AGILE: GIScience Series, 3, 10, 2022. https://doi.org/10.5194/agile-giss-3-10-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Benchmarking Invasive Alien Species Image Recognition Models for a Citizen Science Based Spatial Distribution Monitoring

Tom Niers [1,*], Jan Stenkamp [1,*], Nick Pascal Jakuschona [1,*], Thomas Bartoschek [1], and Sven Schade [2]

[1]Institute for Geoinformatics, University of Münster, Münster, Germany
[2]European Commission, Joint Research Centre (JRC), Ispra, Italy
[*]These authors contributed equally to this work.

Correspondence: Tom Niers (tom.niers@uni-muenster.de), Jan Stenkamp (jan.stenkamp@uni-muenster.de), Nick Jakuschona (nick.jakuschona@uni-muenster.de)

Abstract. Recent developments in image recognition technology including artificial intelligence and machine learning led to an intensified research in computer vision models. This progress also allows advances for the collection of spatio-temporal data on Invasive Alien Species (IAS), in order to understand their geographical distribution and impact on the biodiversity loss. Citizen Science (CS) approaches already show successful solutions how the public can be involved in collecting spatio-temporal data on IAS, e.g. by using mobile applications for monitoring. Our work analyzes recently developed image-based species recognition models suitable for the monitoring of IAS in CS applications. We demonstrate how computer vision models can be benchmarked for such a use case and how their accuracy can be evaluated by testing them with IAS of European Union concern. We found out that available models have different strengths. Depending on which criteria (e.g. high species coverage, costs, maintenance, high accuracies) are considered as most important, it needs to be decided individually which model fits best. Using only one model alone may not necessarily be the best solution, thus combining multiple models or developing a new custom model can be desirable. Generally, cooperation with the model providers can be advantageous.

Keywords. Image Recognition, Invasive Alien Species, Citizen Science, Machine Learning, Species Distribution

1 Introduction

Invasive Alien Species (IAS) are one of the main drivers of biodiversity loss globally. Thus, there is a need to collect spatio-temporal data on the occurrence of IAS to be able to drive conclusions concerning their spatial distribution and to tackle their negative effect on the equilibrium of the local ecosystems. The data required for consistent monitoring of IAS can only be collected with great effort, as it has to be available in large quantities and of corresponding quality. The involvement of volunteers to collect spatial and thematic data to produce Volunteered Geographic Information has been proven to be a successful approach (Goodchild, 2007). Adding a research question on the potential impact or causes of changes in biodiversity to the partly trivial crowdsourcing approach leads to Citizen Science (CS). CS can have a crucial impact on moving the collection of spatio-temporal data about IAS forward. There are already several systems (e.g. Invasive Alien Species Europe App EASIN (2022)) that allow citizens to record their findings of invasive species and thus participate in the monitoring of IAS.

Like the collection of the data itself, its validation and analysis is complex and time-consuming. At this point, the usage of Artificial Intelligence (AI) can significantly facilitate the monitoring of IAS. Both, in terms of recognizing IAS and understanding their spatio-temporal distribution. This work focuses on the automated image-based species recognition of IAS of European Union concern. We investigate several AI-based species recognition models, and analyze their usability for an integration in a CS application enabling user conducted monitoring of IAS. The contribution complements a detailed technical report that was already published (Jakuschona et al., 2022a). Here, we present the scientific perspective of our work, synthesize our main findings, and add more recent results. To do so, we respectively list related work, describe our approach of analyzing the models, share our findings, and address how our analysis might be taken forward.

2 Related Work

Citizen Science can have a crucial effect on supporting the evaluation of species distributions. As Feldman et al. (2021) found, the number of papers utilizing CS approaches for species distribution models (SDMs) doubled compared to the number of overall SDM papers in the 2010s. Not only does CS contribute to the understanding of species distributions in general, but also in particular of IAS (Crall et al., 2015). This is achieved by filling observation gaps of underrepresented species or geographical regions, which can not be covered by experts. Many CS initiatives collecting IAS observations make use of information and communication technologies to improve the data collection. Johnson et al. (2020) analyzed 26 of such initiatives that led to scientific discoveries. Most of them are using GPS to record the location of the observations and allow citizens to take or upload photos within the provided application. Some of these discoveries included the modeling of the spatial distribution of the invasive species or the clarification of spatio-temporal patterns of the invasion.

As an information technology in CS applications, the use of AI can be beneficial, e.g. to help volunteers with identification tools determining unknown species. These could otherwise be underrepresented in the data collection due to the citizens biases (Schermer and Hogeweg, 2018). Beside the data collection, also the engagement and the data validation can be promoted by machine learning (Lotfian et al., 2021). An example was stated by Lotfian et al. (2019) in a case study, where image recognition is used to review the species classification of citizens. Sightings are first filtered by their location and the period the species are usually seen. If a species is common for the region and the time of the day and year, the observation is passed to an image recognition model. If the model determines the same species as the citizen, the observation is automatically validated, if not it has to be resubmitted by the citizen and if it still not passes the auto-filtering it is reviewed by an expert. This process helps to reduce the workload for the experts in the validation process but also encourages the citizens to continue their fieldwork by giving real-time feedback on their observations.

