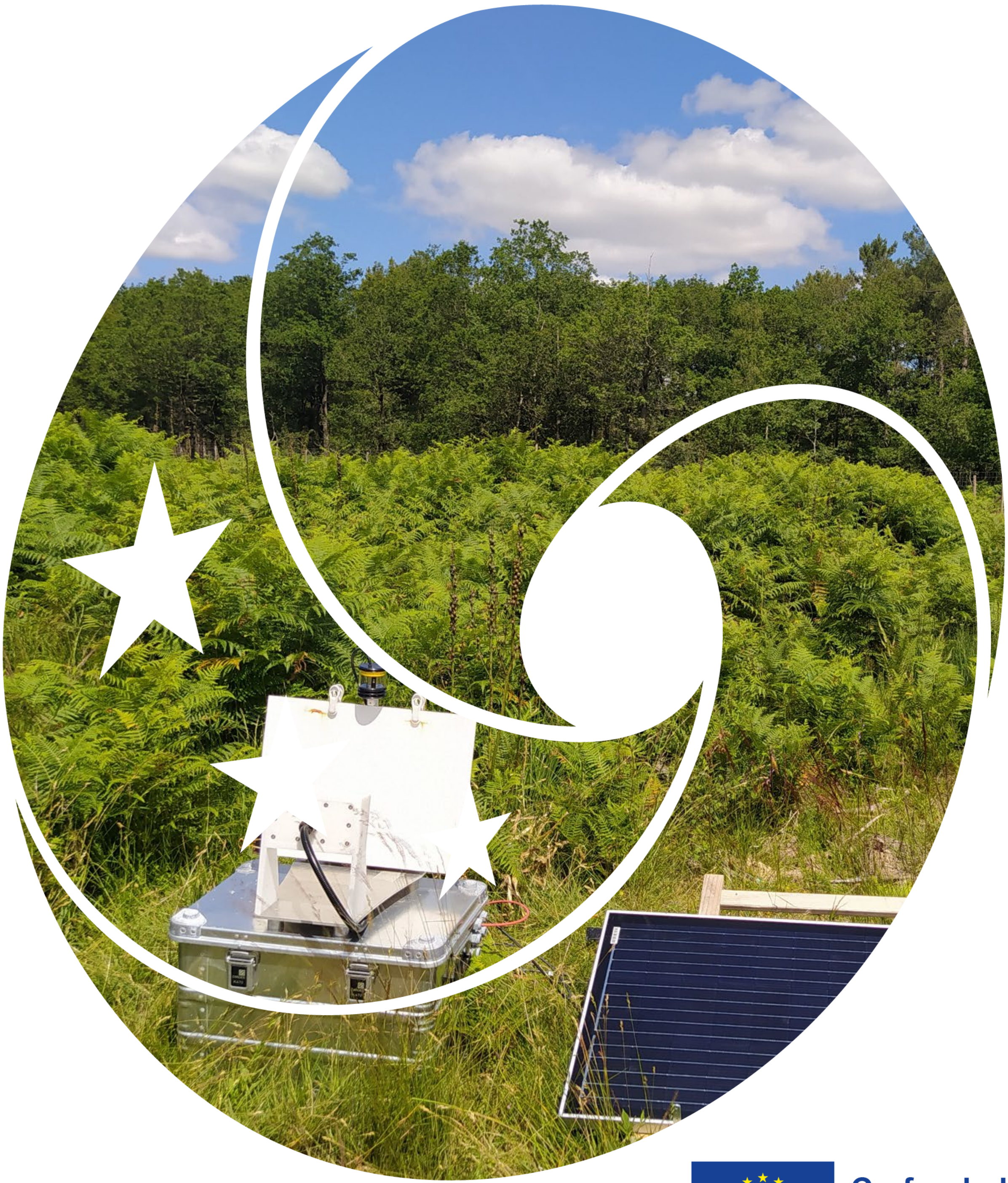
biodiversa+
European Biodiversity Partnership

