European
Commission

Evaluating Image-based Species Recognition Models suitable for Citizen Science Application to Support European Invasive Alien Species Policy

Authors

Jakuschona, N., Niers, T., Stenkamp, J., Bartoschek, T.

Editors

Schade, S., Cardoso, A.C.

2022

Joint Research Centre

Contact information
Name: Sven Schade
Email: s.schade@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC128240

PDF ISBN 978-92-76-46721-2 doi:10.2760/97305

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Jakuschona, N., Niers, T., Stenkamp, J. and Bartoschek, T., Evaluating Image-based Species Recognition Models suitable for Citizen Science Application to Support European Invasive Alien Species Policy, Schade S. and Cardose A.C. (eds.), Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-46721-2, doi:10.2760/97305, JRC128240.

Contents

Acknowledgements

This report and the associated datasets were produced with financial support from the European Commission - Joint Research Centre (JRC). We would like to thank the involved JRC Staff, in particular Sven Schade, Ana Cristina Cardoso, Irena Mitton and Fabiano Antonio Spinelli for their constant support and input to deliver this work and report. Also, we would like to thank the model providers, especially the people behind the contact points, who were very helpful and responsive during the evaluation and research, in particular Patrick Mäder und Jana Wäldchen from Flora Incognita, Ondřej Veselý from Plant.id, Vincent Kalkman from Naturalis, Dylan Verheul from Observation.org, the FishID team from Griffith University and the Global Wetlands Project, in particular Sebastian Lopez-Marcano and many more. Further, we would like to express our thanks to the Institute for Geoinformatics at the University of Münster, for providing us with necessary facilities for the research.

Authors

Nick Jakuschona, Tom Niers, Jan Stenkamp, Thomas Bartoschek
Institute for Geoinformatics, University of Münster

Editors

Abstract

Recent developments in image recognition technology and its application to automated species identification led to an increase in the research of computer vision models. These models play a growing role, especially for the detection and tracking of Invasive Alien Species (IAS) as one of the main drivers of biodiversity loss globally. Here, Citizen Science (CS) is a very promising and already successful approach of involving the public in IAS recording with the help of mobile applications (apps). However, these apps often use computer vision models specialized for distinct classes of organisms or habitats, but not for locally relevant invaders. Our work evaluates image-based species recognition models suitable for use in CS apps to meet the purposes of the European Invasive Alien Species policy. The report includes a state of the art analysis of current species recognition models by identifying criteria and requirements for their use and detecting relations between their providers. It describes a methodology for testing selected models against the IAS list of union concern, a candidate list, and local lists for European regions. The results show that no existing model could detect all species on the above mentioned lists, but several models, such as the iNaturalist API and the Microsoft AI for Earth model, show high accuracies throughout different classes of organisms. The report closes with recommendations on the future use of these models in CS apps - by either collaborating with model providers to add missing species, or by training open source models with additional image data to meet the European purpose.

1 Introduction

Invasive Alien Species (IAS) constitute one of the main drivers of biodiversity loss globally. Citizen Science (CS) can play a role in influencing the way in which society values biodiversity at the same time as contributing to collect valuable data. The involvement of the public in IAS recording through innovative CS initiatives using mobile applications (apps) already complement several existing official observation systems within EU Member States. However, data quality issues may hinder the use of CS data for early warning, management and control of IAS. Artificial Intelligence (AI)-based species recognition models have been identified as valuable tools to support the validation of IAS in Europe.

The JRC Proof of Concept project "CS data supporting IAS policy in Europe" aims at consolidating the framework for CS IAS data in Europe. Towards this objective the aim is inter alia to explore automated solutions for validation of IAS records in support of citizen science. This entails the identification, testing and provision of species recognition models to support the validation of IAS in Europe. It is possible to consider three different lists of IAS in Europe. First the official list of Union concern[1], published by the European Commission. Second, the list of candidate species for the list of IAS of Union concern[2]. And third the local lists[3] (e.g. Malta, Iberian Peninsula). The outmost goal is to improve the current data validation process of the JRC app through automation.

In this project we have identified, investigated and reviewed existing species recognition models based on literature research, search in online code repositories and the World Wide Web, and tested them in light of future use within the above mentioned work. We 1) identified existing species recognition models; 2) selected and tested existing species recognition models for their suitability to recognize invasive alien species; and 3) recommend the most suitable models to recognize invasive alien species of Union concern.

This report presents the outcomes of our work as an expert groups (under the Horizon Europe pool 043681), which was set up with the following tasks:

1. Reviewing the state of the art of image-based species recognition models, the results are presented in Section 2.
2. Testing of selected image-based species recognition models, as presented in Section 3.
3. Recommending image-based species recognition models in support of citizen science, which is provided in Section 4.

Supplementary information is provided in the Annex.

[1] List of Invasive Alien Species of Union concern, https://ec.europa.eu/environment/nature/invasivealien/list/index_en.htm, Annex B.1

[2] List of candidate IAS of Union concern, Annex B.2

[3] Local lists of IAS, Annexes B.3 - B.6

2 State of the art report on image-based species recognition models

Reviewing the state of the art of image-based species recognition models means, first of all, to identify existing species recognition models that use one or more images as input and identify the species represented on this image. After a general identification of the models, it is necessary to find out whether these models are applicable for the given use case and to measure or compare this on the basis of appropriate criteria. Accordingly, we describe how we reviewed the state of the art of image-based species recognition models and give a detailed overview in terms of the species supported, access and use conditions, and technical requirements regarding the models, considered relevant. However, we not only analyze the models themselves, but also briefly describe the model providers and how they are linked to each other.

2.1 Overview of available models

In order to be able to make statements about the state of the art of image-based species recognition models, one must first get an idea of the approaches and strategies currently pursued in the field of image-based species recognition. To fulfill this sub-task, the initial step was to collect relevant information in a variety of ways.

First, a classic literature search on scholarly publications was conducted. Therefore, we investigated journals and outcomes of conferences decisively relevant to this topic published in the last five years, including but not restricted to the journals Remote Sensing for Environment[4], Biodiversity Information Science and Standards[5], Sustainability[6], Methods in Ecology and Evolution (MEE)[7], Ecology and Evolution[8], IEEE Conference on Computer Vision and Pattern Recognition[9] and Journal of Animal Ecology[10]. The publications considered relevant are either articles describing specific models or survey papers. A list of publications reviewed in this literature search can be found in the References section. Second, in addition to searching the literature in specific journals and conferences, a general web search was conducted. For this purpose, we searched for relevant models in blogs that thematically deal with species recognition, especially with image-based approaches, as well as directly in form of a Google search. Third, we searched on hosting platforms for software development like GitHub[11] for software repositories including relevant species recognition models using appropriate keywords. Fourth, contacts to experts in the domain were used to directly obtain information on the current state of the art. In any case, all kinds of models were considered. Both in terms of delivery mode (e.g. download, API) or class of organism (e.g. multiple, birds, plants, animals, mammals).

As a result of the efforts described above, a list has been drawn up which is not elaborated further but which lists all models and developments in the field of image-based species recognition which might be of interest (see Table 1). In particular, models with a high species coverage stand out, which partly correlates with a large geographic range. One example for such a model is the one developed by iNaturalist with 38,000 species, another is provided by Pl@ntNet with 36,870 species. In both models, the covered species are spread worldwide. Geographic Range and Number of Species were thus also our initial selection criteria to decide if further research with the models would be considered useful. Accordingly, on the one hand, all models covering a number of at least 2,000 species were included in our subsequent studies. On the other hand, attention has been paid to ensure that selected models can be applied in Europe and that corresponding species are covered. A list of the models that were further investigated, selected according to the previously described criteria, is given by Table 2. Within this selection, there are three exceptions i.e. the models from FishID[12], FishVerify[13] and Wildlife Insights[14]. Despite the low species coverage, efforts were made to gain access to these models or possible APIs as they were considered as interesting due to the fact that they are trained on a certain class of organisms that is otherwise not comparable in our set of models to be studied. However, further investigation

[4] Remote Sensing for Environment Journal, https://www.journals.elsevier.com/remote-sensing-of-environment
[5] Biodiversity Information Science and Standards Journal, https://biss.pensoft.net/
[6] Sustainability Journal, https://www.mdpi.com/journal/sustainability
[7] Methods in Ecology and Evolution Journal (MEE), https://www.britishecologicalsociety.org/publications/journals/methods-in-ecology-and-evolution/
[8] Ecology and Evolution Journal, https://www.britishecologicalsociety.org/publications/journals/ecology-and-evolution/
[9] IEEE Conference on Computer Vision and Pattern Recognition, https://ieeexplore.ieee.org/xpl/conhome/1000147/all-proceedings
[10] Journal of Animal Ecology, https://www.britishecologicalsociety.org/publications/journals/journal-of-animal-ecology/
[11] GitHub, https://github.com/
[12] FishID: an online platform that automates the analysis of underwater video footage with deep learning algorithms, https://globalwetlandsproject.org/tools-2/fishid/
[13] FishVerify: a tool to identify fish species on images using artificial intelligence, https://www.fishverify.com/
[14] Wildlife Insights: a platform which enables species identification by the usage of AI models, https://www.wildlifeinsights.org/about-wildlife-insights-ai

was ruled out because no models were accessible at FishVerify and Wildlife Insights despite contact requests, and no ready-to-use models are provided generally by FishID.

Table 1: Long List. Long list of potential interesting image-based species recognition models, with indication of the name of the model, the corresponding source, and the geographic spread and number of species covered.

Model	Link	Geographic Region	Number of Species
iNaturalist API	https://www.inaturalist.org/pages/computer_vision_demo	worldwide	38,000
Pl@ntNet	https://plantnet.org/	worldwide	36,870
NIA	https://observation.org/apps/obsidentify/	central Europe (mainly Netherlands + Belgium)	22,302
Plant.id	https://plant.id, https://github.com/Plant-id/Plant-id-API	United States, Europe, India and Australia	12,183
iNaturalist Competition 2021	http://arxiv-export-lb.library.cornell.edu/pdf/2103.16483	worldwide	10,000
PantCLEF Competition 2018	https://www.imageclef.org/PlantCLEF2021	worldwide	10,000
Merlin Bird ID	https://merlin.allaboutbirds.org	worldwide	8,000
Microsoft AI for Earth	https://github.com/Microsoft/SpeciesClassification	worldwide	5,266
Flora Incognita	https://floraincognita.com/	worldwide (focus Europe)	4,803
Wildlife Insights	https://www.wildlifeinsights.org/about-wildlife-insights-ai	worldwide	1,052
FishVerify	https://www.fishverify.com/	mostly USA	500
Camera traps active learning	https://doi.org/10.1111/2041-210X.13504	Africa	100
MLWIC II R package	https://doi.org/10.1002/ece3.6692	USA	58
Different models published by Willi et al. 2019 (University of Minnesota)	https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13099	Africa and USA	51
Camera traps (Serengeti) deep learning	https://doi.org/10.1073/pnas.1719367115	Africa	48

MLWIC R package	https://doi.org/10.1111/2041-210X.13120	USA	27
DeepFish	https://www.sciencedirect.com/science/article/pii/S0925231215017312?via%3Dihub#bib10	n.a.	23
FishID	https://globalwetlandsproject.org/tools-2/fishid/	worldwide	n.a.
Candide	https://candidegardening.com/GB/identify-plants	n.a.	n.a.

Box 1. Datasets dependencies

Notably, within our investigations, we came across different datasets that can be used to train image-based species recognition models. An incomplete list of some of these datasets is added to Annex C. Most of the models contain images of animals or plants that are annotated with the species name represented in the image. Sometimes additionally to the species name, the models also provide, bounding boxes, describing where the object is placed within the image.

It is noticeable that many models, especially of those mentioned in research, are trained on similar datasets, in particular on camera trap images. A reason could be that camera traps are of special interest to be equipped with automated species detection to effectively monitor the occurrence and behavior of animals. Thus, it is also worth mentioning that mainly animals are covered by such datasets and fewer plants. Additionally, the small number of large datasets leads to a limited diversity of image recognition models. Many models are for instance trained on the Snapshot Serengeti (Swanson et al. 2015, Norouzzadeh et al. 2018, Willi et al. 2018) and the NACTI (Norouzzadeh et al., 2021) dataset. This also leads to a limited geographical coverage of models, thus it was difficult to find models trained for European, South American or Asian species.

Some providers of datasets worth mentioning are GBIF[15] which we also used to find images and LILA BC[16] which lists several useful datasets. GBIF also provides access to the images from iNaturalist and for example gives the opportunity to download the iNaturalist research-grade observations as a Darwin Core Archive[17]. Taking images from GBIF to test the models led to some issues because the images could also be used for the training of the tested models, thus they would have an advantage recognizing the depicted species. How the images were selected is described in more detail in Section 3.1.

2.2 Selection criteria for this work

With the limited selection described before, we then took a closer look at the individual models, especially in terms of species supported, access and use conditions, and technical requirements. The tables below describe the criteria that our selection of species recognition models will be investigated on. The main criteria (Table 2) are of high interest and necessary information to decide on each model's applicability. Further criteria (Table 3) add a technical dimension to the analysis, focusing on training images, updates and request specifications.

15 GBIF (Global Biodiversity Information Facility): An international network and data infrastructure aimed at providing open access to data about all types of life on Earth, https://www.gbif.org/

16 LILA BC (Labeled Information Library of Alexandria: Biology and Conservation): A repository for datasets related to biology and conservation, as a resource for machine learning, https://lila.science/

17 Darwin Core Archive: Darwin Core is a standard facilitating the sharing of information about biological diversity. The Darwin Core Archive is the structured collection of files according to the standard, https://dwc.tdwg.org/

Table 2: Main Criteria. This table lists the main criteria, i.e. criteria which are of high interest and summarise necessary information to decide on each model's applicability.

Criteria	**Description**
Species	Species covered by the model, i.e. the species the model is trained on; number of species and a list of the species (scientific names).
IAS	Invasive alien species (IAS) covered by the model, i.e. the IAS included in the model training; number of IAS and a list of the species (scientific names); split by IAS of Union concern (most important criterion), the candidate species for the list of IAS of Union concern, and the local lists (e.g. Malta, Iberian Peninsula).
Class of organism	Describing, if a model is specifically trained for a certain kind of organism (e.g. plant, mammal, bird, fish).
Habitat	Describing, if a model is specifically trained for a certain kind of habitat (terrestrial, marine, freshwater). The models are checked to see if they cover a certain habitat. If at least one species with the corresponding habitat is covered by a model, the habitat is considered as covered.
Geographic Region	Describing if a model is specifically trained for species of a geographic region (e.g. Europe, North America).
Expandability	Describing if a model can be extended by training it with additional images.
Transparency	Describing if a model is transparent? (e.g. Do we know how and for which species it is trained or is it kind of a black box?).
Accessibility	Describing if a model is available for download, as an API or in different ways.
Cost	Costs for license (e.g. per request, year, one-time download…).
License	License under which the model is published.
Accuracy (Golden Standard)	Describing the accuracy of the model when it is applied to golden standard test images (accuracy for Top-1 and Top-5 suggestion of the model), both for IAS of Union concern and IAS candidate for the Union concern list.
Accuracy (User Observations)	Describing the accuracy of the model when it is applied to noisy user observations (accuracy for Top-1 and Top-5 suggestion of the model), both for IAS and candidates.
Accuracy (Same species)	Describing the accuracy of a model, applied to images of species covered by all models (for better comparison).

Table 3: Additional Criteria. This table lists the further criteria, which add a technical dimension to the analysis focusing on training images, updates and request specifications.

Criteria	Description
Training Images	The total number of images and the corresponding dataset used for training the model.
Updates	Describing the update cycle of the model.
Requests	Describing the request specification of the model in terms of the number of input images, number of predictions, request limitations (temporal, moneywise) and if a score for the predictions is given.