Further, use cases of image recognition models within CS applications are especially in the field of camera trap observations, where a lot of images are collected without any human input, that have to be classified afterwards. For example Choiński et al. (2021), Tabak et al. (2019), Norouzzadeh et al. (2021) and Norouzzadeh et al. (2018) trained neural networks to classify camera trap images of mammals. The first for images from Poland, the second from North America and the last two from Africa. The images from Africa were obtained from the Snapshot Serengeti dataset. This is part of a long-term CS project with 225 cameras deployed across the Serengeti National Park in Tanzania. The images are published on a website and can there be annotated by volunteers and afterwards serve as a source for machine learning or computer vision research (Swanson et al., 2015). Beside the models for recognition of mammal species, computer vision techniques can also be used to detect plant species (Goëau et al., 2021; Wäldchen et al., 2018) or to recognize animal individuals (Miele et al., 2021).

Such machine learning models identifying species in CS images can be enhanced by incorporating metadata in the classification process, like the location or the weather (Terry et al., 2020). Also participants in the LifeCLEF challenges in 2018 (Joly et al., 2018) and 2020 (Joly et al., 2020) improved the species identification process by including location information. They were used to reduce the list of species candidates to be considered and thus sped up the classification and increased the accuracy.

Another species classification challenge, providing a respective annotated dataset, is the iNaturalist competition (Van Horn et al., 2018), which was conducted the fourth time in 2021. The datasets of these challenges contain images collected by citizens via the iNaturalist platform. In the competition, citizens are encouraged to create an image classification model with these datasets. The organizers also benchmarked different representation learning methods for two natural world image collections (Van Horn et al., 2021).

In addition to the benchmarking of machine learning methods, there have also been reviews of image recognition approaches in ecology. Wäldchen and Mäder (2018) stated that, although there are various attempts of creating species identification models in research, just a few useful applications are available, while Weinstein (2018) reviewed 187 computer vision applications in animal ecology. Many of the considered applications are either smartphone applications or results of research on datasets with a small number of species. Solely smartphone applications were evaluated by Jones (2020), who tested automatic plant identification apps and their performance on identifying plants common in Britain.

3 Research Methodology

In order to assist citizen scientists in monitoring IAS, a promising approach is to detect the species by an image-based recognition. Thereby the recognition can be performed in an automated manner by using ML methods. Computer vision models are trained using annotated images of specific species in order to subsequently predict the species for a given image. To find appropriate models fitting our use case, an intensive literature search was carried out. Publications and outcomes relevant to this topic were investigated, a web search on GitHub and in corresponding blogs was conducted, and contacts to experts in this domain were used to directly obtain information. Some of the keywords relevant for the respective search queries were "image recognition", "species", "identification", "machine learning", "image based", "automated" and "computer vision". The articles found with this keywords were investigated in more detail and the attributes of described or referenced models were evaluated. Generally, we considered image-based species recognition models that can be integrated either directly as a model or as an API in a system which supports citizens in monitoring IAS.

To be relevant for our subsequent research, the models had to include at least 2,000 species and be applicable in Europe, covering the relevant species. In this chapter we want to describe which criteria we used for our further analysis of the models, which images we used to test and how we calculated the accuracy of the predictions the models returned.

3.1 Assessment Criteria for Species Recognition Models

For our investigations, we defined certain criteria to make the models assessable and comparable (see Table 1).

One set of the criteria describes the coverage of the model. This includes the species that can be recognized by the model, both the number of overall species and specifically IAS, the class of organisms and the geographic region which the model is trained on. A second set addresses the availability of the model. On the one hand, whether the model can be further trained by additional images and one can transparently learn how the model works and with which species it was trained. On the other hand, in which format the model is available, e.g. as download or API, and whether there are corresponding costs. Third, the accuracy is a crucial part of our criteria, i.e. in how many cases the model correctly recognizes the species for a test dataset of images. Besides we defined two further technical criteria. These address details concerning the update cycle of the model and the specification of the requests which can be processed by the model.

In order to obtain all necessary information about the models with respect to the criteria, we used various sources. First, we used the information provided directly by the provider on their website. We have linked this source for each model in Table 3. Second, we searched for additional sources such as further documentation or blog posts, which we referenced when the information was mentioned in the text. Third, if still not sufficient information was available, we obtained it through direct contact with the model providers.

Table 1. Criteria

Criteria	Description
Species	Overall 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).
Class of Organism	Describing, if a model is specifically trained for a certain kind of organism (e.g. plant, mammal, bird, fish).
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 other ways.
Cost	Costs for usage (e.g. per request, year, one-time download).
Accuracy	Describing the accuracy for Top-1 and Top-5 suggestions of a model, when it is applied to a test dataset of images.
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, money-wise) and if a score for the predictions is given.