EUROPEAN PARTNERSHIP

TRANSNATIONAL IMAGE-BASED MONITORING SCHEMES FOR INVASIVE ALIEN SPECIES

Lessons learned from the first two years of the Biodiversa+ pilot on “invasive alien species”



Co-funded by the European Union

Document Information

Grant Agreement number:	101052342
Project acronym	Biodiversa+
Project full name	The European Biodiversity Partnership
Biodiversa+ duration:	7 years
Biodiversa+ start date	Start date: 1st October 2021
More information about Biodiversa+	Website: www.biodiversa.eu Email: contact@biodiversa.eu @Biodiversa.eu Biodiversa+

Deliverable title:	2nd mid-term report “Monitoring invasive alien species with image-based methods”
Authors:	Toke Thomas Høye, Lars Dalby, Kavi Mellerup (MoE_DK)
Contributors/reviewers:	Jacopo Breschi (BOZEN), Barbora Šingliarová (SAS), Alix d’Audeville (OFB), Arnaud Albert (OFB), Afonso Prestes (FRCT), Lars B. Pettersson (SEPA), Zdeněk Račanský (NCA_CZ), Bram D’hondt (VL O); Petteri Vihervaara (MoE_FI)
Work package:	WP2 Promote and support transnational biodiversity monitoring
Task or sub-task title:	Sub-task 2.6.2 sub-pilot on invasive alien species
Lead partner:	MoE_DK
Date of publication:	April 2025
Cover page illustration	© Alix d’Audeville

Table of contents

What is Biodiversa+

The European Biodiversity Partnership, Biodiversa+, supports excellent research on biodiversity with an impact for policy and society. Connecting science, policy and practise for transformative change, Biodiversa+ is part of the European Biodiversity Strategy for 2030 that aims to put Europe's biodiversity on a path to recovery by 2030. Co-funded by the European Commission, Biodiversa+ gathers 81 partners from research funding, programming and environmental policy actors in 40 European and associated countries to work on 5 main objectives:

1. Plan and support research and innovation on biodiversity through a shared strategy, annual joint calls for research projects and capacity building activities
2. Set up a network of harmonised schemes to improve monitoring of biodiversity and ecosystem services across Europe
3. Contribute to high-end knowledge for deploying Nature-based Solutions and valuation of biodiversity in the private sector
4. Ensure efficient science-based support for policy-making and implementation in Europe
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context.

More information at: https://www.biodiversa.eu/

Executive Summary

The European Biodiversity Partnership, Biodiversa+, is testing ways of harmonising biodiversity monitoring at a pan-European scale. This work is conducted through some pilots. In line with the Biodiversa+ priority on "Invasive Alien Species", the Biodiversa+ pilot on “Monitoring invasive alien species with image-based methods” was launched in 2023. It aims to develop and evaluate scalable methods using novel technologies for monitoring invasive alien species on a continental scale. This report describes the outcomes of work done during 2024 as well as some initial perspectives for long term monitoring.

During 2024, the pilot involved 11 active partners across Europe. These included all partners from the first year of the pilot and VL O (Government of Flanders, Belgium), represented by INBO. All partners participated in the plant module and the insect module of the pilot. Both modules employ image-based monitoring approaches to the monitoring of invasive alien species. The plant module trials a camera system (CamAlien) mounted on a car to map invasive alien plant species along roads, while the insect module trials insect camera traps (AMI traps) for monitoring invasive alien moth species attracted to light. During the 2024 season, each partner has collected data in their own country. The data collection in the plant module has mostly progressed as planned and the most relevant subset of the resulting data has been processed using a data management and data processing pipeline developed in collaboration with Pl@ntNet. Data collection in the insect module has also progressed well in 2024. In October 2024, it was decided to further extend the IAS pilot for three more years (2025-2027). A plan for potential implementation of these tools in transnational monitoring of invasive alien plant and insect species will be developed as part of this 2nd extension of the pilot.