Box 2. Communication with Model Providers

In order to obtain the relevant information on the criteria, we have exhausted all sources. Besides a general search on the model providers websites and documentations, we not only tested the APIs and models in order to obtain further information we also contacted some of the model providers directly. Here it was quite interesting to see differences in the style of communication. In some cases, no contact was necessary at all, for example in the case of the iNaturalist 2021 Competition, as all the information was already available. A contact was necessary in quite a few cases to clarify the details concerning a model for our criteria as it was the case for Microsoft AI for Earth, Pl@ntNet, and Plant.id. Otherwise, however, it was necessary to ask for information in order to get access to APIs, models and corresponding documentation at all, which was the case for the iNaturalist API, NIA and Flora Incognita. In yet other cases, no successful contact was made despite increased attempts. This then had the consequence that further investigation of the model was not possible as it was the case for Merlin Bird ID, for instance.

It should be noted that direct contact with the model providers was a time-consuming and labor-intensive affair. Not only that sometimes no answers were given at all, and that several contacts at different places were necessary, but also partly questions remained unanswered. Nevertheless, for the most part, the direct contacts have been worthwhile, both in terms of newly acquired information and regarding access to the models and corresponding APIs.

2.3 State of the art analysis

Table 4 visualizes the information obtained on the criteria, defined in Section 2.2, assigned to each of the models. In the following, we will briefly introduce each model in principle and highlight any special characteristics that can be derived from the criteria.

Table 4: Middle List. This table shows the most important criteria summarized in dependence on the respective model. The full table can be found using this link. At certain points, criteria were summarized or not shown here due to representation reasons.

Model	IAS covered	Candidates covered	Local Lists covered	Class of organism	Accessibility	Input images	No. of results	Score given	Updates
iNaturalist API	57	25	161	Plants + Animals	API, 200 free requests / month	1	depending on score, most time 10	no	Regular updates (~2 times / year).
iNaturalist 2021 Competition	43	16	97	Plants + Animals	Different models (ready for inference or fine-tuning) in different formats freely available for download.	1	adjustable	yes	No. But yearly competitions with different topics.
Microsoft AI for Earth	34	11	77	Plants + Animals	Free model in different formats ready to use available for download.	1	adjustable	yes	No.
NIA	41	8	132	Plants + Animals	Easy access only per smartphone app.	>=1	10	yes	Planning to extend the model to further countries, first especially in Northern Europe.
Pl@ntNet	26 (out of 36 plants)	4 (out of 6 plants)	101 (out of 112 plants)	Plants	API, 500 free requests per day.	1-5	depending on score	yes	Updated every month with new training data and/or new training architecture.
Flora Incognita	25 (out of 36 plants)	4 (out of 6 plants)	73 (out of 112 plants)	Plants	Easy access only per smartphone app.	1-3	10	yes	Regular updates; planned update in 2022 would include most missing IAS.
Plant.id	26 (out of 36 plants)	5 (out of 6 plants)	91 (out of 112 plants)	Plants	API, costs per request depending on number of requests.	1-5	depending on score	yes	Regular updates; missing invasive species are on their nice-to-have roadmap.

PantCLEFCompetition 2018	11 (out of 36 plants)	1 (out of 6 plants)	51 (out of 112 plants)	Plants	Model available for download, but we were not able to run tests on it.	n.a.	n.a.	n.a.	n.a.
Merlin Bird ID	5 (out of 5 birds)	1 (out of 1 birds)	8 (out of 8 birds)	Birds	Only smartphone app; no access to any API granted.	n.a.	n.a.	n.a.	Continuous improvements of the training database.

2.3.1 Models recognizing animals and plants

First, we reviewed with iNaturalist API, iNaturalist 2021 Competition, Microsoft AI for Earth and Nature Identification API (NIA) four image based recognition models which recognize both animals and plants.

2.3.1.1 iNaturalist API

The iNaturalist API is developed by iNaturalist[18], a joint initiative by the California Academy of Science and the National Geographic Society. Mainly, the iNaturalist team is developing a social network where anyone can share their images recorded of any seen species in nature. The contributed results are discussed as in a social network and contribution is made to the scientific data collection in the field of biodiversity monitoring. At the same time, observation images are used to train the iNaturalist Computer Vision Model. If a taxon reaches a number of 100 verifiable observations, the taxon is included in training the model[19]. The iNaturalist team aims to update the model twice a year with new species and training images[20]. The model is used in the application "Seek"[21] provided by iNaturalist, which can be used for the identification of plants and animals. Direct access to the Computer Vision Model is enabled via an API, after a corresponding request to the providers. Whereby free access is limited to a number of 200 requests a month.

2.3.1.2 iNaturalist 2021 Competition

The iNaturalist 2021 Competition is part of the FGVC8 (Fine-Grained Visual Categorization) workshop at the CVPR (Conference on Computer Vision and Pattern Recognition). It is organized by members of Cornell University, Google, Caltech, University of Edinburgh, Brigham Young University and University of Massachusetts[22]. iNaturalist is the main sponsor[23] and provides the competition with a dataset containing 2.7 million images of 10,000 species[24]. Over a period of three months, participants were challenged to develop an image classification that can recognize as many species as possible from the dataset provided. Within the context of this competition, Van Horn G et al. (2021) have compared the current dataset and others of the last competitions and tested them with different models. From this benchmark, we selected and studied the model best suited to our use case, which was the iNat2021 Supervised model. Thus, the model is not directly an outcome of the competition but was created with the corresponding dataset. It is freely available online in different formats, with the possibility for fine-tuning[25].

2.3.1.3 Microsoft AI for Earth

With the program "AI for Earth"[26] Microsoft provides funding for projects that use AI to develop alternative methods for monitoring and modeling Earth's natural systems. In the context of this 50 million USD five year program, the AI for Earth team also developed the so-called "Species Classification API"[27]. However, the Species Classification API is no longer considered as an active project by the AI for Earth team. Although the corresponding GitHub repository[28] still exists, the corresponding information on the website and an API including a demo are no longer accessible. Accordingly, no further updates can be expected, but the model used for the API is available for download in various formats on GitHub.

---

[18] iNaturalist, https://www.inaturalist.org/pages/about

[19] iNaturalist blog article on their approach which images are included in the training data of a new computer vision model, https://www.inaturalist.org/blog/31806-a-new-vision-model (Retr. 2021-12-13)

[20] iNaturalist blog article on their future approach on updating their computer vision model, https://www.inaturalist.org/blog/59122-new-vision-model-training-started (Retr. 2021-12-13)

[21] iNaturalist Seek: application which applies the iNaturalist Computer Vision Model. Users can identify plants and animals, and earn corresponding badges, https://www.inaturalist.org/pages/seek_app and https://github.com/inaturalist/SeekReactNative, both retr. 2021-12-14)

[22] Organizers of the FGVC8 workshop, https://sites.google.com/view/fgvc8/organizers (Retr. 2021-12-13)

[23] iNaturalist 2021 Competition, https://www.kaggle.com/c/inaturalist-2021/rules (Retr. 2021-12-13)

[24] Welcome Post of the iNaturalist 2021 Competition with information on the images and species contained in the competitions dataset, https://www.kaggle.com/c/inaturalist-2021/discussion/225230 (Retr. 2021-12-13)

[25] Repository which contains all the resources needed to reproduce the figures and tables that are found in the paper by Van Horn G. et al. (2021), https://github.com/visipedia/newt/tree/main/benchmark (Retr. 2021-12-13)

[26] Microsoft AI for Earth program, https://www.microsoft.com/en-us/ai/ai-for-earth

[27] Article on creating an artificial intelligence platform for the planet with background information concerning the AI for Earth program, https://www.nature.com/articles/d41586-017-08675-7 (Retr. 2021-12-13)

[28] GitHub repository of the Species Classification API, https://github.com/microsoft/SpeciesClassification (Retr. 2021-12-13)

2.3.1.4 *Nature Identification API*

Nature Identification API (NIA), is a joint effort by Observation International[29], Naturalis[30] and Intel Corp. Observation International. It is a non-profit foundation providing a worldwide platform for the storage and validation of nature information. The software offered by Observation International is, among others, Observation.org[31], offering a way to share observation data including a species registry, as well as ObsIdentify[32] the corresponding app, which also enables an image-based species recognition. The Naturalis Biodiversity Center, as well as Intel Corp. are cooperation partners. Through direct contact with the Observation International team, access to NIA was possible for our testing purposes. The access to NIA was enabled via the API of Observation.org. As the used model is currently limited to species occurring in the Netherlands and Belgium, there is the plan to extend NIA continuously by further species. Observation International is part of an arising consortium of European biodiversity portals including portals in the United Kingdom, Belgium, Netherlands, Norway, Sweden and Denmark, whereby there shall be a joint model which covers species in the mentioned countries. Broad-based tests for such a model are scheduled to take place in the middle of next year. Depending on the success of these tests, the consortium will then be established on a long-term basis.

2.3.1.5 *Comparison between models recognizing animals and plants*

Looking at the species coverage of the models that cover both plants and animals as organisms, the iNaturalist API achieves the highest coverage (compare Table 4). With 57 of 66 IAS of Union concern and 25 of 30 candidates for IAS, the iNaturalist API covers the largest part of species. Likewise for the local list, the iNaturalist API achieves the highest coverage with 161 out of 234 species. In all three species categories (IAS, candidates, local lists), the other models have a significantly lower coverage of species. While the iNaturalist API covers at least more than ⅔ of the species in all categories, this is not the case for all other models that can also recognize plants and animals in any category. Although the iNaturalist Competition (43 of 66 IAS) and NIA (41 of 66 IAS) models still come quite close to ⅔ coverage in terms of IAS of Union concern, the models by the iNaturalist 2021 Competition, Microsoft AI for Earth, and NIA only reach a little more than one-half coverage in all other categories at most (excluding the local list coverage of NIA with 132 of 234 species).

Concerning the requests, all investigated animal and plant models accept exactly one image as input for which predictions will be made, excluding NIA, which also accepts multiple images as single input. As a result, ten predictions are usually returned, whereby the number is adjustable for the models of the iNaturalist Competition and Microsoft AI for Earth. All of the previously mentioned models, apart from the iNaturalist API, provide a score in addition to the prediction, which usually describes the probability for the prediction being correct. A list which shows exactly which model covers which species can be found in Annex B.

2.3.2 Models recognizing only plants

Second, we reviewed with Pl@ntNet, Flora Incognita, Plant.id and PantCLEF Competition 2018 four image based recognition models which recognize plants but no animals.

2.3.2.1 *Pl@ntNet*

The Pl@ntNet[33] project, implemented by a consortium including CIRAD, INRA, INRIA, IRD and the Agropolis Foundation, is a tool which supports the image-based identification of plants for both, amateurs and professionals. The main goal of the project is to facilitate the sharing of information on plants and enable drawing appropriate conclusions with regard to research, business and private activities. Like in a social network, citizen scientists can share their plant observations, which are then identified by the community, assuming they are of appropriate quality. The images of these observations are then used to train a model, which is available via the so-called Pl@ntNet API. Thereby, 500 identification requests per day are free of charge[34], whereby higher

29 Observation International, https://observation-international.org/en/
30 Naturalis, https://www.naturalis.nl/en
31 Observation.org, https://observation.org/
32 ObsIdentify, https://observation.org/apps/obsidentify/
33 Pl@ntNet, https://identify.plantnet.org/
34 Pl@ntNet pricing information, https://my.plantnet.org/pricing (Retr. 2021-12-13)

request numbers are enabled for specific non-profit scientific purposes on a cost-free basis[35]. The model behind the API is updated monthly both, in terms of training data and new training architecture.

2.3.2.2 Flora Incognita

The Flora Incognita[36] research group at the Max Planck Institute for Biogeochemistry in Jena and the Technical University Ilmenau develops methods and technologies that enable automated monitoring of biodiversity. This project basically delivers the two applications “Flora Capture” and “Flora Incognita”[37] Whereby Flora Capture is an application to collect plant images without extracting them from nature, which then serve as data basis for the Flora Incognita application. Collected images are evaluated and identified by a team of experts and then included in the dataset to train an image-based species recognition model, deployed within the Flora Incognita application. The service which enables the species recognition in the Flora Incognita application is called Flora Incognita identification service[38]. A service which allows the identification of plant images with predictions for the most likely species and a corresponding confidence value for each of them. Besides the Pl@ntNet API, Flora Incognita has the peculiarity that the database is filled not only with images of the entire plant, but also with images depicting only the leaf or flower. Thus, even if plant or flower images of different species are very similar, you can identify the species with the leaf. Besides the integration in the Flora Incognita application, the Flora Incognita identification service can be integrated in other applications e.g. the “Flora Helvetica”[39] application. Such access to the Flora Incognita identification service can be enabled through direct contact to the project management of the Flora Incognita research group, however access is associated with a certain authentication effort. The model behind the service is updated regularly, whereby the upcoming update is expected to include extensive improvements in terms of IAS coverage.

2.3.2.3 Plant.id

Plant.id is a project developed by the team of the company FlowerChecker[40], whereby the main goal is to facilitate the monitoring of invasive and endangered species for a wide range of usage scenarios from business to private use. One of the main developments in the Plant.id project is the Plant.id API[41], a machine learning plant identification system, based on the technologies TensorFlow[42], Python and AWS. For corresponding images, the API returns results with predictions for the species depicted in the image as well as additional information on the species like potential plant diseases. The Plant.id API is integrated in the both applications Planta[43] and FlowerChecker[44]. Using the API has a base price of 0.05 € per request, with discounts for a higher number of requests. However, there is the option for a beneficial relationship on a monthly basis for NGOs which have a clear use case and a potential for cooperation on marketing level. The Plant.id API is continuously improved in terms of species coverage, machine learning technologies and additional information about the species. The IAS which are currently not recognized in the model of the Plant.id API, have been added to the nice-to-have roadmap. In addition to the Plant.id API, the FlowerChecker team develops further systems like Plant.id Sky (a system to identify weed on fields via UAV) or Plant.id Sensor (a system to identify weed via a sensor on agricultural machines)[45].

2.3.2.4 PlantCLEF Competition 2018

LifeCLEF[46] is a campaign with the aim to boost advances in the domain of species identification. Thereby the overriding goal is to improve the knowledge about species in terms of identity, geographic distribution and its evaluation to enable sustainable development of humanity and the conservation of biodiversity. LifeCLEF is organized by the Conference and Labs of the Evaluation Forum (CLEF), a series of independent peer-reviewed

---

35 Pl@ntNet terms of use, https://my.plantnet.org/terms_of_use (Retr. 2021-12-13)
36 Flora Incognita, https://floraincognita.com/
37 Blog article, describing the both applications developed by the Flora Incognita team, https://floraincognita.com/blog/2019/06/ (Retr. 2021-12-13)
38 Description of the Flora Incognita service, https://floraincognita.com/flora-projects/
39 Flora Helvetica, https://www.flora-helvetica.ch/
40 FlowerChecker, http://flowerchecker.com/
41 Plant.id API, https://web.plant.id/plant-identification-api/
42 TensorFlow: Open source platform for machine learning, https://www.tensorflow.org/
43 Planta, https://getplanta.com/en
44 FlowerChecker application, http://flowerchecker.com/#flowerchecker-app
45 Plant.id with different products, https://web.plant.id/
46 LifeCLEF, https://www.imageclef.org/LifeCLEF2022

workshops on a broad range of challenges and a set of benchmarking activities. PlantCLEF is again one of the challenges carried out within the LifeCLEF campaign. The challenge of PlantCLEF is to assign, via an automated image-based species recognition, user observations taken from Pl@ntNet to herbarium entries taken from the French Guiana IRD Herbarium[47] and from iDigBio[48]. As can already be seen from the origin of the herbarium sheets, PlantCLEF's focus in recent years has been on South America. However, since this region is not of high interest for this work's use case, we concentrated our analysis on the PlantCLEF challenge 2018 which was conducted under the name ExpertLifeCLEF 2018[49]. This challenge was more of a competition between experts and machines. A set of observations were identified by experts, and considered to be correctly assigned to a species, i.e., the golden standard. Now the challenge was to classify correctly as many of the golden standard images identified by the experts as possible with an automated image-based species recognition. The recognition was based on a training dataset consisting of the images of the previous LifeCLEF campaigns (Goëau et al., 2021). Milan Šulc, the winner of this challenge shared the model[50] he developed during this challenge (Šulc et al., 2018). However, since PlantCLEF is a competition, it is not expected that the models will be updated on a regular basis, but there are further models available for the different competition years[51].