3.2 Approach to Evaluate Model Accuracy

In order to determine how accurately the models perform, we have developed a specific test strategy. According to our use case (see Section 1), our test focuses on the IAS of Union concern specified by the European Union Regulation 1143/2014 (Regulation, 2014), and candidate species that probably will be added to the list in the future. We received the information about IAS candidates directly by the European Commission's Joint Research Centre (JRC), especially by the team supporting the European Alien Species Information Network (EASIN). In total, the number of species considered for testing was 96.

3.2.1 Images Used for Testing

For each species, the test dataset is composed of six images in total. One half of these are so called golden standard images and the other half are user observation images. The golden standard images are images on which the species is clearly represented and recognizable for experts. In contrast, user observations are images taken by actual users of species observation apps, which are usually of lower quality and do not always fully represent the species.

The golden standard images were largely taken from a dataset provided to us by the JRC. However, for some species there were only two images available in this dataset and for the candidate species no images at all. In these cases, images from the Commonwealth Agricultural Bureaux International (CABI, 2022) database were used. The images available in this database are typically of high quality, as they are imported from scientific papers. In case there were no images for a particular species available in the CABI database, images from the Global Biodiversity Information Facility (GBIF, 2022) were used. Thereby, images from the category "preserved specimens" were prioritized. In addition the images had to be of corresponding quality to count as golden standard.

Concerning the observation images, we used a dataset, likewise provided by the JRC. This contains validated images that were uploaded by users to the "Invasive Alien Species Europe" Smartphone Application (EASIN, 2022). In case there were not enough images available for a species within this dataset, images provided by GBIF were used and if available prioritized from the category "human observations". For some species even then not enough images were available, so that other trustworthy sources were used (see Section 3.3). Selecting the observation images, attention was paid that species are still recognizable by a human at all, and that the images differ from each other, e.g. showing different parts of the species.

Table 2. Exemplary Classification of IAS Images

Image	Prediction	Probability in %
	**Myocastor coypus**	80.50
	Ondatra zibethicus	7.51
	Castor canadensis	6.19
	Lontra canadensis	0.96
	Microtus pennsylvanicus	0.72
	Asclepias speciosa	49.80
	Asclepias syriaca	41.40
	Asclepias eriocarpa	3.09
	Wyethia mollis	1.36
	Calotropis procera	0.50
	Terrapene ornata	17.53
	Caiman crocodilus	13.16
	Apalone spinifera	8.95
	Stigmochelys pardalis	7.37
	Lithobates septentrionalis	5.59

The correct prediction is written in bold. The accuracies calculated with the examples in the table are: Top-1 accuracy: 33.33 % (1 out of 3 correct in the first prediction), Top-5 accuracy: 66.67 % (2 out of 3 correct in the first five predictions). The species depicted in the third example is *Lithobates catesbeianus*.

3.2.2 Calculation of Model Accuracy

In order to be able to draw conclusions about the accuracy of the models, all images included in the test dataset were classified by the models.

Typically, for a given image, the models return species as a prediction, including the probability that the species depicted on the image actually matches the prediction. Thereby, it differs by model how many predictions are returned. To determine the accuracy of the models, we decided to calculate the Top-1 and Top-5 accuracy. If the first prediction of a model is correct, it is a Top-1 prediction and if one of the first five predictions of a model is correct, it is within the Top-5 predictions. The calculation of the Top-5 accuracies can be interesting to decide if the model is helpful to show citizens five predictions from which they can indicate the observed one. Thus, the percentage given by the Top-1 or Top-5 accuracy indicates for how many of the images contained in the test dataset the prediction was correct or within the Top-5, respectively. An example for this calculation can be seen in Table 2.

The accuracies, both Top-1 and Top-5 were first calculated for each model independently. Whereby, the test dataset was truncated for each model to the species the model was trained on, since otherwise the model would not be able to detect the correct species. This approach leads to the fact that a different number of species and therefore different images were used to calculate the accuracies of the different models. To increase comparability, the accuracies were calculated a second time with an identical dataset for each model. For this purpose, the test dataset was trimmed to the species covered by all models. However, we differentiated between a dataset with species which are covered by all models, one with species which are covered by all hybrid models (those that recognize both, plants and animals) and one with species covered by all plant models. Thus, within the three categories all/hybrid/plant, all models have exactly the same input dataset, which leads to a better comparability.

3.3 Data and Software Availability

Research code supporting this document is accessible on GitHub. All repositories are available in the GitHub organization EibSReM (Jakuschona et al., 2022b). The three repositories we created are additionally accessible via Zenodo (Niers et al. (2022), Jakuschona et al. (2022c), Stenkamp et al. (2022)). Each of the repositories contains a README file with instructions on how to execute the script. Thereby, the scripts are either executable as python code or as Jupyter Notebook (in case of the API requests). All three repositories can be accessed under the European Union Public Licence V. 1.1. A detailed description of all work steps is available as JRC report (Jakuschona et al., 2022a). In addition to a description of the work steps and further details, the appendix of the report contains all values for the calculation of the accuracies as well as tables which summarize the model coverage of the species. For the image data used, there is a table in Appendix C of the report that lists all sources. Unfortunately, the images cannot be made available to the public for copyright reasons, but the images can be provided by the authors upon request.