The pilot was coordinated by the Ministry of the Environment Denmark through Aarhus University as third party, and was conducted with ten active partners: The Autonomous Province of Bolzano, Italy (BOZEN); Denmark (MoE_DK); France (OFB); Sweden (SEPA); Czech Republic (NCA CZ); Slovakia (SAS); Bulgaria (ExEA); Croatia (MESD); Autonomous Region of the Azores, Portugal, (FRCT); Flanders Region, Belgium (VL O); and Israel (MoEP).

Table of acronyms

AMI traps	"Automated Monitoring of Insects" traps based on a prototype from a research team at Aarhus University, Denmark and further developed into a standardised hardware system by UKCEH
API	Application Programming Interface
CamAlien	Camera system for monitoring invasive alien plant species along roadside verges. Produced by The AI Lab (https://theailab.dk/)
EUNIS	European Nature Information System (https://eunis.eea.europa.eu/)
ERDA	Electronic Research Data Archive, an open science data storage facility at Aarhus University, Denmark (https://www.erda.au.dk/)
EUNIS	European Nature Information System (https://eunis.eea.europa.eu/)
GBIF	Global Biodiversity Information Facility (https://www.gbif.org/)
GDPR	General Data Protection Regulation. Regulation (EU) 2016/679
IAS	Invasive Alien Species
JRC	Joint Research Centre
STING	Science and Technology for Pollinating Insects
SPRING	European project on "Strengthening pollinator recovery through indicators and monitoring"

Introduction

Evidence is mounting of widespread biodiversity decline across the globe. This gives a stark warning for the perilous state of our planet, yet the evidence base remains biased to a few regions and a few species groups which are feasible to monitor. General conclusions on the status of biodiversity are complex given their diverse ecologies and high variability between and within taxon groups, over time and geographic regions. Repeatable sampling methods that can automate and expand the extent and resolution of biodiversity monitoring are urgently needed to provide robust estimates of long-term trends. Closing the knowledge gap has never been more important.

Biodiversa+, the European Biodiversity Partnership, aims at promoting and supporting transnational biodiversity monitoring, by building a transnational network of harmonised biodiversity monitoring schemes on specific priority topics. One of these Biodiversa+ priorities focuses on invasive alien species (IAS) (Basille et al. 2023). As a way of supporting this harmonisation work on this priority, the Biodiversa+ partners agreed to launch in January 2023 a pilot on monitoring invasive alien species with image-based methods. The project was initially planned for one year (2023), then extended one more year (2024) and in October 2024 extended for three more years (2025-2027)[1]. This report concerns the work done in year 2 (2024). The pilot gathers 11 countries: Italy - Autonomous Province of Bolzano with BOZEN, Denmark with MoE_DK, France with OFB, Sweden with SEPA, Czech Republic with NCA_CZ, Slovakia with SAS, Bulgaria with ExEA, Croatia with MESD, Portugal - Autonomous Region of the Azores with FRCT, Belgium with VL O, and Israel with MoEP.

The aim of this pilot is to pave the way for and evaluate scalable methods using novel technologies for monitoring invasive alien species at the geographical scale of the partnership. More specifically, the pilot will: a) Implement a coordinated international image-based monitoring scheme for invasive alien plant and insect species across biogeographical regions, b) Improve image recognition for invasive alien plant and insect species, c) Map benefits and constraints to real-time mapping of invasive alien plant and insect species and presentation in decision support tools, and d) showcase the workflow of image-based monitoring for early detection of invasive alien species.

In the context of Biodiversa+, this task is particularly linked to Biodiversa+ work on the harmonisation of protocols, methods and databases and promotion of novel technologies to monitor biodiversity.

The specific objectives of the second year (2024) of the pilot were:

1. Each active partner has collected ten hours of image data along road sections with known occurrences of invasive alien plant species during the 2024 growing season.
2. Each active partner has had three insect camera traps in operation for at least three months during the 2024 growing season.
3. A database hosting all image data and associated metadata is available by the end of the project.
4. An automated upload and image-processing workflow to localise and identify plant species and macro moths of invasion concern across the countries involved in the pilot project.