2.3.2.5 Comparison between models recognizing only plants

Looking at the species coverage of the models that cover only plants as organisms, one must first consider that the total number of species covered will obviously be significantly lower than for models that can detect both animals and plants. As can be seen in Table 4 in all three species categories (IAS, candidates, local lists) the models provided by Pl@ntNet, Flora Incognita, and Plant.id all cover at least ⅔ of the species except Flora Incognita for the local list, whereas it is with 73 species still close to ⅔ coverage. The coverage of the 2018 PantCLEF Competition model is significantly lower. Regarding IAS of Union concern this model does not even cover ⅓ of the species, for the candidate species just one out of six species and for the local lists not at all one half. In terms of species coverage of IAS and candidates, the three models Pl@ntNet, Flora Incognita and Plant.id are quite similar. Pl@ntNet and Plant.id cover with 26 of 36 possible species one IAS more than Flora Incognita and Plant.id covers with five of six candidate species one species more than Pl@ntNet and Flora Incognita. There are some major differences between the three models in terms of the local lists. Pl@ntNet covers slightly more than 90 %, Plant.id 91 of 112 species and Plant.id just under ⅔ of all species on the local lists. Alternatively, looking only at the total count of the three species categories, Pl@ntNet would have the highest coverage with 131 out of 154 species. A list which shows exactly which model covers which species can be found in Annex B.

Concerning the requests, the three models Pl@ntNet, Flora Incognita and Plant.id provide, in addition to the prediction for an image, a certain score, which usually describes the probability for the prediction being correct. For the models provided by Pl@ntNet and Plant.id at least one but a maximum of five images are possible as input, whereby the number of returned predictions depends on the respective score. Depending on the quality of the images, Flora Incognita needs one to three images as input, but the result always contains ten predictions. Due to the previously described low species coverage of the PlantCLEF Competition model and difficulties in getting the model to run (see Section 3), we were unable to collect information on the requests.

2.3.3 Models recognizing only birds

Third, we reviewed with Merlin Bird ID one image-based recognition model which recognizes birds but no further animals or plants.

2.3.3.1 Merlin Bird ID

Merlin Bird ID[52] is a freely available application developed by the Cornell Lab of Ornithology, with the goal to provide a quick identification help for all levels of bird watchers. The most interesting function of the app for this work's use case is called Photo ID[53]. The team of the Cornell Lab of Ornithology uses computer vision and

47 French Guiana IRD Herbarium, http://publish.plantnet-project.org/project/caypub
48 iDigBio (US National Resource for Advancing Digitization of Biodiversity Collections), https://www.idigbio.org/
49 ExpertLifeCLEF 2018, https://www.imageclef.org/node/231
50 Model by Milan Šulc, http://ptak.felk.cvut.cz/personal/sulcmila/models/LifeCLEF2018/
51 Different PlantCLEF models, 2021: https://github.com/NeuonAI/plantclef2021_challenge, 2019: https://github.com/datvo06/PlantCLEF2019MRIM, and 2018: http://ptak.felk.cvut.cz/personal/sulcmila/models/LifeCLEF2018/
52 Merlin Bird ID, https://merlin.allaboutbirds.org/
53 Photo ID, https://support.ebird.org/en/support/solutions/articles/48000966224-merlin-photo-id (Retr. 2021-12-13)

machine learning approaches to facilitate a photo identification model that runs on mobile devices. Thereby the Macaulay Library[54] serves as the basis for the training images. In addition to the Merlin Bird ID app, the Cornell Lab of Ornithology also operates the eBird platform (also available as smartphone application), a birding community to share sightings and find birds. The observations shared by users in eBird are thus collected in the Macaulay Library and serve as training data for the Photo ID function of the Merlin Bird ID app. In this way, the training data and thus the model behind Photo ID are continuously updated. To ensure that this is also applied in the Merlin Bird ID app, Photo ID is offered as an additional free in-app download to keep the model always up to date, independent of usual app updates. Basically, Photo ID can only be used via the app, not via a separate model download or API usage. Even after numerous attempts, no contact could be established with the Merlin Bird ID team and thus no direct access to the corresponding model could be achieved. Besides Bird ID, Merlin Bird ID provides additional functions. The app is basically kept small in terms of download size and users can download the relevant species for their region of interest via so-called "bird packages". However, these are not required to use Bird ID. But with these packages Merlin Bird ID provides, among other things, the function of a bird diary or the possibility to identify a user observation by asking the user certain questions regarding the observation (e.g. where, when, seize, appearance, activity). In addition, Sound ID is integrated in the Merlin Bird ID application, a feature to identify recorded bird calls.

2.3.3.2 Comparison between models recognizing only birds

Looking at the species coverage of the model that covers only birds, we must first note that with Merlin Bird ID only one suitable model was found during our research described in Section 2.1. It covers all birds occurring in our species lists (five birds in IAS, one bird in candidates, eight birds in local lists). However, as described at Merlin Bird ID and Communication with Model Providers, since no direct access to the model is possible, not even through a contact to the Merlin Bird ID team, we could not collect any information on possible requests.

2.3.4 Relationships within and between model providers

During the analysis of the state of the art models, it became apparent that not only the providers themselves use different interwoven apps and systems to enable image-based species recognition, but also that different providers cooperate with each other. In the following, we will discuss these relationships.

2.3.4.1 Relationships within model providers

A closer analysis of the various image-based species recognition models reveals that they often follow a specific procedure. One basic requirement for the application of a model is previous training. So in a sense, a model has to learn how to recognize species in images. Accordingly, one trains the model with a dataset of images representing species whose identity is known. Thus, the model learns what a certain species looks like. Once the model has learned through training what a species looks like, new images can be used for which the model is then able to predict the depicted species, based on the knowledge gained in training. Thus, for prediction, a training basis, in the form of a dataset with verified species images, where the represented species is confirmed to be correctly assigned, is required.

A large number of the providers of image-based species recognition models collect this training data themselves. The reasons for this can be manifold e.g. there can be difficulties in terms of usage rights or there is a need for continuous updates as species populations also change. In addition, the dataset and thus the images must be adapted to the corresponding use case, for example, if species should be recognized later on user observation images that were taken in the wild, or if herbarium plants need to be recognized. There are a lot of factors that influence the design of the training dataset. Accordingly, many of the model providers have decided to create the datasets themselves and to organize the necessary image search adapted to their use case.

One solution for this collection of images can be a crowdsourcing platform following the citizen science approach. To run through an example, a model provider operates a platform on which citizens can share their species observations in the form of images. Besides sharing, the quality of the images can be assessed and the shown species can be recognized and confirmed by the community. If sufficient images are available in the appropriate quality (i.e. resolution, verified species assignment) and number, they can be used to train an image-

[54] Macaulay Library (wildlife media archive), https://www.macaulaylibrary.org/

based species recognition model. This model, trained with the images from one platform, can then be used on another platform for automated image-based species recognition.

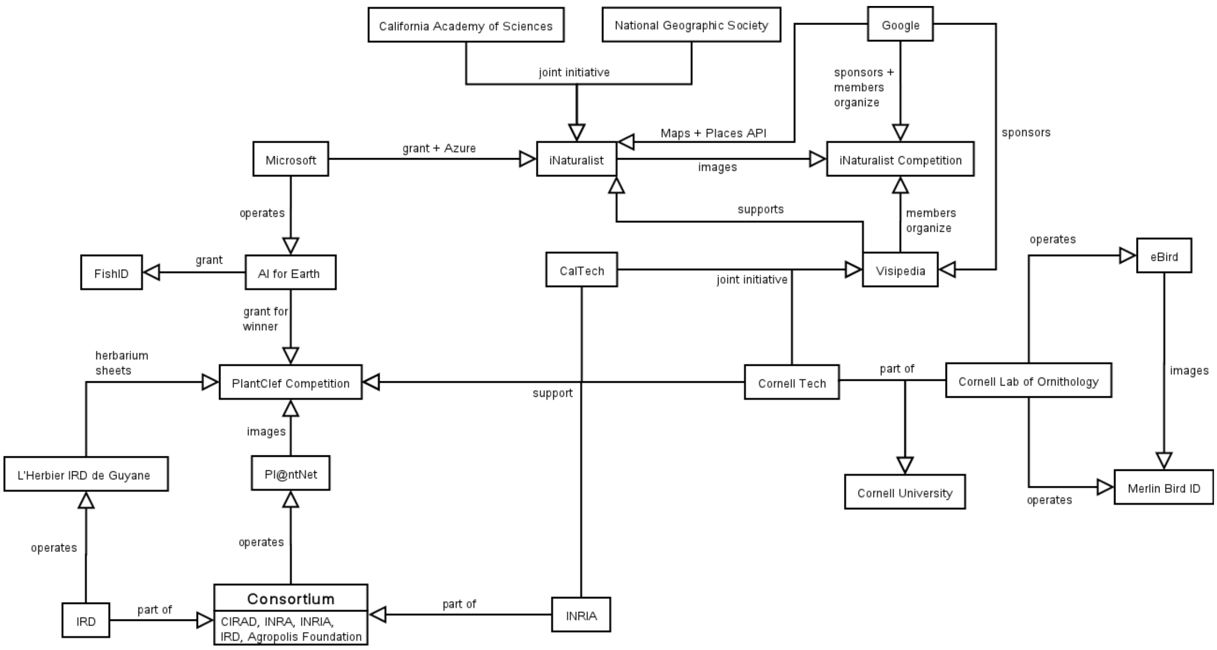
This approach has been used by several providers of the models we studied. Thus, for example, the images collected on the iNaturalist platform are used to train the model of the API once a species reaches 100 verifiable observations. Similar for Flora Incognita and Merlin Bird ID, the images are collected with an extra application (in the case of Flora Incognita: Flora Capture and in the case of Merlin Bird ID: eBird). This procedure is likewise for NIA with Observation.org, Plant.id with FlowerChecker and Pl@ntNet.

It can therefore be considered a common practice of model providers to collect the training data that is necessary for the operation of an image-based species recognition model themselves. This can be done by applications or platforms that are not obviously related to the actual image-based recognition feature.

2.3.4.2 Relationships between model providers

Besides connections within model providers, there are also some between the providers themselves. Figure 1 shows these relationships in the form of an Entity-Relationship (ER) model. In the following, we describe the different relationship types and describe some of the relationships in detail.

Figure 1: Provider Relationships. The figure shows a simplified ER model illustrating the relationships between the various model providers, organizations and applications. The model providers and corresponding organizations are represented by rectangular boxes, the relationships between them by arrows. The arrows always have a direction (triangle at one end of the relationship), which indicates in which direction the relationship works. Relationships and organizations are defined by keywords. To give an example, the organization Cornell Lab of Ornithology operates the two applications eBird and Merlin Bird ID.



One of the most obvious types of relationships is financial support. An example here is Google, who not only sponsored the iNaturalist Competition[55], but also Visipedia[56] by a research award. Besides Microsoft, with its AI for Earth program, provides financial support for projects that use AI and monitor earth's natural systems. This program supports the organizations FishID and iNaturalist in the form of grants but also the winners of the PantCLEF Competition with a 5,000 USD cloud credit[57].

55 Information on the iNaturalist Computer Vision Model, https://www.inaturalist.org/pages/computer_vision_demo (Retr. 2021-12-13)

56 Visipedia, https://vision.cornell.edu/se3/projects/visipedia/

57 Microsoft AI for Earth supported a PhD student with a 10,000 USD grant to accelerate the work on FishID (https://news.griffith.edu.au/2020/01/21/microsoft-ai-grant-sharpens-phds-eye-on-fish-monitoring/, Retr. 2021-12-13). iNaturalist is

Another typical form of relationships is given in the form of business structures. So that one organization is part of another, for example. This is the case for the Cornell Lab of Ornithology and Cornell Tech, which are both part of Cornell University. Similar for the organizations IRD and INRIA which are both part of a Consortium, which operates the platform Pl@ntNet[58]. This dependency indicates the next type of relationship, namely that a consortium or organization operates a platform or application. This is also the case for IRD which operates the L'Herbier IRD de Guyane to collect herbarium sheets, or eBird and Merlin Bird ID, which are both operated by the Cornell Lab of Ornithology.

What can also be partly contained in the relationships explained in the previous paragraph are concrete collaborations between organizations. iNaturalist, for example, is a joint initiative by the California Academy of Sciences and the National Geographic Society[59]. The same is the case for Visipedia which is a joint initiative on visual information by two vision groups at Caltech and Cornell Tech[56]. In addition, there are also collaborations that are not directly a joint initiative, but nevertheless a mutual support. For example, iNaturalist received support from Vispedia at the beginning of the development of their computer vision model in 2016[55]. Or else the PlantCLEF Competition, which is supported by INRIA, CalTech and Cornell Tech. In general, the competitions are supported by employees of affiliated organizations[60]. Thus, the iNaturalist Competition is supported by members of Google and Visipedia[61].

Another form of relationship is given in the form of technical support. One example here is Azure, the cloud computing platform of Microsoft, which enabled iNaturalist to operate a computer vision model[62]. Another example is Google, which grants iNaturalist free access to the Places APIs and their Maps[59]. Besides providing appropriate images is a common relationship. Not only within organizations like between eBird and Merlin Bird ID, but also as provision for competitions. Thus, iNaturalist provides images for the iNaturalist Competition, and Pl@ntNet as well as L'Herbier IRD de Guyane provide images for the PlantCLEF Competition.

With these examples, it becomes clear that a variety of relationships exist between the different model providers or organizations and applications. These can take the form of financial support, the sharing of datasets or technical products, cooperation or general business structures. Certainly, our description and also the ER model shown in Figure 1 do not claim to be complete, since they describe only relationships that we noticed during our analysis. Nevertheless, they provide a general impression of the linkage of the various model providers. The various collaborations show the efforts of the different organizations to move forward in the domain of image-based species recognition. This is also confirmed by the competitions organized to provide new approaches in this domain.

a partner of the AI for Earth program and received a grant (https://www.microsoft.com/en-us/ai/ai-for-earth-partners?activetab=pivot_1:primaryr3, https://ai4edatasetspublicassets.blob.core.windows.net/grantee-profiles/iNaturalist_US_Bio___AI4E%20Grantee%20Profile.pdf, both retr. 2021-12-13). The winner of each LifeCLEF challenge will be offered a cloud credit grant of 5,000 USD (https://www.imageclef.org/PlantCLEF2021).

58 Pl@ntNet credits, https://identify.plantnet.org/credits

59 iNaturalist informs about their joint initiative (https://www.inaturalist.org/pages/about) and their further partners (https://www.inaturalist.org/pages/partners).

60 The PlantCLEF competition is supported by employees of affiliated organizations, in this case members of CIRAD (https://www.imageclef.org/LifeCLEF2022).

61 The iNaturalist Competition is part of the FGVC workshop which is organized by members of Google and Visipedia (https://sites.google.com/view/fgvc8/organizers).

62 Microsoft enabled iNaturalist to use Microsoft Azure (https://ai4edatasetspublicassets.blob.core.windows.net/grantee-profiles/iNaturalist_US_Bio_AI4E%20Grantee%20Profile.pdf, Retr. 2021-12-13).

3 Models suitable to improve the validation of Invasive Alien Species records

As mentioned in Section 2, accuracy is an important criterion to evaluate the models. To analyze the accuracy of a model, we tested the models with images of IAS and checked whether the different image-based species recognition models can detect the correct species. In this section, we describe how the models can be accessed, our test methodology and the results of our tests.

In the end, we tested seven of the before in detail inspected nine models:

- iNaturalist API
- iNaturalist 2021 Competition
- Microsoft AI for Earth
- NIA
- Pl@nt Net
- Flora Incognita
- Plant.id

We did not evaluate the PlantCLEF 2018 model and the model by Merlin Bird ID. For the PlantCLEF model, we were not able to run it on the available infrastructure in the given time. As fundamental problem the model was provided as TensorFlow checkpoints and not as a frozen model (as, for example, the Microsoft AI for Earth model). We tried to convert it to a frozen model by restoring the checkpoints[63], but TensorFlow could not restore the checkpoints correctly. We contacted the model creator, who suggested using the TF-Slim package[64] to resolve the technical issues. Since the model was created in 2018 with an old version of TF-Slim and TensorFlow, most functions were deprecated now (in November 2021). Thus, we installed the TensorFlow and TF-Slim versions that were used to train the model. To make the images accessible for the checkpoint model, we had to convert them to tensors. Due to several errors, especially while converting the images to tensors with the old packages, we were not able to get the checkpoints running for the classification of the images. Given our time constraints and the limited coverage of IAS by the PlantCLEF model, we concentrated on the other models instead. For the Merlin Bird ID model, as mentioned in Box 2, we did not receive any answers to our questions from the Merlin Bird ID team, and therefore could not access the model for testing.

