey. PLoS Negl Trop Dis **5**(4), e1018 (2011)
40. Mühling, M., Franz, J., Korfhage, N., Freisleben, B.: Bird species recognition via neural architecture search. In: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
41. NIPS Int. Conf.: Proc. Neural Information Processing Scaled for Bioacoustics, from Neurons to Big Data (2013), http://sabiod.org/nips4b

42. Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J.: A deep active learning system for species identification and counting in camera trap images. Methods in Ecology and Evolution **12**(1), 150–161 (2021)
43. Picek, L., Ruiz De Castañeda, R., Durso, A.M., Sharada, P.M.: Overview of the snakeclef 2020: Automatic snake species identification challenge. In: CLEF task overview 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
44. Picek, L., Durso, A.M., Ruiz De Castañeda, R., Bolon, I.: Overview of SnakeCLEF 2021: Automatic snake species identification with country-level focus. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
45. Picek, L., Šulc, M., Matas, J., Heilmann-Clausen, J., Jeppesen, T.S., Læssøe, T., Frøslev, T.: Danish fungi 2020 – not just another image recognition dataset (2021)
46. Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., et al.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography **40**(8), 913–929 (2017)
47. Roll, U., Feldman, A., Novosolov, M., Allison, A., Bauer, A.M., Bernard, R., Böhm, M., Castro-Herrera, F., Chirio, L., Collen, B., et al.: The global distribution of tetrapods reveals a need for targeted reptile conservation. Nature Ecology & Evolution **1**(11), 1677–1682 (2017)
48. Seneviratne, S.: Contrastive representation learning for natural world imagery: Habitat prediction for 30,000 species. In: CLEF working notes 2021, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2021, Bucharest, Romania. (2021)
49. Shiu, Y., Palmer, K., Roch, M.A., Fleishman, E., Liu, X., Nosal, E.M., Helble, T., Cholewiak, D., Gillespie, D., Klinck, H.: Deep neural networks for automated detection of marine mammal species. Scientific reports **10**(1), 1–12 (2020)
50. Towsey, M., Planitz, B., Nantes, A., Wimmer, J., Roe, P.: A toolbox for animal call recognition. Bioacoustics **21**(2), 107–125 (2012)
51. Trifa, V.M., Kirschel, A.N., Taylor, C.E., Vallejo, E.E.: Automated species recognition of antbirds in a mexican rainforest using hidden markov models. The Journal of the Acoustical Society of America **123**, 2424 (2008)
52. Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., Belongie, S.: The inaturalist species classification and detection dataset. CVPR (2018)
53. Villacis, J., Goëau, H., Bonnet, P., Mata-Montero, E., Joly, A.: Domain adaptation in the context of herbarium collections: a submission to plantclef 2020. In: CLEF working notes 2020, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2020, Thessaloniki, Greece. (2020)
54. Villon, S., Mouillot, D., Chaumont, M., Subsol, G., Claverie, T., Villéger, S.: A new method to control error rates in automated species identification with deep learning algorithms. Scientific reports **10**(1), 1–13 (2020)
55. Wäldchen, J., Mäder, P.: Machine learning for image based species identification. Methods in Ecology and Evolution **9**(11), 2216–2225 (2018)
56. Wäldchen, J., Rzanny, M., Seeland, M., Mäder, P.: Automated plant species identification—trends and future directions. PLoS computational biology **14**(4), e1005993 (2018)
57. Wood, C.M., Kahl, S., Chaon, P., Peery, M.Z., Klinck, H.: Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. Methods in Ecology and Evolution **12**(5), 885–896 (2021)

58. Youshan Zhang, B.D.D.: Weighted pseudo labeling refinement for plant identification. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum (2021)
59. Yu, X., Wang, J., Kays, R., Jansen, P.A., Wang, T., Huang, T.: Automated identification of animal species in camera trap images. EURASIP Journal on Image and Video Processing (2013)