[1] Biodiversa+ biodiversity monitoring pilot: https://www.biodiversa.eu/biodiversity-monitoring/pilot/

5. Visualisations of the locations, identity, and time of observation of at least 100 plant and insect species have been prepared by the end of the project.
6. Dialogue with rail, road, and river shipping authorities in at least five countries have been established and trials applying the image recording platform (CamAlien) on trains and or boats in at least three countries have been carried out.
7. A data partnership agreement has been established with the EASIN (https://alien.jrc.ec.europa.eu/easin) for publishing data from the pilot scheme in their database.

This 2024 interim report describes the implementation steps taken in this pilot towards the transnational monitoring of invasive alien species, including a description of the equipment, taxonomic focus, site selection, data collection and image analysis. Further, the report summarises data management practices in the project and gives an overview of the first results and lessons learnt. The report ends with sections on the perspectives for implementation in long-term transnational monitoring. This report complements the living document describing the data management plan and the work plan for the 2nd extension of the project during 2025-2027.

1. Implementation

The IAS pilot consists of two modules: a plant module and an insect module. Both modules employ image recognition using computer vision and deep learning models for cost efficient and rapid detection of species of concern. These methods rely on training data of the species of concern to function effectively. Such data is available in the Global Biodiversity Information Facility (GBIF) as well as in databases of various national and regional pilot projects. The pilot program will be relevant in the context of introduction sites as well as for the collection of additional training data to improve image recognition models.

1.1 Equipment

Invasive alien plants are monitored with a custom-built camera system mounted on a car. During driving, the camera records continuously and is capable of recording images without motion blur at driving speeds up to 110 km/h. A web application accessible on a smartphone can be used to monitor camera settings in real time and view photos. An earlier version of the system used in the pilot is described in Dyrmann et al. (2021). The equipment used in this project is named CamAlien and is produced by the company The AI Lab. In addition, a series of Youtube videos give detailed descriptions of the operation of the system.

Insects are monitored with Automated Monitoring of Insects (AMI) traps, which are standardised camera-enabled light traps described by Bjerge et al. (2021) and https://www.insectmonitoring.org/. The AMI insect camera trap attracts nocturnal insects with UV light (LepiLED) and records insects landing on a white board. The AMI traps can be deployed with grid power or using solar panels and batteries. The traps run automatically on a predetermined schedule. The AMI trap images provide a rich basis for large-scale species inventories and abundance time series.

1.2 Taxonomic focus and site selection

1.2.1 Plant module

The recording of images at high speed requires that focal plants are clearly visible and identifiable from images recorded at a distance of 5–10 metres. This has led to a primary focus in the pilot on larger herbaceous or climbing, shrubby or woody species. Each partner was asked to provide a list of relevant plant species for their region. This was done to focus efforts on a manageable number of species. During the first year of the project, a collaboration with Pl@ntNet was established to build the image recognition pipeline around their facility.

The CamAlien camera system is designed for roadside verge monitoring. This is relevant in the context of the establishment and spread of invasive alien plant species, as roads and other transport networks are considered one of their primary dispersal routes. The camera system is capable of recording images along any road, but for consistency in the pilot scheme it was recommended to focus data collection on highways. During the 2024 field season, tests were also made using CamAlien to perform recordings from trains and along canals and rivers.

1.2.2 Insect module

The project is developing a broad taxon classifier as well as a species-level classifier for Lepidoptera. The classification models will allow for the species identification of moth species, which are identifiable from images recorded with the AMI trap cameras. This constrained the taxonomic focus to the subset of Lepidoptera species often referred to by naturalists as macro moths. These are typically, but not always, larger and with more distinct morphological traits.

Regarding site selection, there is limited knowledge about the distribution of invasive alien insect species. As such, the AMI traps have the potential to reveal the occurrence of species not known to the region. Ideally, potential establishment sites for invasive alien species should be selected for the deployment of AMI traps. At the same time, there are logistical considerations with the deployment of AMI traps as they require a power source (e.g. grid or solar) and they are costly equipment with a risk of unwanted interference with the traps or even theft. Given these considerations, the work plan specified that primary sites should be botanical gardens or ports. For some partners, it was necessary to widen the scope to include other potential introduction sites for invasive alien insect species, such as plant nurseries or urban sites.