3.1 Access to the selected models

The seven remaining models were accessed in different ways. The models from the iNaturalist 2021 Competition and Microsoft AI for Earth were accessible as “raw” models. Thus, we deployed them ourselves on local machines. The Flora Incognita model was tested directly by the model providers, with a set of images provided by us. We had to perform the test in this way because the model owners expressed major difficulties in granting access to the API. As reasons they mentioned different security policies and issues related to user management. Since the model providers preferred the option of performing the tests themselves, we have to assume that the results we received (as raw data in JSON-format via mail) are correct. All other models were accessed through respective APIs. Furthermore, again for Flora Incognita model, it was necessary to convert the test images to jpg format. For all other models .jpg or .png formats could be used.

3.1.1 Self-deployed models

The iNaturalist 2021 Competition and the Microsoft AI for Earth models were the only self-deployed models, i.e. both were available for download and could be deployed locally by ourselves. We downloaded them, studied their documentation and examined the available code, which both were provided via GitHub. To deploy both models on a local infrastructure we adapted existing code for our needs, where required.

[63] TensorFlow function to load model checkpoints, https://www.tensorflow.org/api_docs/python/tf/compat/v1/train/Saver#restore

[64] TF-Slim: Lightweight library for defining, training and evaluating complex models in TensorFlow, https://github.com/google-research/tf-slim

In the case of the iNaturalist 2021 Competition model, mainly the *run_pytorch_server.py*[65] script, which was originally created for the iNaturalist 2017 Competition, had to be adapted. To execute it, the respective models need to be downloaded from another GitHub repository, and the path to the model needs to be adapted within the script. The adapted scripts are also made available via GitHub[66].

The Microsoft AI for Earth model behaves similarly. Here, the scripts from the official Microsoft species classification repository[67] are adapted and narrowed down to the scripts needed for our purposes. These scripts also offer the opportunity to avoid downloading the models beforehand, but to access the required model only while running the script. Again, the adapted scripts are uploaded to GitHub[68].

Both models are accessed with Python scripts and it was necessary to install different packages. To avoid any technical issues with different package versions, we recommend creating individual virtual environments, for example by using Anaconda[69]. The READMEs of the repositories include more details on how to exactly run the code and where to download the models.

Once set-up correctly, both types of scripts allow to hand over a folder of images to be classified (tested), and both write the results into a .txt or .csv file. Comparing both models, it was noticeable that the inference by the Microsoft model took longer than the one by the iNaturalist 2021 Competition model.

3.1.2 APIs

The models of iNaturalist, NIA, Pl@ntNet and Plant.id were only accessible via respective APIs. The APIs have different functions and the requests have to be designed differently for each API. To access the APIs, the model providers published different documentations and example scripts, which explain how to use the API. We provide an overview of the functionality of the APIs in Table 5.

Pl@ntNet and Plant.id provide very detailed documentation, explaining how the requests must be designed, which parameters can be added, and how to interpret the responses. For these both APIs there were even ready-to-use code examples online available.

The iNaturalist API provides substantial documentation. Here it was possible to adjust the parameters that can be included in a request via an online tool. There was also the possibility to send an example request and inspect the results. Code examples were provided, too, but they unfortunately did not work for us, i.e we had to build our own request from scratch. For example, we had to try different image encodings to find out the right encoding for the API. Furthermore, the documentation and coding examples are not publicly available, but only through an approved account, which we had to set up for our testing purposes.

For NIA, only documentation about the authentication and requests to Observation.org is available, but not about the endpoint which gives access to the Computer Vision (CV) model. Upon request, we got an example curl[70] request for accessing the CV model through the API.

Notably, the images sent to the various APIs have to be encoded in different formats. For Plant.id the images must be encoded as a string in base64 format. For all other APIs the images have to be sent as “multipart/form-data”.

Most APIs allow to add additional information to the request (besides the actual image).

For the iNaturalist API, the user can add further input variables, such as location and time, which help to classify the image. Further, the user can specify the language in which the common name should be returned. The API returns a list of results, which includes the scientific name, the common name and whether the detected species was already seen nearby, in the case a location was specified.

65 Python script to run the iNaturalist 2017 Competition model, https://github.com/deblagoj/iNaturalist-API/blob/master/run_pytorch_server.py (Retr. 2021-12-14)
66 Adapted python script, which is also working for the iNaturalist 2021 Competition model, https://github.com/EibSReM/iNaturalist_Competition (Retr. 2021-12-14)
67 GitHub repository of the Microsoft AI for Earth species classification API, https://github.com/microsoft/SpeciesClassification (Retr. 2021-12-14)
68 Adapted repository to let the Microsoft AI for Earth model run locally, https://github.com/EibSReM/MicrosoftSpeciesClassification (Retr. 2021-12-14)
69 Anaconda: Python package manager, https://www.anaconda.com/
70 curl: Command line tool for transferring data with URLs, https://curl.se/

For Pl@ntNet, also the response language can be specified. Besides, it is mandatory to specify the organ displayed. There were options for “leaf”, “flower”, “bark” and “fruit” at the time we conducted the tests, meanwhile they also added the option “auto”. The users can also specify whether they want to receive similar images, to the uploaded one. Next to these similar images, the users get the scientific name, the common name and the genus and family name of the identified plant.

For Plant.id, as with the iNaturalist API, it is also possible to specify the location, time, and response language. Furthermore, using the Plant.id API, it is decidable whether a fast answer or a higher accuracy or something in between is preferred. For our test, we used the “in between” default option. As a special feature, it is also possible to ask for further details of the plant and similar images. Plant.id provides information about the edible parts of the plant, and synonyms of the species name. It can also identify whether the plant has a disease and provides further details about the disease. The API returns the scientific and common name as well.

Since we do not have documentation for NIA, we do not know whether it is possible to add further variables to the request. As response, we just received the scientific names for the suggestions.

Table 5: This table shows different characteristic of the used APIs. It links the different documentations and coding examples. Further, it shows how the image must be encoded, which variables can be added to the request and which data about the suggestions is included in the response.

Model	**Documentation**	**Coding examples**	**Image input**	**Further input variables**	**Response data**
iNaturalist API	https://rapidapi.com/inaturalist-inaturalist-default/api/visionapi/ (only visible with approved account)	https://rapidapi.com/inaturalist-inaturalist-default/api/visionapi/ (only visible with approved account)	multipart/form-data	- location - time - response language	- scientific name - common name - already seen nearby
Pl@ntNet	https://my.plantnet.org/account/doc#openapi	https://github.com/plantnet/my.plantnet/blob/master/examples/post/run.py	multipart/form-data	- organ on image - response language - data which should be added in response	- scientific name - common name - family name - similar images
Plant.id	https://github.com/flowerchecker/Plant-id-API/wiki	https://github.com/flowerchecker/Plant-id-API/blob/master/python/sync_identification_example.py	encoded as string in base64	- location - time - response language - data which should be added in response - evaluation behavior	- scientific name - common name - plant details (e.g. edible parts, synonyms) - disease details - similar images
NIA	https://waarneming.nl/api/v1/docs/authentication-oauth2.md (only authentication)	n.a	multipart/form-data	n.a.	- scientific name

3.2 Methods to test the identified models

To determine how accurate the models can identify IASs of Union concern we developed a test strategy. We focused on the Invasive Alien Species of Union concern specified by the European Union Regulation 1143/2014, and candidate species that may be added to the list in the future (the next update of the Union list is expected in the 1st semester of 2022). The models were tested with images of these 96 species. For each species, we selected six images, which were used for testing. In three of the six images the species is well represented (including their most characteristic features) and easy to identify for an expert. These images are referred to as the "golden standard" images. The other three images we call "observation photos". These are mainly images which have been uploaded by real users to different applications, such as the "Invasive Alien Species Europe" smartphone application[71] or the Platform NatureMapr[72]. In contrast to the golden standard pictures, these pictures usually had a lower quality, and the species was not represented with all details. All images were taken from several providers with a different prioritization. The selection process is explained in the following section.

The JRC provided us with a dataset of pictures for each IAS of Union concern. From this dataset the first three images of each species were used as golden standard images. However, for some species only two golden standard pictures were available. Further, for the candidate species, no images at all were available in this provided dataset. For these species we searched for further images on the Commonwealth Agricultural Bureaux International (CABI)[73] database. The images from CABI are imported from scientific papers and therefore typically consist of a good quality and represent the species well. If there were not enough images on CABI, we searched on the Global Biodiversity Information Facility (GBIF) for further pictures. For the golden standard images, we prioritized images from "preserved specimens" instead of "human observations". To count as a golden standard, these images had to be of high quality and the species had to be easily recognizable.

For the observation photos, the JRC provided us validated pictures that were uploaded to the "Invasive Alien Species Europe" smartphone application. From this dataset, we took the three observation images if available. Sometimes not enough images were uploaded to the application. In this case we searched on GBIF for sufficient images. In contrast to the golden standard images, here "human observations" are prioritized over "preserved specimens". For a few species, there were also no images available on GBIF, so we had to search for images from further trustworthy sources on the internet. For the selection of all observation images, we chose images, where the species should be recognizable for a human at all. Additionally, the images should differ from each other, e.g. by showing different parts of a plant. This means for example instead of three images of the leaves, one of the leave, one of the blossoms and one covering the entire plant are used.

On GBIF, it is possible to find images from different providers, such as NatureMapr or Cornell Lab of Ornithology. We list all providers, we took images from, in Annex D[74]. Also, images uploaded to iNaturalist and Pl@ntNet can be found on GBIF. Since the computer vision models of these providers could have been trained with these images, we excluded them during our image search on GBIF. Otherwise, the models trained on the selected test images could have been advantaged during the testing. We did this for both the golden standard and observation images.

To classify all images by the models, we created three different repositories which contain the prepared scripts. All repositories are collected in a GitHub Organization[75]. Two repositories contain scripts to run the models locally, one each for the iNaturalist Competition[76] and Microsoft AI for Earth[77]. The third repository[78] contains a Jupyter Notebook[79], which sends the images to the different APIs and saves the result in a readable file. For all scripts, the images to test must be placed in one folder and the results are saved in a .csv file. With the selected images and scripts, we performed the tests for the different models. For all models, we just used images of

[71] "Invasive Alien Species Europe" smartphone application, https://visitors-centre.jrc.ec.europa.eu/en/media/tools/tracking-invasive-alien-species-europe-mobile-app

[72] NatureMapr, https://naturemapr.org/home

[73] Commonwealth Agricultural Bureaux International (CABI) database, http://www.cabi.org/isc

[74] Do not hesitate to ask the authors to get the exact image test set for reproducibility purposes.

[75] Github Organization which comprises all used repositories, https://github.com/EibSReM (Retr. 2021-12-14)

[76] Github Repository containing the scripts to run the iNaturalist Competition Model, https://github.com/EibSReM/ iNaturalist_Competition (Retr. 2021-12-14)

[77] Github Repository containing the scripts to run the Microsoft AI for Earth Model, https://github.com/EibSReM/ MicrosoftSpeciesClassification (Retr. 2021-12-14)

[78] GitHub Repository containing the Jupyter Notebook, which allows communication with different computer vision APIs, https://github.com/EibSReM/RequestCollectionComputerVisionAPIs (Retr. 2021-12-14)

[79] Jupyter Notebook: Web application for creating and sharing documents that contain code, visualizations, and text, https://jupyter.org/

species covered by the model. As a particularity for the Pl@ntNet API, where it was mandatory to specify the part of the plant depicted in the image, we had to add this information to the request manually.

All models returned as a result a list of suggested species. For the models we run on our machines, there was a parameter to adjust the number of suggestions. Some APIs always returned a fixed number of suggestions (ten for iNaturalist API, NIA and Flora Incognita). Other APIs returned the results in dependence to the probability. For example, if a model "was sure" that it detected the right species, there was just one suggestion, otherwise up to ten suggestions (Plant.id, Pl@ntNet API). Most models also included the score of the suggestion. However, we did not consider this score, because it can have different meanings - most times it is the probability, but not always (Flora Incognita) and sometimes we even did not receive any score (iNaturalist API). With the mentioned scripts we let all models classify the selected test images[80]. As result, we received for each model and tested image a list of identified species.

3.3 Calculating the accuracy of models

To evaluate the accuracy of the model, we analyzed the results. As first step, we checked on which rank the correct species was identified by the model. For example, for one image of the species *Acridotheres tristis*, we received suggestions in the order *Acridotheres tristis*, *Manorina melanocephala* and *Psilorhinus morio*, which then results in rank 1 for this image. As another example, for an image of the species *Ailanthus altissima* we got the suggestions in the order *Rhus typhina*, *Rhus lanceolata* and *Ailanthus altissima*, which then results in rank 3 for this image. If the correct species was not under the best five suggestions, the image got the rank 10. Out of these ranks, we calculated two different accuracies, Top-1 and Top-5. The Top-1 accuracy shows the percentage of images, where the first suggestion of the model was correct. To calculate this value we, divided the number of images that have rank 1, through all images tested. For the Top-5 accuracy, we divided the number of images which had rank 5 or better, through all images tested to get the percentage of images, which were classified correctly under the first five suggestions. First, we calculated the accuracies separately for the golden standard and user observation images of the IAS of Union concern. Second, we calculated the accuracy over all IAS of Union concern images, i.e. golden standard and user observations together. We did the same for images of the candidate IAS. Last, we also calculated the accuracies for all images from the candidates, as well as from the IAS of Union concern. As a result, it is possible to inspect how the accuracies of the models differ between golden standard and observation images, as well as the differences between images of listed IAS and candidates.

These Top-1 and Top-5 accuracies relate to all IAS covered by the model. Since the models covered a different number of IAS (for example, iNaturalist API with 82 species (Union concern + Candidates) in contrast to NIA with 49 species), the accuracies were calculated with a different number of images. For the iNaturalist API, we used 492 images (82 * 6) and for NIA just 294 images. Also, the covered species differed, and thus different images were used. The first Top-1 and Top-5 accuracy values give a good overview of how accurate the model is in general. However, it is difficult to compare the models among each other, because of the different test sets. Therefore, we checked which species are covered by all models. We did this in a more detailed way and identified which species are covered by all hybrid models (those which recognize both plants and animals), and which species are covered only by plant models. We again calculated the Top-1 and Top-5 accuracies as above, but only with the set of species covered by all/hybrid/plant models. This time all models had the same input images, thus the accuracies are better comparable between the models.

3.4 Results

After receiving the predictions of all images, we calculated the accuracies as mentioned above. All calculated accuracies are listed in Annex A. A smaller overview of the accuracies, we indicated as the most interesting, can be found in Table 6. In the second and the third column of Table 6, we added information about the number of species covered by each model. In the fourth column, we provide the Top-1 accuracies for the golden standard images of the currently listed IAS of Union concern. These images with a good quality were mainly provided by the JRC. The Microsoft AI for Earth Model reached the highest accuracy, with a value of 94.12 %. This means for 94.12 % of the images we tested, the model was able to identify the species correctly with the first suggestion. Second comes iNaturalist Competition (79.84 %), followed by Plant.id (79.49 %), Flora Incognita

80 All APIs were tested on November 23 and 24, 2021.

(75.32 %) and the iNaturalist API (73.68 %). These models are close together. The accuracies for the Pl@ntNet API[81] (53.85 %) and NIA (52.03 %) are not as high as the ones from the other providers.