4 Results

As a result of the literature and web research, we indicated seven models as applicable for detecting IAS of Union concern, since they cover enough species, have an appropriate geographic focus and can be integrated in another application. These seven models we tested can be inspected in Table 3.

The models can be separated in two groups. The first group includes models that can only identify plants (Flora Incognita, Pl@ntNet-API and Plant.id API), the second comprises models able to identify plants and animals (iNat2021, iNaturalist API, Microsoft and NIA) which we call hybrid models. For these seven models, the assessment criteria described in Section 3.1 were inspected in detail. In this chapter, we are going to describe the results for each model regarding the criteria. Starting with the species coverage, followed by the accessibility and usability and concluding with the model accuracies.

4.1 Coverage of Species

For the species coverage, we inspected the overall amount of species covered, the coverage of IAS of Union concern and the geographic region covered by each model. An overview of the species coverage can be found in Table 4.

Starting with the geographic region, the iNaturalist, Pl@nt-Net and Plant.id API cover species from the whole world. The same applies for the iNat2021 and Microsoft model. Also, Flora Incognita is covering species from the whole world, whereby they focus on species from Europe (Mäder et al., 2021). NIA concentrates on species from Belgium and the Netherlands (Observation.org, 2022).

The iNaturalist API is able to identify 38,000 species (Shepard, 2021), the most species among all models. Followed by Pl@ntNet-API, which is able to identify 29,615 plant species (Pl@ntNet, 2022). In terms of the other plant models, Plant.id covers 12,128 species and Flora Incoginta 4,803 (Mäder et al., 2021). NIA covers 22,302 species of plants and animals, iNat2021 10,000 (Van Horn et al., 2021) and Microsoft 5,266 (Microsoft, 2022). Out of the 96 IAS of Union concern, including the candidates, the iNaturalist API covers with 82 (iNaturalist, 2022) the most species. The other hybrid models cover fewer IAS, as iNat2021 covers 59 species, NIA 49 species and Microsoft 45 species. The plant identification models cover a similar amount of the 42 IAS that are plants. Plant.id is able to identify 31 plants, Pl@ntNet-API 30 and Flora Incognita 29.

Further, it is interesting that twelve species are covered by all models and five species are not covered by any model. A closer look at the plant models reveals that 21 species of the 42 plants are covered by all models and just six by none. Whereas the iNaturalist API covers with 36 IAS plants as much as all plant models combined. The species not covered by any model are mostly subspecies, like *Pueraria montana var. lobata*. A further finding is, that the models cover species of a similar domain. Bird, mammal and plant species in our list are covered well by at least one of the models. However, species living in the water or invertebrates are covered less.

4.2 Model Accessibility and Usability

Besides the species coverage, the accessibility and usability of the models is an essential criterion. Here we have two different kinds of model accessibility. For the first kind, the raw model is directly available. This is the case for the iNat2021 and the Microsoft model. Here, the model can be downloaded and deployed on an own machine. These models are also free to use. The second kind of models can be accessed through an API, which was the case for all other models we evaluated. The Pl@ntNet-API can be accessed directly and offers up to 500 free requests per day. Access to the Plant.id, iNaturalist, NIA and Flora Incognita APIs can be granted upon request by the model providers. The integration and use of Plant.id API and iNaturalist API is usually connected with costs, whereas the integration of Flora Incognita and NIA is for free. However, the costs for using the APIs can be discussed with the model providers. In addition to an API usage, the models from Pl@ntNet, Plant.id, iNaturalist and Flora Incognita also can be accessed in other ways with different costs, for example for free by using provided smartphone applications.

Apart from the accessibility, also the usage of the models differs. A first difference is the number of input images. Microsoft, iNat2021 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. Further, the numbers of species suggested by the model differ. For the downloaded models, it is possible to change this number. NIA and Flora Incognita always return ten results. For the iNaturalist API, the Pl@ntNet-API and Plant.id the number of results depend on the score of the predictions. For example, if the model predicts one species with a high score, only one suggestion is returned. A distinctive feature is that Pl@ntNet-API and Flora Incognita offer the user the possibility to indicate which part of the plant is shown on the image. Further, it is noticeable, that the iNaturalist API is the only API which not returns a score.