1.3 Data collection

1.3.1 Plant module

Each partner collects images during continuous recording for ten hours of driving. While driving, an expert in identification of invasive alien plant species (not the driver of the car) scans the roadside verge for the invasive alien plant species of concern. The camera equipment allows for tagging of image sequences when species of concern are observed. By pressing a button, a tag is added to images recorded during the previous few seconds. By tagging subsets of the images with observed species of interest, the co-pilot can help drastically minimise the search for data to evaluate model performance and ultimately feed additional data to the training of better image recognition models. Once satisfactory model performance has been achieved, future monitoring of invasive alien plant species will only require the driver and no expert in plant identification and can thus be deployed at scale.

1.3.2 Insect module

Each active partner deploys three AMI traps. The AMI traps record images automatically every night on a predefined schedule (see section 3.1). During the 2024 field season, the ecologically relevant context of each AMI trap location was characterized through vegetation surveys. The goal was to assess the vegetation and habitats from each site with an AMI trap in three steps (at 5 m, all species; 25 m, woody species; 100 m, EUNIS habitats).

1.4 Image analysis

1.4.1 Plant module

The data upload and image analysis pipeline have been integrated in the data management system of the pilot to ensure that all images uploaded by the partners are managed automatically and in a consistent manner. First, the partners upload their data to the Electronic Research Data Archive (ERDA, erda.au.dk) at Aarhus University (AU). After the user has uploaded an image, an API developed by the pilot coordinators is called and as a first step it will determine if the image name is parsable. Next, the metadata is stored in the database and the image moved to storage. Lastly, a call to Pl@ntNet is made and the results returned are written to the database. A detailed flow diagram of the steps involved in this process is presented in Fig. 1. Collaborators at Pl@ntNet have prepared a dedicated adaptation of the Pl@ntNet application programming interface (API) involving custom-adapted parameters. This pilot was the only user of this facility, but since its start there have been additional requests and it is likely to be maintained in the future. This process involves calling the API several hundred times for each image to assess the likelihood of occurrence of the species of interest in multiple regions of each image and summarising the outcomes of these calls into image-level metrics of confidence scores as the maximum score across all calls to the API.