Next, in the fifth column, we present the Top-1 accuracies of the listed IAS for the observation images. Here, the models were tested with the same number of images as for the golden standard images. All lose accuracy compared to the golden standard images. One reason for this could be that the quality of these images is lower than the quality of the golden standard images. Observation photos have been taken with a smartphone and often not cover the complete species on the image. Microsoft AI for Earth still provides the highest accuracy with 83.33 %. Second in terms of the user observations is the iNaturalist model with 70.18 %. So, it also performs well on observation photos. The reason for this might be that the model was trained on photos from users. In third place for the observations is Plant.id (65.38 %), followed by iNaturalist Competition (64.34 %) and Flora Incognita (60.49 %). These models performed way better with golden standard images than with user observations. The Pl@ntNet API (52.56 %) performed as good as with the golden standard images, but still has a lower accuracy than the other models. NIA for this test again has the lowest accuracy.

In column six, we show the accuracies over all tested images, including golden standard and user observation images of candidates and already listed IAS. Here, it is important to mention that the models performed less accurate for the candidates than for the listed species, especially for the golden standard images. These accuracies are provided in Annex A. The reason for this might be that we were not provided with images for the candidates by the JRC. The quality of images we took from CABI or GBIF was lower than the quality of the images provided by the JRC. For all images, again, Microsoft AI for Earth (78.89 %) has the highest accuracy followed by the model from Plant.id (70.97 %) and Flora Incognita (66.67 %). Next, we have the iNaturalist API with 65.68 % followed by iNaturalist Competition (62.12 %). This shows that the iNaturalist API works better than the iNaturalist Competition model with photos which have not the best quality. Over all photos, the iNaturalist API performed better, whereas for the golden standard images, the iNaturalist Competition model did. The lowest percentage of images were correctly identified by Pl@ntNet API (54.44 %) and NIA (41.16 %).

In the next column, we provide the Top-5 accuracy over all images. These values show the percentages of images, where the correct species was under the first five suggestions the model provided. The Top-5 accuracies of the models are higher than the Top-1 accuracies. Especially, models covering only plants show a better accuracy. One reason for this could be the existence of plants looking very similar, for example, for an image of *Heracleum Mantegazzianum,* Plant.id suggested: *Heracleum Maximum*, *Heracleum Sphondylium* and *Heracleum Mantegazzianum*. The correct result was on rank 3. Thus, the model was quite able to detect that the plant is of the genus *Heracleum*, but could not detect the correct species on rank 1. For the Top-5 accuracy, Plant.id has the highest accuracy with (93.01 %), followed by Microsoft AI for Earth (91.48 %), Flora Incognita (88.17 %), iNaturalist API (84.55 %) and iNaturalist Competition (84.18 %). Even the Pl@ntNet API has significantly increased its accuracy from Top-1 to Top-5, but still cannot catch up to the others with 73.33 %. The lowest accuracy achieved NIA with 51.02 %.

As already mentioned above, we used only images of species covered by the model. We had different test sets for the models with diverse images. To really compare the models on the same test set, we created three different tests. In the first set there are images of animals and plants, which are covered by all hybrid models (iNaturalist API, iNaturalist Competition, Microsoft AI for Earth and NIA). This test set comprises 30 species (candidates + listed), so 180 images (three gold + three user observations per species). The Top-1 accuracies of this test set can be seen in the eighth column. Also, for this test set Microsoft AI for Earth has the best accuracy with 86.11 %, followed by the iNaturalist API (73.89 %), iNaturalist Competition (69.74 %) and lastly NIA (46.11 %).

We did the same for the models, only covering plants (Pl@ntNet API, Flora Incognita and Plant.id). This test set comprises 21 plants, so 126 images. The results can also be inspected in the eighth column below the black line. The best accuracy was again achieved by Plant.id with 73.81 % closely followed by Flora Incognita with 72.80 %. The lowest accuracy achieved the Pl@ntNet API with 54.78 %.

Interestingly, all the accuracies for these test sets are higher than the accuracies for all the images. The reason for this might be that species covered by all models are easier to identify than other species. Another reason could be that for these species, a lot of training images are available. We also created a test set of species, which contains species covered by all tested models. This test set contains twelve plants. The results can be

[81] After we completed our work (and thanks to our findings), the people in charge of the Pl@ntNet platform recognised that the my.plantnet.org API – which was subject to our testing - had a serious bug at the time of the evaluation and that this bug did not exist in the Pl@ntNet web and mobile applications. Higher accuracies can be expected once this bug will be fixed.

found in Annex A. In Annex A, you can also find the different Top-1 and Top-5 accuracies separated, in listed and candidate IAS and in golden standard and user observation images.

Table 6: Short list models with IAS covered and certain accuracies. All accuracies are percentages. T1 means, that the Top-1 accuracy was calculated, whereas T5, means the Top-5 accuracy. The G in parenthesis stands for the test set of golden standard images, the U for user observation images. IAS means that images of the currently listed IAS were used, candidates indicates that images of the candidate species were used. With "same species", we mean that the same test set was used for all models and all species were covered by each model (in this case distinguished between images for hybrid (above the bold line) and plant (below the bold line) models). For instance, in the fourth column this means: Top-1 accuracy of golden standard images from listed IAS.

		Candidates covered	**Accuracy**				
		IAS + Candidates	**IAS**		**IAS + Candidates**		**Same species**
Model	**IAS covered**	**T1 (G+U)**	**T1 (G)**	**T1(U)**	**T1 (G+U)**	**T5 (G+U)**	**T1 (G+U)**
iNaturalist API	57	25	73.68	70.18	65.85	84.55	73.89
iNaturalist 2021 Competition	43	16	79.84	64.34	64.12	84.18	69.74
Microsoft AI for Earth	34	11	94.12	83.33	78.89	91.48	86.11
NIA	41	8	52.03	43.90	41.16	51.02	46.11
Pl@ntNet API	26 (out of 36 plants)	4 (out of 6 plants)	53.85	52.56	54.44	73.33	54.78
Flora Incognita	25 (out of 36 plants)	4 (out of 6 plants)	75.31	60.49	66.67	88.17	72.80
Plant.id	26 (out of 36 plants)	5 (out of 6 plants)	79.49	65.38	70.97	93.01	73.81

4 Recommendations on the future use of image-based species recognition models in support of citizen science

After investigating different aspects of image-based species recognition models in the previous sections, we summarize the results in this section and we give recommendations on their future use in support of citizen science to especially recognize invasive alien species in Europe. The recommendations are split into different parts. First, the best models regarding species coverage will be determined, then the models are evaluated on their usability and accessibility and afterwards the model performance is analyzed to recognize the correct species, i.e. the accuracy. At the end, overall recommendations are given, depending on different scenarios.

All recommendations and conclusions are formulated in relation to the seven models on the short list (see Section 3).

4.1 Species coverage

The iNaturalist API includes the most species with 38,000, increasing to 47,000 species with their upcoming update. Microsoft (5,266) and Flora Incognita (4,803) cover the fewest species and the others lay in between. Pl@ntNet (36,870), NIA (22,302), Plant.id (12,183) and Flora Incognita are also updated regularly, i.e. they are worth to observe further, whereby Microsoft and iNaturalist 2021 Competition are not.

Out of the 66 IAS of Union concern, the iNaturalist API covers the most species in their model with 57 (64 after the update), followed by the iNaturalist Competition model, NIA and the Microsoft model (Annex B). The plant species recognition models Pl@ntNet, Plant.id and Flora Incognita all cover a similar number of species with 25 or 26 out of 36 plant species. Additionally, they cover very similar plants since 18 plants are covered by all three models and four plants are covered by none of them (*Ehrharta calycina*, *Heracleum sosnowskyi*, *Pueraria montana var. Lobata*, *Salvinia molesta*) with three of these four plants covered by the iNaturalist API that is not specialized on plants. One plant (*Pueraria montana var. lobata*) is covered by no model. The same holds for two animals (*Arthurdendyus triangulatus*, *Procambarus fallax f. virginalis*). The reason for *Pueraria montana var. Lobata* not to be included in any model could be that it is a subspecies and some models might not include all varieties but just the species. Depending on the differences between the subspecies it might thus be an opportunity to recognize the species instead of the subspecies as IAS in citizen science use cases.

Following these species coverages it would not be very useful to combine different models in order to get a better coverage since even with a combination not all species would be covered. Especially, it would not help a lot to combine different plant models with each other or a plant model with the iNaturalist API. After the update of the iNaturalist API *Pueraria montana var. Lobata* and *Vespa velutina nigrithorax* will be the only IAS of Union concern not covered by this model, keeping *Pueraria montana var. lobata* the only species not covered by any model since *Vespa velutina nigrithorax* is covered by NIA. The easiest way to cover all IAS would probably be to contact iNaturalist and ask them to include the two missing species in their next update, since they also have several observations of them in their database.

For the local lists of IAS (Appendices B.3 - B.6) we identified a similar situation: the iNaturalist API covers the most species and Microsoft the least, but here some differences in the plant models emerge. Pl@ntNet includes 101 out of 112 species, Flora Incognita 73 and Plant.id 91. Additionally, they now cover some plants that cannot be specified by the other models. However, again, several species (43 out of 233) are not covered by any of the models. The IAS candidates are also covered best by the iNaturalist API with 25 out of 30 and the plant models covering four or five of the six plant species. Four candidate species are not covered by any of the models. Looking at the species from the local lists and candidates not covered by any model, it is noticeable that they are mostly fish and invertebrates, in particular invertebrates living in the water (e.g. see Annex B.3 Danube list). Also some plants (*Mesembryanthemum lancifolium, Phytophthora cinnamomi, Reynoutria × bohemica, Rugulopteryx okamurae*) are among them, but they seem not to have an obvious similarity as they are different plant types (flowering plants, fungus-like, algae) and might not be covered for different reasons.

Summarizing, all models cover similar species of their domain, but it is difficult to find a reason for this. It is noticeable that species living in the water, e.g. *Perccottus glenii* or *Arthurdendyus triangulatus,* are covered less often than others, probably because there is not so much footage of them. But also more camera-accessible species, such as some plants (e.g. *Ehrharta calycina*, *Pueraria montana var. lobata*) are covered rarely or not at all. Mammals and birds are in general covered well, by at least one of the models.

All in all, it can be recommended to use the iNaturalist API to cover as many species as possible. It includes most of the IAS and also the highest number of species in general offered by all models. Moreover, the iNaturalist Computer Vision Model is updated regularly. Another opportunity would be to train an own custom model to cover all IAS species. Therefore, the Microsoft AI for Earth species recognition model or the iNaturalist 2021 Competition model could be used as a basis. Both of their public GitHub repositories, including links to the models, are published under the MIT License and thus free to use and modify. Among them the iNaturalist 2021 Competition model covers more species, which will not change in the future since both models are not updated anymore.

Another finding is that it would not be very useful to combine the investigated models to cover more species. However, beyond the investigated models, it might be helpful to combine different models, for example including a model specialized on aquatic species could increase the species coverage because many of the not covered species are living in the water. To achieve this a contact to the provider of FishID can be of interest as they create custom models for fish and water creatures but did not provide a model we could test in more detail yet. To use such a combination of models would make it necessary to create another preceding model that detects a more abstract class of species (e.g. fish, plant, invertebrate, bird, and mammal) and distinguishes between them to start the detailed species recognition with a specialized model afterwards.

4.2 Accessibility and usability

There were mainly two different kinds of models available for our testing purposes: Models freely available for download and deployment on an own machine and models available through an API. The Microsoft and the iNaturalist Competition model are among the former, the others are among the latter. The APIs from iNaturalist, NIA, Pl@ntNet and Plant.id are available with different request limits and costs for further requests. We were not granted access to the Flora Incognita identification service because it was too much effort for our purposes in the provider's opinion, but they did the inference themselves for us and sent us the raw data of the results. The different access types also allow different kinds of adaption of the model. The models available for download can also be adapted independently, i.e. they can be trained on further species. One drawback associated with this is that the models are oftentimes (in our case all examined models) not updated anymore by the providers and it can be a high effort to train the models. However, the benefit of such a self-trained model is, there are no external costs that need to be paid regularly, e.g. on a monthly basis, once the model is finished. While the APIs are not available to be adapted independently, the providers of Flora Incognita, NIA and Plant.id offered different collaborations to improve their model and make it applicable to the IAS use case. The modalities would depend on the exact needs (of the JRC) and case-by-case discussions. The models from iNaturalist and Pl@ntNet are regularly updated depending on validated user observations.

The models also differ in the number of images that can be analyzed for one prediction and the information given as results. Microsoft, iNaturalist Competition and iNaturalist API use exactly one input image for a prediction, Flora Incognita uses one to three, the Pl@ntNet API and Plant.id take one to five images and NIA can use one or more images. The number of results (the amount of species "guessed") for the iNaturalist API, the Pl@ntNet API and Plant.id is depending on the highest score of the predictions, i.e. if the model is determining a species with a high percentage, fewer species are listed than for cases the model was unsure about the species, hence having low scores for the first predictions. NIA and Flora Incognita predictions provide ten results every time and the number of results for the downloaded models can be adjusted. One special attribute of the Pl@ntNet API and Flora Incognita models is, it can be determined which part of a plant is photographed to improve the prediction. For the Pl@ntNet API it is mandatory to specify which part of the plant is depicted, Plant.id also takes images without having this information. Regarding the results, iNaturalist has the special property of not returning a score with the prediction but just the ranking, different from the others, which all also return the score.

In summary, the recommendation regarding the accessibility and usability is to use the models available for download, iNaturalist Competition and Microsoft AI for Earth. They are free to use, allow unlimited numbers of requests and responses and can be adapted independently to add further species.

4.3 Accuracies

Two extremes are noticeable regarding the accuracies. In most cases, the Microsoft model has the highest accuracies and NIA the lowest. Especially for the golden standard images of the IAS, Microsoft performed well and got a bit poorer for the observations and candidates. For the candidates and observations the iNaturalist

API, the Pl@ntNet API, Plant.id and Flora Incognita have all similar values. It has to be considered that these images were all selected by ourselves and might lead to some biases that influence the accuracy. For example, if a model covers a species with bad image availability it is more difficult to recognize the correct species, which decreases the accuracy. NIA has the lowest accuracy in all categories and as it also only covers species from Belgium and Netherlands it can currently not be recommended for any use case of Union concern.

For the models recognizing only plants, Plant.id has the highest accuracies for almost all categories, closely followed by Flora Incognita which has slightly poorer accuracies most of the time. It was also tested how the models performed on exactly the same images for the twelve (plant) species that were covered by all models. Again, Microsoft and Plant.id have the highest Top-1 accuracies and NIA the poorest. The other models lie in between in the order from best to worst: Flora Incognita, iNaturalist API, Pl@ntNet API, iNaturalist Competition.

Evaluating the accuracies, also the total number of species that can be recognized by a model has to be considered. Since we only tested the models on the species they covered, a model with fewer overall species has a lower chance of making a wrong prediction since not so many other species could be recognized. Thus, the accuracies of models with higher numbers of species like the iNaturalist API or the Pl@ntNet API have to be ranked higher and the good accuracies of Microsoft and Flora Incognita with fewer species should be ranked lower.

Altogether, the Microsoft AI for Earth species recognition model outperformed the other models and is recommended to determine species in images most accurately. As an alternative to reach high accuracies while covering more species, different models could be combined. For example, the iNaturalist Competition model, covering many species, can be combined with Flora Incognita's high accuracy model. Additionally, the provider of Flora Incognita said they wanted to cover almost all plant IAS with their next update in 2022. However, still some species would not be covered with such combinations of models. Keeping in mind the overall high number of species covered also the iNaturalist API has good accuracies. Just considering the accuracy of the currently covered species it can be recommended to use the Microsoft model. To add some more species without losing much accuracy, the Microsoft model could be combined with Plant.id, as the plant model with the highest accuracies.

4.4 Overall recommendations

Depending on what has to be achieved by the models there are different options that can be chosen, including some compromises. Possible achievements are, to cover all IAS, to have no external costs or to put the least effort in the development of the model. Different advantages and disadvantages of the models are listed in Table 7.