Table 3. Model Overview

Name	Source	Description
iNaturalist 2021 Competition (iNat2021)	Van Horn et al. (2021)	Van Horn et al. (2021) created different models based on the iNaturalist 2021 competition dataset and benchmarked them. From this benchmark, we selected the iNat2021 Supervised model as the most sufficient for our use case. Thus, the model is not directly an outcome of the competition, but was created with the corresponding dataset.
iNaturalist API	https://www.inaturalist.org/pages/computer_vision_demo	The iNaturalist API is developed by iNaturalist, a joint initiative by the Californian Academy of Science and the National Geographic Society. It is able to identify plant and animal species.
Microsoft AI for Earth Species Classification	https://github.com/Microsoft/SpeciesClassification	In the context of the AI for Earth program, the so-called "Species Classification API" was developed. It covers animal and plant species.
Nature Identification API (NIA)	https://observation.org/pages/nia-explain/	The Nature Identification API, is a joint effort by Observation International, Naturalis and Intel Corp. It is able to identify plant and animal species.
Flora Incognita	https://floraincognita.de/	The Flora Incognita research group at the Max Planck Institute for Biogeochemistry in Jena and the Technical University Ilmenau develop the Flora Incognita identification service, which allows the identification of plant images with predictions for the most likely species.
Pl@ntNet-API	https://my.plantnet.org/	The Pl@ntNet 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 providers have advised us to use their API-project "the-plant-list" for the identification.
Plant.id API	https://plant.id/	Plant.id is a project developed by the team of the FlowerChecker company, 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.

Table 4. Species Coverage of Models

Name	Species Coverage	IAS Coverage	
		Union concern (66)	Candidates (30)
iNat2021	10,000	43	16
iNaturalist API	38,000	57	25
Microsoft Species Classification	5,266	34	11
NIA	22,302	41	8
Flora Incognita	4,803	25 (out of 36 plants)	4 (out of 6 plants)
Pl@ntNet-API	29,615	26 (out of 36 plants)	4 (out of 6 plants)
Plant.id API	12,183	26 (out of 36 plants)	5 (out of 6 plants)

4.3 Updates of Models

Besides the current state of the models, it is important to inspect whether the models are updated regularly. The both downloaded models, will not be developed further while the other models are getting automatically updated by the model providers. If special changes e.g. concerning the species coverage are wanted, these can be requested from the model owners. The iNaturalist API is getting a big update soon, by increasing the covered species to 47,000 taxa (Shepard, 2021). After this update, they plan to implement two updates per year. The providers of NIA currently work on a completely new model, which will probably get published in 2022 and cover around 30,000 species for Northern Europe. Subsequently, it is planned to update the model annually. Pl@ntNet performs monthly model updates, increasing the species coverage or improving the model architecture. The model developed by Flora Incognita is updated regularly. In the beginning of 2022 they plan a bigger update for the model, which will then cover almost all plant IAS of Union concern. Also, the model owners from Plant.id perform regular updates and expect to cover all plant IAS of Union concern in summer 2022.

4.4 Accuracy of Models

One of the most important criteria to evaluate the models is the accuracy of these. Like described in Section 3.2.2, we performed two different tests. First, we tested all IAS which are covered by the models, which means that the test sizes are different for each model. Second, with same images from species covered by all models. An overview of the accuracies can be inspected in Table 5.

Like described in Section 3.3 and Section 4.2 the models had to be accessed in different ways. As peculiarity, the model from Flora Incognita was tested by the model providers, since it was too much effort to get access to the API for our purposes in the provider's opinion. We received the raw data of the results from them.

The highest Top-1 accuracy achieved Microsoft with 78.89 %, followed by the model developed by Plant.id (70.97 %) and Flora Incognita (66.67 %). Behind, iNaturalist API generated a score of 65.68 % followed by Pl@ntNet-API with 63.33 % and iNat2021 (62.12 %). The lowest percentage of images correctly identified achieved NIA (41.16 %). For the Top-5 accuracy, Plant.id had the highest accuracy with (93.01 %), followed by Microsoft Species Recognition (91.48 %) and Flora Incognita (88.17 %). Pl@ntNet-API (85.00 %), iNaturalist API (84.55 %) and iNat2021 (84.18 %) achieved similar accuracies. The lowest accuracy achieved NIA with 51.02 %.

The second test set contained twelve species, covered by all models. The highest Top-1 accuracies for this set were provided by Microsoft and Plant.id API, both with 80.56 %. Followed by, Flora Incognita (75.00 %), PlantNet-API (72.22 %), iNaturalist API (61.11 %), iNat2021 (58.00 %) and NIA (27.78 %). As only plant species are covered by all models, we also split this test for hybrid and plant only models. As result, the hybrid test set consisted of 30 x 6 images and the plant set of 21 x 6 images. Also, for the hybrid test set the Microsoft model achieved the best accuracy with 86.11 %, followed by the iNaturalist API (73.89 %), iNat2021 (69.74 %) and lastly NIA (46.11 %). For the plant set, the accuracies of the models are very similar. Pl@ntNet-API had 67.83 %, Flora Incognita 72.80 % and Plant.id 73.81 %.

5 Discussion & Conclusions

In this chapter, we are going to discuss the achieved results and will draw conclusions regarding the advantages of the different models. Our results indicate different strengths of the models. Depending on which criterion the focus is set, a model is more or less suitable.