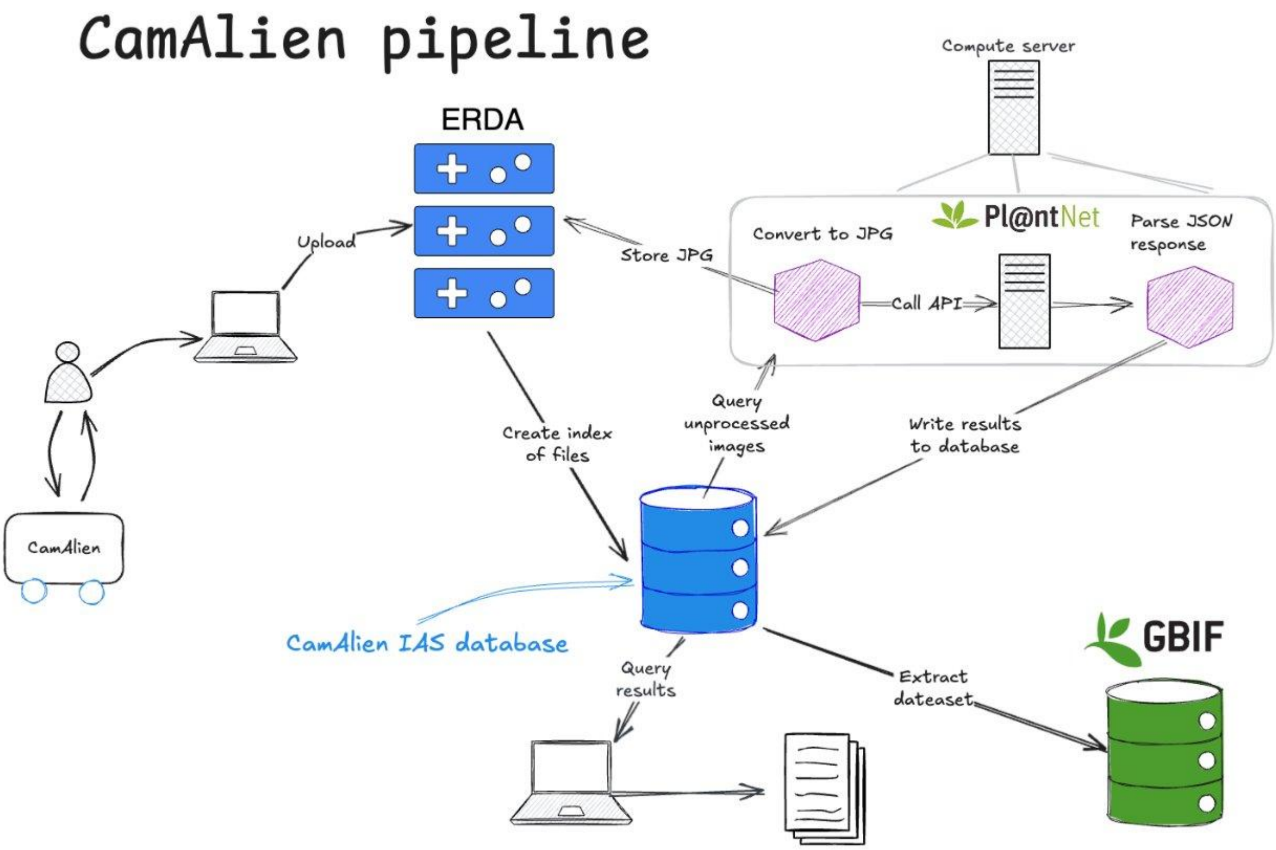


Figure 1: Analysis pipeline for the plant module, describing the process for each image recorded with the CamAlien system from the plant module uploaded to the Electronic Research Data Archive (ERDA) at Aarhus University and the interaction with the Pl@ntNet API.

1.4.2 Insect module

The extraction of species level data from the AMI trap images involves three key steps: localisation, broad taxon classification and species level classification. All three steps of the pipeline are complex tasks for which currently three different deep learning models are involved. Like for the plant module, partners upload their data to the Electronic Research Data Archive (ERDA)[2] at Aarhus University. After the user has uploaded an image, it is indexed and the path to the image and its metadata is written to our AMI database. Through a selection of localisation and classification models, images can be processed and the results are again written to the AMI database. A detailed flow diagram of the steps involved in this process is presented in Fig. 2.

The localisation model is assumed to work generically for images from any region, but the current set of classification models are specific to particular biogeographic regions. For this project, additional work on improving the localisation model is carried out and a new regional classification

[2] ERDA: https://erda.au.dk/

model has been developed. A first version of both models is now available. The new localisation algorithm ("Flatbug"; Svenning et al. in prep.) for use on the AMI trap images is a YOLOv8 instance segmentation model. It has been pre-trained on images from several other automatic insect localization projects. To improve its performance, all insects from >150 AMI trap images have been manually annotated and used to train this model. The model will be further improved but is already very promising (Fig. 3). Using instance segmentation for insect localization offers a major advantage compared to bounding boxes. When two or more insects are close to each other, bounding boxes can end up containing several insects. It is especially the case for the largest ones. Using polygons allows to extract each insect more precisely, which at the end, will facilitate the classification process.