To cover all IAS no investigated model on its own or a combination of models is sufficient. To achieve full coverage, either a custom model has to be trained or a collaboration with one of the model providers has to be established. To train an own model, the Microsoft AI for Earth is a good basis, since it is freely available and has the best accuracies. The downside is, it is not updated in the future and covers less species than the other adaptable model from the iNaturalist 2021 Competition. An opportunity for a collaboration with Naturalis, as one of the providers of NIA, was offered to us during our work. Naturalis mentioned that an initiative for invasive species could be part of their consortium, which is updated regularly with new European countries. However, this already shows their drawback of currently covering only Belgium and the Netherlands, extending to Northern Europe in the next year. Additionally, we found a relatively low accuracy in recognizing IAS in the images we used. Another provider that offered a collaboration is Plant.id. Working together with them only makes sense if different models for the recognition of plants and animals are used, but if this is the case, Plant.id is an option with high accuracies and a high species coverage. A well suited compromise could be to reach out to iNaturalist and ask them to include the few missing IAS in the next update of their Computer Vision model.

To avoid external costs an own model can be trained. Therefore, the Microsoft model is a good basis which can be fine-tuned by training further species. Yet, we did not investigate what has to be done to create a custom model out of the given models. For this purpose, some steps would be to collect a sufficient number of good images for each species, label them and train the model with them. Maybe it would also mean to remove the top layer of the current model and to train all the species already included in the model again. Another issue with training an own model is the similarity of IAS with non-invasive species. It again means a high effort to determine species that are similar to IAS and also train the model on them, including the gathering of images. Maintaining an own model can also include creating the model from scratch. Therefore different existing datasets, e.g. from iNaturalist, Microsoft or GBIF (Annex C) can be used to get around the effort of collecting images for all the species separately. Thus, avoiding external costs leads to high internal costs to develop and maintain an own model.

Having the least effort to implement a model means to use one of the existing models, without further adaptations. To achieve this, the two options are first to use one of the free models and just deploy them as they are, to avoid costs, or second to use the iNaturalist API which is covering the most species. Correspondingly, the disadvantage of the 'free' models is the low species coverage, with the iNaturalist Competition model covering more species than the Microsoft model, but the Microsoft model having higher accuracies. The disadvantages of the iNaturalist model are the constant fees that apply.

As the overall summary we would recommend to use the iNaturalist API for image-based species recognition of IAS and to get in contact with iNaturalist and set up a collaboration, to include the missing species as fast as possible. This would mean an easy and fast solution without the need of putting a high effort in the implementation and maintenance of an own model. The other option is to maintain an own model to avoid fees and the dependency on an external provider by simultaneously controlling the species covered by the model with the drawback of a high effort.

Table 7: Models and their main advantages and disadvantages as well as the consequential recommendations.

Model	Advantages	Disadvantages	Recommendation
iNaturalist API	Most species, good accuracies also for observations	Fees	Best choice; ask for collaboration to add missing IAS
iNaturalist Competition	Free, adaptable	Not updated	Combine with other models for higher accuracy and more species or for fine-tuning
Microsoft	Best accuracies, free, adaptable	Not updated, few species	Use as basis for training of custom model
NIA	Offered collaboration	Worst accuracies, low geographical coverage	Check for improvements after update; do not use without improvements
Pl@ntNet API	Easy access to API, covers most plants from local lists	Fees (after 500 requests a day); worst plant accuracies	Use to cover most local plant IAS
Flora Incognita	High accuracies	Fees, accessibility not clear	Check coverage after update and compare to Plant.id
Plant.id	High accuracies; offered collaboration	Fees	Combine with animal model

4.5 Limitations of this work

Within our research, we faced different limitations, especially regarding the transparency of the models and the preparation of images for our test purposes. Some of them also have to be considered during further steps.

Most of the models that were presented in this report were not available for download to be deployed separately but were just accessible via an API. This includes that it was difficult or sometimes even not possible to find a list of species covered by the model without getting in touch with the providers. In general, it was not clear for most of the models how the input data are used within the models, i.e. which images are used for training, how the input images are augmented or how models are updated and which criteria are applied to include additional species. Some models also allow uploading several images for one prediction but it is not described how the images are used for the prediction, i.e. whether a model can process multiple images at the same time for one prediction or if each image is analyzed on its own and then the predictions are summarized to one value. It behaves similarly for additional information that can be forwarded to some APIs (such as location or season), where it is also not clear how they are included in the prediction. Moreover, the APIs allow just a small number of free requests and after using them it is necessary to pay fees, although some providers allowed a higher amount of free requests for our research. The available pre-trained models were often difficult to deploy since it was not clear how to use them or how far they are trained, and the labels related to the IDs as classification result were mostly not published.

Another issue is the determination of the species covered by a model on the basis of the species name. To check the coverage we used the scientific name and compared the list of IAS with the lists of species covered by each model. But some species have several scientific names (e.g. *Orconectes virilis* = *Faxonius virilis*) and because of this a species might be declared as not covered by a model although it is, but with another name. To tackle this issue we looked up synonyms for scientific names in CABI's Invasive Species Compendium as far as possible and if a model was able to recognize one of the synonyms it was categorized as covering the species. All scientific names are added to the lists in Annex B.

Additionally, it was difficult and sometimes a high effort to find good images to test the models on. In particular gathering images for the IAS candidates was difficult since we were not provided with any images for them and had to look up golden standard and user observation images. This can also be important for the training of a custom model because for that many images of each species are necessary to create a good model. Furthermore, we cannot be entirely sure the images taken from GBIF or other sources are classified correctly although they were labeled as verified. Many of the platforms verify user observations by other users which can lead to wrong verifications, but we had to assume they were right.

4.6 Future work

As a consequence of the results and recommendations mentioned in the previous sections, different opportunities for future work emerge, in particular regarding the evaluation of the presented models.

Beside the Top-1 and Top-5 accuracy of the model predictions on the images of IAS of Union concern and the IAS candidates, further metrics could be calculated. Top-3 or Top-10 accuracies, accuracies for other species than the IAS or the accuracies for the species of the local lists could be investigated. Calculating accuracies for species other than the IAS, in particular species covered by all the models, could be of interest to better compare the performance of the models on the same images. Moreover, it could be evaluated how the models perform on distinguishing between IAS and similar species that are not invasive by classifying both types of species and calculating the Top-1 accuracy. Beforehand, it has to be checked whether the respective species are covered by the model. Beside the performance on non-invasive species the model can also be checked on what is recognized in images not containing any plant or animal and if the model is able to determine that there is nothing of interest in an image.

Another topic to further investigate is the model performance depending on the kind and quality of input images and parameters. For instance some model APIs provide the opportunity to provide multiple images for one classification, add further information like the location or the season or to state which part of a plant is depicted. All these features could be investigated to determine how much they increase the model accuracy. Plant.id and iNaturalist for example mentioned in one of their mails they also train their models on images with bad quality and thus such images can also be classified correctly, but of course it is more difficult, especially if no distinct features of the species are visible.

Furthermore, an interesting future work is to evaluate how the images should look to get the best predictions. For instance properties, such as the image size (i.e. how many pixels), the proportion of area of interest within an image (i.e. how many percent of the pixels of an image are covered by an animal or plant) or the characteristic features of a species could be investigated. The important characteristics to improve the prediction can then be communicated to the user of the model, for example advice could be given on how close the camera has to be to the object or which parts of the object should be photographed.

Additionally, it would be of interest to evaluate the opportunities of a custom model for image-based recognition of IAS. Here, interesting topics could be to determine and distinguish the effort and value it would mean to train an own model from scratch, train a IAS recognition model based on pre-trained weights of a general image recognition model or to fine-tune an existing species recognition model. In general, it can already be said that it would be a high effort to train a custom model and maintain it, caused for instance by the effort to gather enough images for each species to train a useful model. As the models available for adaptation, we tested the Microsoft AI for Earth species recognition model and the iNaturalist 2021 Competition model. Both are available in different formats and require different steps for fine-tuning. The Microsoft model can be downloaded as a model ready for production and it is necessary to restore the training checkpoints of the model to fine-tune it. The iNaturalist Competition model is also available without the top layer and could be used for training directly. To use the models later it is probably necessary to train them again also with the images and species the finished models are already trained on, in addition to the (invasive) species that need to be added.

References

Adriaens, T., Tricarico, E., Reyserhove, L., De Jesus Cardoso, A., Gervasini, E., Lopez Canizares, C., Mitton, I., Schade, S., Spinelli, F.A. and Tsiamis, K., Data-validation solutions for citizen science data on invasive alien species, EUR 30857 EN, Publications Office of the European Union, Luxembourg, 2021, ISBN 978-92-76-42055-2, doi:10.2760/694386, JRC126140. http://dx.doi.org/10.2760/694386

Åkerblom, M., Pasi Raumonen, Raisa Mäkipää, Mikko Kaasalainen, (2017) Automatic tree species recognition with quantitative structure models, Remote Sensing of Environment, Volume 191, 2017, Pages 1-12, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2016.12.002

Choiński, M., Rogowski, M., Tynecki, P., Kuijper, D. P. J., Churski, M., & Bubnicki, J. W. (2021). A First Step Towards Automated Species Recognition from Camera Trap Images of Mammals Using AI in a European Temperate Forest. In K. Saeed & J. Dvorský (Eds.), Computer Information Systems and Industrial Management (pp. 299–310). Springer International Publishing. https://doi.org/10.1007/978-3-030-84340-3_24

Goëau H, Bonnet P, Joly A (2021) AI-based Identification of Plant Photographs from Herbarium Specimens. Biodiversity Information Science and Standards 5: e73751. https://doi.org/10.3897/biss.5.73751

Goëau H, Bonnet P, Joly A (2018) Overview of ExpertLifeCLEF 2018: how far automated identification systems are from the best experts ? CLEF task overview 2018, CLEF: Conference and Labs of the Evaluation Forum, Sep. 2018, Avignon, France.

Jones, H. G. (2020). What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora. AoB Plants, 12(6), plaa052.

Lotfian, M.,Ingensand, J., Brovelli, M.A. The Partnership of Citizen Science and Machine Learning: Benefits, Risks, and Future Challenges for Engagement, Data Collection, and Data Quality. Sustainability 2021, 13, 8087. https://doi.org/10.3390/su13148087

Miele, V., Dussert, G., Spataro, B., Chamaillé-Jammes, S., Allainé, D., Bonenfant, C. (2021) Revisiting animal photo-identification using deep metric learning and network analysis, https://doi.org/10.1111/2041-210X.13577

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. Methods in Ecology and Evolution, 12, 150– 161. https://doi.org/10.1111/2041-210X.13504

Norouzzadeh, M. S., et al. (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning, https://doi.org/10.1073/pnas.1719367115

Schermer M, Hogeweg L (2018) Supporting citizen scientists with automatic species identification using deep learning image recognition models. Biodiversity Information Science and Standards 2: e25268. https://doi.org/10.3897/biss.2.25268

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. Ecology and Evolution, 10(7), 3503– 3517. https://doi.org/10.1002/ece3.6147

Sulc, M., Picek, L., & Matas, J. (2018, September). Plant Recognition by Inception Networks with Test-time Class Prior Estimation. In CLEF (Working Notes). http://ceur-ws.org/Vol-2125/paper_152.pdf

Swanson, A., Kosmala, M., Lintott, C. et al. (2015) Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci Data 2, 150026. https://doi.org/10.1038/sdata.2015.26

Tabak, M. A. et al. (2018) Machine learning to classify animal species in camera trap images: Applications in ecology, https://doi.org/10.1111/2041-210X.13120

Terry, JCD, Roy, HE, August, TA. (2020) Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data. Methods Ecol Evol.; 11: 303– 315. https://doi.org/10.1111/2041-210X.13335

Van Horn G, Cole E, Beery S, Wilber K, Belongie S, Mac Aodha O, et al. (2021) Benchmarking Representation Learning for Natural World Image Collections. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 12884-12893. http://arxiv-export-lb.library.cornell.edu/pdf/2103.16483

Van Horn, G., Mac Aodha, M., Song, Y., Cui, Y., Sun, C., Shepard, A. (2018). The inaturalist species classification and detection dataset. In: Proceedings of the IEEE conference on computer vision and pattern recognition. S. 8769-8778. https://doi.org/10.1109/CVPR.2018.00914

Wäldchen, J., & Mäder, P. (2018). Machine learning for image based species identification. Methods in Ecology and Evolution, 9(11), 2216-2225. https://doi.org/10.1111/2041-210X.13075

Weinstein, BG. (2018) A computer vision for animal ecology. J Anim Ecol.; 87: 533– 545. https://doi.org/10.1111/1365-2656.12780

Willi, M, Pitman, RT, Cardoso, AW, et al. (2019) Identifying animal species in camera trap images using deep learning and citizen science. Methods Ecol Evol.; 10: 80– 91. https://doi.org/10.1111/2041-210X.13099

List of boxes

List of figures

List of tables

Annexes

Annex A: Short List of Models with Accuracies

All accuracies are percentages. T1 means that the Top-1 accuracy was calculated, whereas T5, means the Top-5 accuracy. The G in brackets stands for the test set of golden standard images, the U for user observation images. IAS means that images of the currently listed IAS were used, candidates indicated that images of the candidate species were used. With “same species”, we mean that the same test set was used for all models and all species in this test set were covered by each model. For this, there were three different test sets. First, with species covered by all models, second with species covered by all plant models and third species covered by all hybrid models.

	IAS Accuracy						Candidates Accuracy					
Model	T1 (G)	T5 (G)	T1 (U)	T5 (U)	T1 (G+U)	T5 (G+U)	T1 (G)	T5 (G)	T1 (U)	T5 (U)	T1 (G+U)	T5 (G+U)
iNaturalist API	73.68	94.15	70.18	88.30	71.93	91.23	54.67	70.67	49.33	68.00	52.00	69.33
iNaturalist 2021 Competition	79.84	94.57	64.34	86.05	72.09	90.31	41.67	70.83	43.75	64.58	42.71	67.71
Microsoft AI for Earth	94.12	98.04	83.33	96.08	88.73	97.06	51.52	75.76	45.45	72.73	48.48	74.24
NIA	52.03	59.35	43.90	53.66	47.97	56.50	0.00	20.83	12.50	25.00	6.25	22.92
Pl@ntNet API	53.85	74.36	52.56	71.79	53.21	73.08	66.67	75.00	58.33	75.00	62.50	75.00
Flora Incognita	75.31	97.53	60.49	82.72	67.90	90.12	33.33	66.67	83.33	83.33	58.33	75.00
Plant.id	79.49	96.15	65.38	91.03	72.44	93.59	60.00	86.67	66.67	93.33	63.33	90.00

Model	IAS covered	Candidates covered	IAS + Candidates Accuracy						Accuracy same species					
									Top-1 (Gold + obs, incl candidates)			Top-5 (Gold + obs, incl candidates)		
			T1 (G)	T5 (G)	T1 (U)	T5 (U)	T1 (G+U)	T5 (G+U)	Plant models	Hybrid models	Hybrid & Plant models	Plant models	Hybrid models	Hybrid & Plant models
iNaturalist API	57	25	67.89	86.99	63.82	82.11	65.85	84.55	/	73.89	61.11	/	87.78	79.17
iNaturalist 2021 Competition	43	16	69.49	88.14	58.76	80.23	64.12	84.18	/	69.74	58.00	/	86.18	78.00
Microsoft AI for Earth	34	11	83.70	92.59	74.07	90.37	78.89	91.48	/	86.11	80.56	/	93.89	90.28
NIA	41	8	43.54	53.06	38.78	48.98	41.16	51.02	/	46.11	27.78	/	54.44	44.44
Pl@ntNet API	26 (out of 36 plants)	4 (out of 6 plants)	55.56	74.44	53.33	72.22	54.44	73.33	54.78	/	58.33	71.30	/	69.44
Flora Incognita	25 (out of 36 plants)	4 (out of 6 plants)	69.89	93.55	63.44	82.80	66.67	88.17	72.80	/	75.00	87.20	/	88.89
Plant.id	26 (out of 36 plants)	5 (out of 6 plants)	76.34	94.62	65.59	91.40	70.97	93.01	73.81	/	80.56	92.86	/	94.44

Annex B: Species covered

Tables of species covered by the image-recognition models and APIs, divided into different regions. First IAS, listing all invasive alien species currently of Union concern, second the candidates that might be added to the IAS during the next update and then the local lists. Yes means the species is covered by the model (with one of the listed scientific names), no means it is not.