The iNaturalist API and Pl@ntNet-API cover the most species and show sufficient accuracies. Therefore, they could be used to observe a high number of species and get a general overview of the distribution of these species. In case the priority is to correctly identify IAS, the models from Microsoft and Plant.id could be used. Both models have the best accuracies, but cover fewer species overall. However, here it is important that visually similar species also have to be covered by the model. To cover a certain type of organism (e.g. plant, mammal, bird), it can be beneficial to use a model which is especially trained for this organism. Therefore, if the focus should be on plants, it can make sense to choose a plant model with a high IAS coverage even though the accuracy is a bit worse than with the hybrid models. If the geographical region is prioritized, then it can also be interesting to use for instance NIA which is trained on species from Belgium and the Netherlands. Besides the accuracy and species coverage of the model, its accessibility is a decision criterion. The models from iNat2021 and Microsoft can be downloaded for free as raw models. To use the other models, it is necessary to contact the provider to discuss costs, expandability or integration methods concerning the APIs. So if the model should be integrated directly as a static element in an own system to avoid passing the images to other providers, it might be useful to use iNat2021 and Microsoft. Whereas then, no updates regarding the species and accuracies can be expected by the providers. If the priority is to have a model that is continuously updated with the latest IAS, it is advisable to consider to use the models available via API, whereby then communication to the model providers can become necessary and costs may apply.

Thus, which model fits best depends very strongly on the criteria the focus is on. If all criteria have a certain importance and none stands out, a combination can be a possible solution for the use case described in Section 1. For example, the iNaturalist API model has a high species coverage, both in general and specifically in terms of IAS, which can even be further increased by the combined use of a plant model. Here also arises the possibility to use the models one after the other. For example, if a hybrid model initially returns a plant as a prediction, a specific plant model can be used to double-check and thus improve the accuracy.

In general, it can be helpful to arrange with the providers to improve the model for the individual requirements, as all API providers offered collaboration possibilities. Besides using an existing model, developing a new custom one is also an option. Training a custom model is on the one hand a solution to cover all species of interest but on the other hand connected to high costs of implemen-

Table 5. Accuracies of Models

Name	Top-1 in %	Top-5 in %	Top-1 same species in %		
			All models	Hybrid	Plant only
iNat2021	64.12	84.18	58.00	73.89	-
iNaturalist API	65.85	84.55	61.11	69.74	-
Microsoft Species Classification	78.89	91.48	80.56	86.11	-
Nature Identification API (NIA)	41.16	51.02	27.78	46.11	-
Flora Incognita	66.67	88.17	75.00	-	72.80
Pl@ntNet-API	63.33	85.00	72.22	-	67.83
Plant.id API	70.97	93.01	80.56	-	73.81

Note: All APIs were tested on November 23 and 24, 2021. However, since Pl@ntNet-API discovered a bug after our test series, this API was tested again on February 17, 2022 after the bug was fixed.

tation, especially because of the high effort of collecting the necessary images. Additionally, besides the species of interest, it would also need to cover a lot of other species to distinguish between invasive and non invasive ones. To limit the effort, the iNat2021 or Microsoft model could be used as a basis and adapted.

To sump it up, depending on which criterion is most relevant, different models are suitable, whereby a combination of models, a collaboration with model providers or a new development can also be appropriate.

5.1 Limitations

In our research, we encountered several limitations, particularly concerning the transparency of the models, the naming of the used species and the availability of suitable testing data.

Most of the models we tested were not available for download but as API. As a consequence, due to the API access, the model is not always transparent. Although it is possible to send requests to the API and receive a corresponding response, it is often not possible to understand what is happening in between, the model seems to be a black box. It is unclear how and with which species and images the model was trained, what is happening with the images sent in the request and how further information is processed. This information is only available if it is provided by the operator of the API. Additionally, the number of requests to the APIs may be limited, although most API providers allow free requests, at least for a small scope. However, even if the models are fully available, as is the case for the downloadable models, it can be difficult to make them usable, due to missing documentation.

Another difficulty is the fact that some species have several scientific names. Although we have tried to cover all synonyms of scientific species names, it can happen that during testing a species is declared as not covered by a model, although it is covered but with a synonym. Finding suitable images for the test dataset required high effort. On the one hand, the images had to have the appropriate quality and should partly differ from each other. On the other hand, images that originate from the model providers can not be used in a test dataset. Testing models on images they were trained on, gives the corresponding model an advantage, which has to be avoided. Certainly, we have tried with our best knowledge and conscience to avoid this double use, but it can still happen that individual images were used for both, training and testing. Among other things, this may be due to the fact that there are a large number of connections among the model providers. The different providers support each other in the development of the models, provide financial aid and are sharing their training images. Additionally, we had to make the assumption that all of our test images we received externally depicted the species stated by the provider. Sometime this might not be the case, as platforms like GBIF often provide user-verified images.