Models: PI (Plant.id), PN (Pl@ntNet API), FI (Flora Incognita), iNC (iNaturalist 2021 Competition), NIA, iNA (iNaturalist API), Mic (Microsoft AI for Earth Species Classification).

Annex B.1 IAS of Union concern

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Acacia saligna	plant	yes	yes	yes	yes	no	yes	no
Acridotheres tristis	bird	no	no	no	yes	yes	yes	yes
Ailanthus altissima	plant	yes	yes	yes	yes	yes	yes	yes
Alopochen aegyptiacus	bird	no	no	no	yes	yes	yes	yes
Alternanthera philoxeroides	plant	yes	yes	yes	yes	no	yes	yes
Andropogon virginicus	plant	yes	no	yes	yes	no	yes	no
Arthurdendyus triangulatus	invertebrates	no	no	no	no	no	no	no
Asclepias syriaca	plant	yes	yes	yes	yes	yes	yes	yes
Baccharis halimifolia	plant	yes	yes	yes	yes	yes	yes	yes
Cabomba caroliniana	plant	yes	yes	yes	no	yes	yes	no
Callosciurus erythraeus	mammal	no	no	no	yes	yes	yes	no
Cardiospermum grandiflorum	plant	no	yes	no	no	no	yes	no
Cortaderia jubata	plant	yes	no	no	yes	no	yes	yes
Corvus splendens	bird	no	no	no	yes	yes	yes	yes
Ehrharta calycina	plant	no	no	no	no	no	yes	no
Eichhornia crassipes / Pontaderia crassipes	plant	yes	yes	yes	yes	yes	yes	yes
Elodea nuttallii	plant	no	yes	yes	no	yes	yes	no
Eriocheir sinensis	invertebrates	no	no	no	no	yes	yes	no
Gunnera tinctoria	plant	yes	yes	yes	yes	yes	yes	no

Gymnocoronis spilanthoides	plant	no	yes	no	no	yes	no	no
Heracleum mantegazzianum	plant	yes	yes	yes	yes	yes	yes	no
Heracleum persicum	plant	no	no	yes	no	no	no	no
Heracleum sosnowskyi	plant	no	no	no	yes	no	yes	no
Herpestes javanicus	mammal	no	no	no	no	no	no	yes
Humulus scandens	plant	yes	no	no	no	no	yes	no
Hydrocotyle ranunculoides	plant	yes	yes	yes	yes	yes	yes	yes
Impatiens glandulifera	plant	yes	yes	yes	yes	yes	yes	yes
Lagarosiphon major	plant	no	yes	yes	no	yes	no	no
Lepomis gibbosus	fish	no	no	no	yes	yes	yes	yes
Lespedeza juncea var. sericea (=Lespedeza cuneata)	plant	yes	no	yes	yes	no	yes	yes
Lithobates catesbeianus	amphibian	no	no	no	yes	yes	yes	yes
Ludwigia grandiflora	plant	yes	yes	no	no	yes	no	no
Ludwigia peploides	plant	yes	yes	yes	yes	yes	yes	yes
Lygodium japonicum	plant	yes	yes	yes	yes	no	yes	yes
Lysichiton americanus	plant	yes	yes	yes	yes	yes	yes	yes
Microstegium vimineum	plant	yes	yes	yes	yes	no	yes	yes
Muntiacus reevesi	mammal	no	no	no	no	yes	yes	no
Myocastor coypus	mammal	no	no	no	yes	yes	yes	yes
Myriophyllum aquaticum	plant	yes	yes	yes	yes	yes	yes	yes
Myriophyllum heterophyllum	plant	no	yes	yes	no	yes	yes	no
Nasua nasua	mammal	no	no	no	yes	no	yes	no
Nyctereutes procyonoides	mammal	no	no	no	no	yes	yes	no
Ondatra zibethicus	mammal	no	no	no	yes	yes	yes	yes

Orconectes limosus (= Faxonius limosus)	invertebrates	no	no	no	no	yes	yes	no
Orconectes virilis (= Faxonius virilis)	invertebrates	no	no	no	yes	yes	yes	no
Oxyura jamaicensis	bird	no	no	no	yes	yes	yes	yes
Pacifastacus leniusculus	invertebrates	no	no	no	yes	yes	yes	yes
Parthenium hysterophorus	plant	yes	yes	yes	yes	no	yes	yes
Pennisetum setaceum / Cenchrus setaceus	plant	yes	yes	no	yes	no	yes	yes
Perccottus glenii	fish	no	no	no	no	no	yes	no
Persicaria perfoliata	plant	yes	no	yes	yes	no	yes	yes
Plotosus lineatus	fish	no	no	no	yes	no	yes	no
Procambarus clarkii	invertebrates	no	no	no	yes	yes	yes	yes
Procambarus fallax f. virginalis	invertebrates	no	no	no	no	no	no	no
Procyon lotor	mammal	no	no	no	yes	yes	yes	yes
Prosopis juliflora	plant	yes	yes	yes	no	no	yes	no
Pseudorasbora parva	fish	no	no	no	no	yes	yes	no
Pueraria montana var. lobata = Pueraria lobata	plant	no	no	no	no	no	no	no
Salvinia molesta	plant	no	no	no	no	yes	yes	no
Sciurus carolinensis	mammal	no	no	no	yes	yes	yes	yes
Sciurus niger	mammal	no	no	no	yes	yes	yes	yes
Tamias sibiricus	mammal	no	no	no	yes	no	yes	no
Threskiornis aethiopicus	bird	no	no	no	yes	yes	yes	yes
Trachemys scripta	reptile	no	no	no	yes	yes	yes	yes
Triadica sebifera	plant	yes	yes	no	yes	no	yes	yes
Vespa velutina nigrithorax	invertebrates	no	no	no	no	yes	no	no

Annex B.2 Candidates for IAS of Union concern

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Ameiurus melas	fish	no	no	no	yes	yes	yes	yes
Ameiurus nebulosus	fish	no	no	no	yes	yes	yes	no
Axis axis	mammal	no	no	no	yes	no	yes	yes
Boccardia proboscidea	invertebrates	no	no	no	no	no	no	no
Callosciurus finlaysonii	mammal	no	no	no	no	no	yes	no
Castor canadensis	mammal	no	no	no	yes	no	yes	yes
Celastrus orbiculatus	plant	yes	yes	yes	yes	yes	yes	yes
Channa argus	fish	no	no	no	no	no	yes	no
Faxonius rusticus	invertebrates	no	no	no	no	no	yes	no
Fundulus heteroclitus	fish	no	no	no	yes	no	yes	no
Gambusia affinis	fish	no	no	no	yes	no	yes	yes
Gambusia holbrooki	fish	no	no	no	yes	no	yes	yes
Hakea sericea	plant	yes	yes	no	yes	no	yes	no
Koenigia polystachya	plant	yes	no	yes	no	no	no	no
Lagocephalus sceleratus	fish	no	no	no	no	no	yes	no
Lampropeltis getula	reptile	no	no	no	yes	yes	yes	yes
Limnoperna fortunei	invertebrates	no	no	no	no	no	no	no
Morone americana	fish	no	no	no	yes	no	yes	no
Perna viridis	invertebrates	no	no	no	no	no	yes	no
Phytolacca americana	plant	yes	yes	yes	yes	yes	yes	yes
Pistia stratiotes	plant	yes	yes	yes	yes	yes	yes	yes
Pterois miles	fish	no	no	no	no	no	yes	no

Pycnonotus cafer	bird	no	no	no	yes	yes	yes	yes
Rugulopteryx okamurae	plant	no	no	no	no	no	no	no
Schizoporella japonica	invertebrates	no	no	no	no	no	no	no
Solenopsis geminata	invertebrates	no	no	no	no	no	yes	no
Solenopsis invicta	invertebrates	no	no	no	yes	no	yes	yes
Solenopsis richteri	invertebrates	no	no	no	no	no	yes	no
Wasmannia auropunctata	invertebrates	no	no	no	no	no	yes	no
Xenopus laevis	amphibian	no	no	no	yes	yes	yes	no

Annex B.3 Local IAS – Danube Region

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Aedes albopictus	insect	no	no	no	yes	no	yes	yes
Alburnus albidus	fish	no	no	no	no	no	no	no
Ameiurus melas	fish	no	no	no	no	yes	yes	yes
Ameiurus nebulosus	fish	no	no	no	no	yes	yes	no
Babka gymnotrachelus	fish	no	no	no	no	no	no	no
Barbronia weberi	invertebrates	no	no	no	no	no	no	no
Borysthenia naticina	invertebrates	no	no	no	no	no	no	no
Branchiura sowerbyi	invertebrates	no	no	no	no	yes	no	no
Carassius gibelio	fish	no	no	no	no	yes	yes	no
Caspihalacarus hyrcanus	invertebrates	no	no	no	no	no	no	no
Caspiobdella fadejewi	invertebrates	no	no	no	no	no	no	no
Chaetogammarus ischnus	invertebrates	no	no	no	no	no	no	no
Chaetogammarus trichiatus	invertebrates	no	no	no	no	no	no	no
Chelicorophium curvispinum	invertebrates	no	no	no	no	yes	no	no
Chelicorophium robustum	invertebrates	no	no	no	no	yes	no	no
Chelicorophium sowinskyi	invertebrates	no	no	no	no	no	no	no
Corbicula fluminalis	invertebrates	no	no	no	no	yes	no	no
Corbicula fluminea	invertebrates	no	no	no	yes	yes	yes	yes
Cordylophora caspia	invertebrates	no	no	no	no	yes	no	no
Coregonus peled	fish	no	no	no	no	no	no	no
Crangonyx pseudogracilis	invertebrates	no	no	no	no	yes	no	no
Craspedacusta sowerbyi	invertebrates	no	no	no	no	no	no	no

Ctenopharyngodon idella	fish	no	no	no	no	yes	yes	no
Dendrocoelum romanodanubiale	invertebrates	no	no	no	no	no	no	no
Dikerogammarus bispinosus	invertebrates	no	no	no	no	no	no	no
Dikerogammarus haemobaphes	invertebrates	no	no	no	no	yes	no	no
Dikerogammarus villosus	invertebrates	no	no	no	no	yes	no	no
Dreissena polymorpha	invertebrates	no	no	no	yes	no	yes	yes
Dreissena rostriformis bugensis	invertebrates	no	no	no	no	no	no	no
Dugesia tigrina	invertebrates	no	no	no	no	no	no	no
Species	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Hypania invalida	invertebrates	no	no	no	no	no	no	no
Hypophthalmichthys molitrix	fish	no	no	no	no	no	yes	no
Hypophthalmichthys nobilis	fish	no	no	no	no	yes	no	no
Jaera sarsi	invertebrates	no	no	no	no	no	no	no
Katamysis warpachowskyi	invertebrates	no	no	no	no	no	no	no
Lepomis gibbosus	fish	no	no	no	yes	yes	yes	yes
Leucos basak	fish	no	no	no	no	no	no	no
Limnomysis benedeni	invertebrates	no	no	no	no	yes	no	no
Manayunkia caspica	invertebrates	no	no	no	no	no	no	no
Melanoides tuberculatus	invertebrates	no	no	no	no	yes	yes	no
Micropterus salmoides	fish	no	no	no	no	no	yes	yes
Neogobius fluviatilis	fish	no	no	no	no	yes	no	no
Neogobius melanostomus	fish	no	no	no	yes	yes	yes	no
Niphargus hrabei	invertebrates	no	no	no	no	no	no	no
Obesogammarus obesus	invertebrates	no	no	no	no	no	no	no

Oncorhynchus mykiss	fish	no	no	no	no	yes	yes	yes
Orchestia cavimana	invertebrates	no	no	no	no	yes	no	no
Pachychilon macedonicus	fish	no	no	no	no	no	no	no
Paramysis lacustris	invertebrates	no	no	no	no	no	no	no
Pectinatella magnifica	invertebrates	no	no	no	no	yes	yes	no
Physella acuta	invertebrates	no	no	no	no	yes	yes	no
Piscicola haranti	invertebrates	no	no	no	no	yes	no	no
Polyodon spathula	fish	no	no	no	no	no	no	no
Ponticola kessleri	fish	no	no	no	no	no	no	no
Pontogammarus robustoides	invertebrates	no	no	no	no	no	no	no
Potamopyrgus antipodarum	invertebrates	no	no	no	no	yes	yes	no
Potamothrix moldaviensis	invertebrates	no	no	no	no	yes	no	no
Proasellus coxalis	invertebrates	no	no	no	no	yes	no	no
Salmo letnica	fish	no	no	no	no	no	no	no
Salvelinus fontinalis	fish	no	no	no	no	yes	yes	no
Scardinius graecus	fish	no	no	no	no	no	no	no
Sinanodonta woodiana	invertebrates	no	no	no	no	yes	yes	no
Synurella ambulans	invertebrates	no	no	no	no	no	no	no
Urnatella gracilis	invertebrates	no	no	no	no	no	no	no

Annex B.4 Local IAS - Sava Ties

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Acer negundo	plant	no	yes	yes	yes	yes	yes	yes
Ambrosia artemisiifolia	plant	no	yes	yes	yes	yes	yes	yes
Amorpha fruticosa	plant	no	yes	yes	yes	yes	yes	yes
Bidens frondosa	plant	yes	yes	yes	yes	yes	yes	yes
Buddleja davidii	plant	yes	yes	yes	yes	yes	yes	yes
Conyza canadensis	plant	yes	yes	yes	yes	yes	yes	yes
Echinocystis lobata	plant	yes	yes	yes	yes	no	yes	yes
Fraxinus americana	plant	yes	yes	yes	yes	yes	yes	yes
Fraxinus pennsylvanica	plant	yes	yes	yes	yes	yes	yes	yes
Gleditsia triacanthos	plant	yes	yes	yes	yes	yes	yes	yes
Oenothera biennis	plant	yes	no	yes	yes	yes	yes	yes
Panicum barbipulvinatum	plant	yes	no	no	no	no	no	no
Paulownia tomentosa	plant	yes	yes	yes	yes	yes	no	yes
Physocarpus opulifolius	plant	yes	yes	yes	yes	yes	yes	yes
Phytolacca americana	pant	yes	yes	yes	yes	yes	yes	yes
Reynoutria japonica	plant	yes	yes	yes	yes	no	yes	no
Reynoutria sachalinensis	plant	yes	yes	yes	yes	no	yes	no
Reynoutria × bohemica	plant	no	no	no	no	no	no	no
Robinia pseudoacacia	plant	yes	yes	yes	yes	yes	yes	yes
Solidago canadensis	plant	yes	yes	yes	yes	yes	yes	yes
Solidago gigantea	plant	yes	yes	yes	yes	yes	yes	no
Spiraea japonica	plant	yes	yes	yes	yes	yes	yes	no

Symphyotrichum lanceolatum	plant	yes	yes	yes	yes	no	yes	no
Symphyotrichum novi-belgii	plant	yes	yes	yes	yes	no	yes	no
Vitis riparia	plant	yes	yes	no	yes	no	yes	yes
Xanthium strumarium	plant	yes	yes	yes	yes	yes	yes	yes

Annex B.5 Local IAS – Iberian Peninsula

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Acer negundo	plant	yes	yes	yes	yes	yes	yes	yes
Acer pseudoplatanus	plant	yes	yes	yes	yes	yes	yes	yes
Achillea filipendulina	plant	yes	yes	no	no	yes	yes	no
Aesculus hippocastanum	plant	yes	yes	yes	yes	yes	yes	yes
Agave americana	plant	yes	yes	yes	yes	no	yes	yes
Aix galericulata	bird	no	no	no	yes	yes	yes	yes
Alburnus alburnus	fish	no	no	no	no	yes	yes	no
Amandava amandava	bird	no	no	no	no	no	yes	no
Amaranthus albus	plant	yes	yes	yes	no	yes	yes	no
Amaranthus blitoides	plant	yes	yes	yes	no	yes	yes	no
Amaranthus hybridus	plant	yes	yes	yes	no	yes	yes	no
Amaranthus muricatus	plant	no	yes	no	no	yes	no	no
Amaranthus powellii	plant	no	yes	no	no	no	yes	no
Amaranthus retroflexus	plant	yes	yes	yes	yes	yes	yes	no
Amaranthus viridis	plant	yes	yes	yes	no	yes	yes	no
Ammotragus lervia	mammal	no	no	no	no	no	yes	no
Artemia franciscana	invertebrates	no	no	no	no	no	no	no
Artemisia verlotiorum	plant	no	yes	no	no	yes	yes	no
Arundo donax	plant	yes	yes	yes	yes	yes	yes	yes
Aster squamatus / Symphyotrichum squamatum	plant	yes	yes	no	no	no	yes	no
Australoheros facetus	fish	no	no	no	no	no	no	no
Azolla filiculoides	plant	yes	yes	yes	yes	yes	yes	no

Bactrocera oleae	invertebrates	no	no	no	no	no	no	no
Bemisia tabaci	invertebrates	no	no	no	no	no	no	no
Bidens aurea	plant	yes	yes	no	no	no	yes	no
Bidens subalternans	plant	no	yes	no	no	no	no	no
Branta canadensis	bird	no	no	no	yes	yes	yes	yes
Bromus willdenowii (Bromus catharticus)	plant	yes	yes	yes	yes	no	yes	no
Cairina moschata	bird	no	no	no	yes	yes	yes	yes
Callinectes sapidus	invertebrates	no	no	no	yes	yes	yes	yes
Carassius auratus	fish	no	no	no	yes	yes	yes	yes
Cherax destructor	invertebrates	no	no	no	no	no	yes	no
Chrysemys picta	reptile	no	no	no	yes	yes	yes	yes
Conyza bonariensis (Erigeron bonariensis)	plant	yes	yes	yes	yes	yes	yes	no
Conyza canadensis / Erigeron canadensis / Erigeron pusillus	plant	yes	yes	no	yes	yes	yes	yes
Conyza sumatrensis (Erigeron sumatrensis)	plant	yes	yes	no	yes	yes	yes	no
Crassostrea gigas	invertebrates	no	no	no	no	yes	yes	no
Crassula helmsii	plant	no	yes	yes	no	yes	yes	no
Crepidula fornicata	invertebrates	no	no	no	yes	yes	yes	yes
Cupressus arizonica	plant	yes	yes	no	no	no	yes	no
Cylindropuntia imbricata	plant	no	yes	yes	yes	no	yes	yes
Cylindropuntia rosea	plant	no	yes	no	no	no	no	yes
Cyprinus carpio	fish	no	no	no	yes	yes	yes	yes
Dama dama	mammal	no	no	no	yes	yes	yes	yes
Datura innoxia	plant	yes	yes	no	no	yes	yes	no
Datura stramonium	plant	yes	yes	yes	yes	yes	yes	yes

Echinochloa hispidula	plant	yes	no	no	no	no	no	no
Egeria densa	plant	yes	yes	yes	no	yes	yes	no
Elaeagnus angustifolia	plant	yes	yes	yes	yes	yes	yes	yes
Elodea canadensis	plant	yes	yes	yes	yes	yes	yes	no
Esox lucius	fish	no	no	no	yes	yes	yes	no
Estrilda astrild	bird	no	no	no	yes	yes	yes	yes
Eucalyptus camaldulensis	plant	no	yes	yes	no	no	yes	no
Eucalyptus globulus	plant	no	yes	yes	yes	no	yes	yes
Eurytoma amygdali	invertebrates	no	no	no	no	no	no	no
Fallopia baldschuanica	plant	yes	yes	yes	no	yes	yes	no
Ficopomatus enigmaticus	invertebrates	no	no	no	no	yes	yes	no
Frankliniella occidentalis	invertebrates	no	no	no	no	no	no	no
Fundulus heteroclitus	fish	no	no	no	yes	no	yes	no
Gambusia holbrooki	fish	no	no	no	yes	no	yes	yes
Gleditsia triacanthos	plant	yes	yes	yes	yes	yes	yes	yes
Gobio lozanoi	fish	no	no	no	no	no	yes	no
Helianthus tuberosus	plant	yes	yes	yes	yes	yes	yes	yes
Ictalurus punctatus	fish	no	no	no	yes	no	yes	yes
Ipomoea purpurea	plant	yes	yes	yes	yes	yes	yes	yes
Isatis tinctoria	plant	yes	yes	yes	no	yes	yes	no
Lernaea cyprinacea	fish	no	no	no	no	no	no	no
Marsupenaeus japonicus	invertebrates	no	no	no	no	no	no	no
Mirabilis jalapa	plant	yes	yes	yes	yes	yes	yes	yes
Misgurnus anguillicaudatus	fish	no	no	no	no	yes	yes	no

Mnemiopsis leidyi	invertebrates	no	no	no	no	yes	yes	no
Myiopsitta monachus	bird	no	no	no	yes	yes	yes	yes
Mytilopsis leucophaeata	invertebrates	no	no	no	no	yes	no	yes
Neovison vison	mammal	no	no	no	yes	yes	yes	yes
Nicotiana glauca	plant	yes	yes	yes	yes	yes	yes	yes
Nymphaea mexicana	plant	yes	yes	no	no	no	yes	no
Oenothera biennis	plant	yes	yes	no	yes	yes	yes	yes
Oenothera glazioviana	plant	yes	yes	no	yes	yes	yes	no
Ophiostoma ulmi / Ophiostoma novo-ulmi / Ceratocystis ulmi	plant	no	no	no	no	yes	no	no
Opuntia dillenii	plant	yes	yes	no	no	no	yes	no
Opuntia ficus-indica	plant	yes	yes	yes	yes	no	yes	yes
Opuntia maxima	plant	no	yes	no	no	no	no	no
Ovis musimon / Ovis orientalis	mammal	no	no	no	no	yes	no	no
Oxalis pes-caprae	plant	yes	yes	yes	yes	no	yes	yes
Paspalum dilatatum	plant	yes	yes	yes	yes	yes	yes	yes
Paspalum paspalodes	plant	yes	no	no	no	no	yes	no
Paysandisia archon	invertebrates	no	no	no	no	yes	yes	no
Perca fluviatilis	fish	no	no	no	yes	yes	yes	no
Phasianus colchicus	bird	no	no	no	yes	yes	yes	yes
Phyla nodiflora / Lippia filiformis / Lippia nodiflora	plant	yes	yes	yes	yes	yes	yes	yes
Phytophthora cinnamomi	plant	no	no	no	no	no	no	no
Pomacea spp.	invertebrates	no	no	no	no	no	yes	no
Populus x canadensis	plant	yes	yes	no	no	yes	no	no
Potamopyrgus antipodarum	invertebrates	no	no	no	no	yes	yes	no

Prays oleae	invertebrates	no	no	no	no	yes	no	no
Psittacula krameri	bird	no	no	no	yes	yes	yes	yes
Rapana venosa	invertebrates	no	no	no	no	yes	yes	no
Rhinella marina	amphibian	no	no	no	yes	no	yes	yes
Rhithropanopeus harrisii	invertebrates	no	no	no	no	yes	yes	no
Rhynchophorus ferrugineus	invertebrates	no	no	no	yes	no	yes	no
Robinia pseudoacacia	plant	no	yes	yes	yes	yes	yes	yes
Ruditapes philippinarum	invertebrates	no	no	no	yes	yes	yes	no
Rutilus rutilus	fish	no	no	no	no	yes	yes	no
Salix babylonica	plant	yes	yes	yes	yes	yes	yes	yes
Sander lucioperca	fish	no	no	no	no	yes	yes	no
Scardinius erythrophthalmus	fish	no	no	no	yes	yes	yes	no
Senecio inaequidens	plant	yes	yes	yes	yes	yes	yes	no
Silurus glanis	fish	no	no	no	no	yes	yes	no
Sophora japonica	plant	yes	yes	yes	no	yes	yes	no
Sorghum halepense	plant	yes	yes	yes	yes	yes	yes	yes
Tomicus destruens	invertebrates	no	no	no	no	no	no	no
Tuta absoluta	invertebrates	no	no	no	no	yes	no	no
Ulmus pumila	plant	yes	yes	yes	yes	no	yes	no
Vinca difformis	plant	yes	yes	no	no	no	yes	no
Xanthium spinosum	plant	yes	yes	yes	yes	yes	yes	no
Xanthium strumarium	plant	yes	yes	yes	yes	yes	yes	yes
Xenopus laevis	amphibian	no	no	no	yes	yes	yes	no
Zygophyllum fabago	plant	yes	yes	no	no	no	yes	no

Annex B.6 Local IAS – Malta

Species scientific name	Category	PI	PN	FI	iNC	NIA	iNA	Mic
Acacia cyclops	plant	yes	no	no	yes	no	yes	no
Achatina achatina	invertebrates	no	no	no	no	no	no	no
Agave americana	plant	yes	yes	yes	yes	no	yes	yes
Agave sisalana	plant	yes	no	yes	no	no	yes	no
Anredera cordifolia	plant	yes	yes	yes	no	no	yes	no
Bambusa vulgaris	plant	yes	no	yes	no	no	yes	no
Cardiospermum halicacabum	plant	yes	yes	yes	yes	no	yes	yes
Carpobrotus acinaciformis	plant	no	no	yes	no	no	yes	no
Carpobrotus edulis	plant	yes	yes	yes	yes	yes	yes	yes
Drosanthemum hispidum	plant	yes	no	yes	no	yes	yes	no
Kalanchoe daigremontiana	plant	yes	no	yes	no	no	yes	no
Kalanchoe delagoensis	plant	yes	yes	no	yes	no	yes	no
Leucaena leucocephala	plant	yes	yes	yes	yes	no	yes	yes
Lissachatina fulica	invertebrates	no	no	no	yes	no	yes	yes
Malephora crocea	plant	yes	no	yes	no	no	yes	no
Mesembryanthemum cordifolium	plant	yes	no	yes	yes	no	yes	no
Mesembryanthemum lancifolium	plant	no	no	no	no	no	no	no
Mirabilis jalapa	plant	yes	yes	yes	yes	yes	yes	yes
Nicotiana glauca	plant	yes	yes	yes	yes	yes	yes	yes
Otala punctata	invertebrates	no	no	no	no	no	yes	no
Pelophylax bedriagae	amphibian	no	no	no	no	no	yes	no
Pennisetum villosum	plant	yes	no	yes	no	yes	yes	no

Ricinus communis	plant	yes	yes	yes	no	yes	yes	yes
Tropaeolum majus	plant	yes	yes	yes	no	yes	yes	yes
Vachellia karroo (Acacia karroo)	plant	no	no	yes	no	no	yes	no
Yucca gloriosa	plant	no	no	yes	no	yes	yes	no

Annex C: Selection of Datasets

During our investigations we came across different datasets containing images that are labeled with the respective species depicted in the images. The providers were gathered and are listed here with the name of the dataset, the name of the provider, a link to the dataset and the number of species and images included in the dataset.

Name	Provider	Link	Number of species / images
iNaturalist Challenge at FGVC 2017	iNaturalist	https://www.kaggle.com/c/inaturalist-challenge-at-fgvc-2017/data	5,089 / 675,000
iNaturalist Competition datasets (values from 2021)	iNaturalist	https://github.com/visipedia/inat_comp	10,000 / 2,686,843
iNaturalist GBIF DarwinCore Archive	iNaturalist / GBIF	http://www.inaturalist.org/observations/gbif-observations-dwca.zip	n.a.
GBIF (different datasets, e.g. from iNaturalist)	GBIF	https://www.gbif.org/dataset/50c9509d-22c7-4a22-a47d-8c48425ef4a7	64,000 datasets
EMammel	EMammel	https://emammal.si.edu/	110 projects
Caltech-UCSD Birds 200	Caltech	http://www.vision.caltech.edu/visipedia/CUB-200.html	200 / 6033
Flower Dataset	Oxford University	https://www.robots.ox.ac.uk/~vgg/data/flowers/	102 / 40-258 per class
NABirds	Cornell Lab of Ornithology	https://dl.allaboutbirds.org/nabirds	400 / 48,000
Project Natick Underwater Video	Microsoft	https://github.com/Microsoft/Project_Natick_Analysis/releases/tag/annotated_data	n.a. / 1,000
Imagery data of paired shrub-open microsites	Noble et al.	http://gigadb.org/dataset/100191	n.a. / 100,000
Mammalweb	Mammalweb	https://osf.io/znm6k/	n.a.
FishnetAI	The Nature Conservancy	https://www.fishnet.ai/home	33 / 86,029
DeepFish	Saleh et al.	https://alzayats.github.io/DeepFish/	20 habitats / 40,000
Snapshot Serengeti	Swanson et al.	https://datadryad.org/stash/dataset/doi:10.5061/dryad.5pt92	40 / 10.8 million
Labeled Information Library of Alexandria: Biology and Conservation	LILA BC	https://lila.science/	n.a. / > 10 million

North America Camera Trap Images (NACTI)	LILA https://lila.science/	https://lila.science/datasets/nacti	28 / 3.7 million
Caltech Camera Traps	Caltech	https://lila.science/datasets/caltech-camera-traps	21 / 243,100
Seaview Survey Photo-quadrat and Image Classification Dataset	XL Catlin Seaview Survey, University of Queensland	https://espace.library.uq.edu.au/view/UQ:734799	n.a. / 1,1 million
Training images PantCLEF for 2017	ImageClef	http://otmedia.lirmm.fr/LifeCLEF/	10,000 / n.a.
Labeled Fish in the wild	NOAA Fisheries	https://swfscdata.nmfs.noaa.gov/labeled-fishes-in-the-wild/	n.a. / 929

Annex D: Image Providers

Below all providers of images used for testing are listed. Most of the images were downloaded via GBIF and the respective source was extracted from there if possible. Additionally, links to the databases are listed, where images of species can be searched by mostly entering the scientific name. When no link was found, the GBIF link was used.

Provider	Link
Arizona State University Biocollections. Laura Steger and Rick Overson Vertebrate Observations	https://csvcoll.org/portal/index.php
Atlas of Living Australia - BioCollect	https://www.ala.org.au/biocollect/
Atlas of Living Australia - Weeds in Australia.	https://weeds.org.au
Biodiversity4all Research-Grade Observations	https://www.biodiversity4all.org/
Biological Records Centre. Mammal records for Europe via the iMammalia app.	https://mammalnet.com/
Botanic Garden and Botanical Museum Berlin Observations	https://www.ggbn.org/bobo/sptool/sptool.php
CFE - Centre for Functional Ecology, Department of Life Sciences, University of Coimbra. Sightings Map of Invasive Plants in Portugal	https://invasoras.pt/en/invasive-species-in-portugal
Estonian Naturalists' Society	https://doi.org/10.15468/bmk3ab
India Biodiversity Portal	https://indiabiodiversity.org/
Lomonosov Moscow State University. Collections of Bioclass, school #179, Moscow	https://doi.org/10.15468/4f0bmt
Miljøstyrelsen / The Danish Environmental Protection Agency. Species recordings from the Danish National portal Arter.dk	https://arter.dk/
Missouri Botanical Garden. Tropicos Specimen Data	https://www.tropicos.org/
National Museum of Natural History, Smithsonian Institution. NMNH Extant Specimen Records (USNM, US)	https://collections.nmnh.si.edu/search/
NatureMapr. Atlas of Life in the Coastal Wilderness	https://atlasoflife.naturemapr.org/home
NatureMapr. Canberra Nature Map	https://canberra.naturemapr.org/
Questagame. Earth Guardians Weekly Feed	https://biocache.ala.org.au/search
SANBI - South African Biodiversity Institute	http://pza.sanbi.org/

Senckenberg. African Plants - a photo guide	http://www.africanplants.senckenberg.de/
The Norwegian Biodiversity Information Centre (NBIC). Norwegian Species Observation Service	https://www.artsobservasjoner.no/
The South African Institute for Aquatic Biodiversity. Occurrence records of southern African aquatic biodiversity	https://doi.org/10.15468/pv7vds
University of Colorado Museum of Natural History. UCM Vertebrate Observations Collection (Arctos)	https://arctos.database.museum/
Vanderbilt University - Bioimages	http://bioimages.vanderbilt.edu/

KJ-01-22-039-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/97305

ISBN 978-92-76-46721-2