A further issue is, that models covering a high number of species are possibly disadvantageous in our accuracy testing procedure. While testing models only with species they cover, a model with a relatively low number of overall species has a lower risk of making a wrong prediction, as the number of species that are available for selection is lower. Therefore, in addition to the tests, it is necessary to consider the number of covered species and, where possible, the ratio of covered species and correspondingly used training images.

5.2 Future Work

Based on the findings described above, several options for future work come up.

First of all, the considered models could be evaluated by further methods. Beside the Top-1 and Top-5 accuracy, additional metrics can be calculated and the conducted tests could be expanded on further species. Currently, just the IAS and candidates of concern specified by the European Union were evaluated, but in addition to that also the models performances on classifying images of IAS with regional relevance or species similar to IAS are of interest. For species similar to IAS, it is important to investigate if they are even covered by the models, i.e. if the model is able to distinguish between the different species. Otherwise, it just classifies a non invasive species as IAS because it does not know the non invasive species. When a model covers both, i.e. the invasive and non invasive species, its performance on distinguishing between them can be additionally evaluated.

Moreover, it is useful to know about the ability of models to incorporate additional information to the image itself and how well this enhances the prediction. Examples are whether they are able to take several images as input for one identification or if they can handle metadata like the location at which a species was observed. It is also interesting to evaluate if a model is able to use the information which part of a plant or animal is depicted or at which season the image was taken. Further, image parameters to be investigated are the angle and distance of the camera to the object, the quality of the image or the area of interest within an image, i.e. the percentage of pixels actually showing the object of interest. Depending on the use case also the model performance on empty images might be worth evaluating, i.e. whether a model is able to determine if there is no animal or plant at all in the image instead of just mentioning some random species.

Instead of using new evaluation methods, also additional models, not considered yet, could be evaluated. In the future there will probably emerge new models, some of the not accessible models could be made available or for other use cases further models are of interest. Several providers of the investigated models also announced updates coming soon which might be worthwhile to reevaluate in the future.

6 Acknowledgments

This contribution complements a detailed technical report that was already published (Jakuschona et al., 2022a). The 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. 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 and 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.

References

CABI: Invasive Species Compendium, Commonwealth Agricultural Bureaux International database (CABI), https://www.cabi.org/isc, 2022.

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

Crall, A. W., Jarnevich, C. S., Young, N. E., Panke, B. J., Renz, M., and Stohlgren, T. J.: Citizen science contributes to our knowledge of invasive plant species distributions, Biological Invasions, 17, 2415–2427, https://doi.org/10.1007/s10530-015-0885-4, 2015.

EASIN: JRC EASIN Invasive Alien Species Europe App, https://easin.jrc.ec.europa.eu/easin/CitizenScience/BecomeACitizen, 2022.

Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., and Fenton, N. J.: Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review, PLOS ONE, 16, e0234 587, https://doi.org/10.1371/journal.pone.0234587, 2021.

GBIF: Global Biodiversity Information Facility (GBIF), https://www.gbif.org/, 2022.

Goodchild, M. F.: Citizens as sensors: the world of volunteered geography, GeoJournal, 69, 211–221, 2007.

Goëau, H., Bonnet, P., and Joly, A.: AI-based Identification of Plant Photographs from Herbarium Specimens, Biodiversity Information Science and Standards, 5, e73 751, https://doi.org/10.3897/biss.5.73751, 2021.

iNaturalist: Taxa Info · iNaturalist, https://www.inaturalist.org/taxa, 2022.

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, European Commission, Joint Research Center, https://doi.org/10.2760/97305, 2022a.

Jakuschona, N., Stenkamp, J., and Niers, T.: EibSReM GitHub Repository, https://github.com/EibSReM, 2022b.

Jakuschona, N., Stenkamp, J., and Niers, T.: MicrosoftSpeciesClassification, https://doi.org/10.5281/zenodo.6190003, 2022c.

Johnson, B. A., Mader, A. D., Dasgupta, R., and Kumar, P.: Citizen science and invasive alien species: An analysis of citizen science initiatives using information and communications technology (ICT) to collect invasive alien species observations, Global Ecology and Conservation, 21, e00 812, https://doi.org/10.1016/j.gecco.2019.e00812, 2020.

Joly, A., Goëau, H., Botella, C., Glotin, H., Bonnet, P., Vellinga, W.-P., Planqué, R., and Müller, H.: Overview of LifeCLEF 2018: A Large-Scale Evaluation of Species Identification and Recommendation Algorithms in the Era of AI, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, edited by Bellot, P., Trabelsi, C., Mothe, J., Murtagh, F., Nie, J. Y., Soulier, L., SanJuan, E., Cappellato, L., and Ferro, N., Lecture Notes in Computer Science, pp. 247–266, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-319-98932-7_24, 2018.

Joly, A., Goëau, H., Kahl, S., Deneu, B., Servajean, M., Cole, E., Picek, L., Ruiz de Castañeda, R., Bolon, I., Durso, A., Lorieul, T., Botella, C., Glotin, H., Champ, J., Eggel, I., Vellinga, W.-P., Bonnet, P., and Müller, H.: Overview of LifeCLEF 2020: A System-Oriented Evaluation of Automated Species Identification and Species Distribution Prediction, in: Experimental IR Meets Multilinguality, Multimodality, and Interaction, edited by Arampatzis, A., Kanoulas, E., Tsikrika, T., Vrochidis, S., Joho, H., Lioma, C., Eickhoff, C., Névéol, A., Cappellato, L., and Ferro, N., Lecture Notes in Computer Science, pp. 342–363, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-58219-7_23, 2020.

Jones, H. G.: What plant is that? Tests of automated image recognition apps for plant identification on plants from the British flora, AoB PLANTS, 12, https://doi.org/10.1093/aobpla/plaa052, 2020.

Lotfian, M., Ingensand, J., Ertz, O., Oulevay, S., and Chassin, T.: Auto-filtering validation in citizen science biodiversity monitoring: a case study, Proceedings of the ICA, 2, 1–5, https://doi.org/10.5194/ica-proc-2-78-2019, 2019.

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

Microsoft: microsoft/SpeciesClassification, https://github.com/microsoft/SpeciesClassification, original-date: 2019-02-04T21:33:38Z, 2022.

Miele, V., Dussert, G., Spataro, B., Chamaillé-Jammes, S., Allainé, D., and Bonenfant, C.: Revisiting animal photo-identification using deep metric learning and network analysis, Methods in Ecology and Evolution, 12, 863–873, https://doi.org/10.1111/2041-210X.13577, 2021.

Mäder, P., Boho, D., Rzanny, M., Seeland, M., Wittich, H. C., Deggelmann, A., and Wäldchen, J.: The Flora Incognita app–interactive plant species identification, Methods in Ecology and Evolution, 12, 1335–1342, publisher: Wiley Online Library, 2021.

Niers, T., Jakuschona, N., and Stenkamp, J.: RequestCollectionComputerVisionAPIs, https://doi.org/10.5281/zenodo.6189239, 2022.

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., and Clune, J.: Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning, Proceedings of the National Academy of Sciences, 115, E5716–E5725, https://doi.org/10.1073/pnas.1719367115, 2018.

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., and Clune, J.: 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, 2021.

Observation.org: Explanation NIA, https://observation.org/pages/nia-explain/, 2022.

Pl@ntNet: World flora: Species - Pl@ntNet identify, https://identify.plantnet.org/the-plant-list/species, 2022.

Regulation: No 1143/2014 of the European Parliament and of the Council of 22 October 2014 on the prevention and management of the introduction and spread of invasive alien species, OJ, L317/35, 2014.

Schermer, M. and Hogeweg, L.: Supporting citizen scientists with automatic species identification using deep learning image recognition models, Biodiversity Information Science and Standards, 2, e25 268, https://doi.org/10.3897/biss.2.25268, 2018.

Shepard, A.: New Vision Model Training Started, https://www.inaturalist.org/blog/59122-new-vision-model-training-started, publisher: iNaturalist, 2021.

Stenkamp, J., Niers, T., and Jakuschona, N.: iNaturalist_Competition, https://doi.org/10.5281/zenodo.6190009, 2022.

Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., and Packer, C.: Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna, Scientific Data, 2, 150 026, https://doi.org/10.1038/sdata.2015.26, 2015.

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., Lukacs, P. M., Moeller, A. K., Mandeville, E. G., Clune, J., and Miller, R. S.: Machine learning to classify animal species in camera trap images: Applications in ecology, Methods in Ecology and Evolution, 10, 585–590, https://doi.org/10.1111/2041-210X.13120, 2019.

Terry, J. C. D., Roy, H. E., and August, T. A.: Thinking like a naturalist: Enhancing computer vision of citizen science images by harnessing contextual data, Methods in Ecology and Evolution, 11, 303–315, https://doi.org/10.1111/2041-210X.13335, 2020.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., and Belongie, S.: The iNaturalist Species Classification and Detection Dataset, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8769–8778, IEEE, Salt Lake City, UT, https://doi.org/10.1109/CVPR.2018.00914, 2018.

Van Horn, G., Cole, E., Beery, S., Wilber, K., Belongie, S., and Mac Aodha, O.: Benchmarking Representation Learning for Natural World Image Collections, arXiv:2103.16483 [cs], http://arxiv.org/abs/2103.16483, arXiv: 2103.16483, 2021.

Weinstein, B. G.: A computer vision for animal ecology, Journal of Animal Ecology, 87, 533–545, https://doi.org/10.1111/1365-2656.12780, 2018.

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

Wäldchen, J., Rzanny, M., Seeland, M., and Mäder, P.: Automated plant species identification—Trends and future directions, PLOS Computational Biology, 14, e1005 993, https://doi.org/10.1371/journal.pcbi.1005993, 2018.