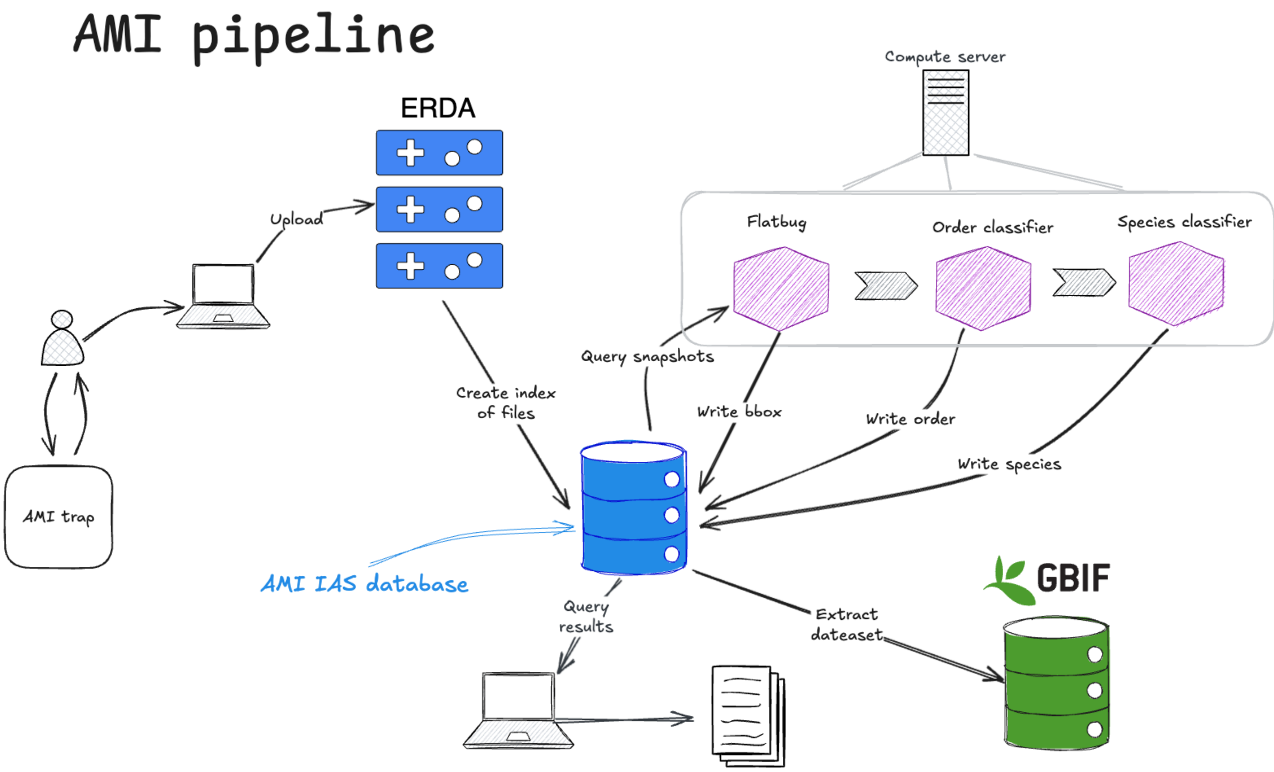


Figure 2: Analysis pipeline for the insect module, describing the process for each image recorded with the AMI system from the insect module uploaded to the Electronic Research Data Archive (ERDA) at Aarhus University and the use of custom-built deep learning models developed as part of the pilot.

For the localisation tasks, a key advantage is that the insects found in the AMI trap images are also represented in global citizen science databases and GBIF. Although these images have many different inherent biases and are taken with different cameras against different backgrounds, they offer an important resource when training image recognition models. In this project, we use five steps to create the species classification model: (1) choosing a backbone model, here we used EfficientNet (Tan & Le 2019), (2) applying an appropriate classification head, (3) collecting and training on citizen science images of the species-of-interest from GBIF, (4) manual annotations of a small number of localised objects from the AMI images and (5) fine-tuning on the manually annotated objects from the AMI images.

The GBIF dataset currently used is based on a list, kindly provided by the Dutch Butterfly Conservation, containing 3235 distinct macro-moth species found in Europe, and records of the countries, which are considered within their range. After conservative name resolution with the GBIF Backbone Taxonomy (i.e. the taxonomic reference that GBIF uses to consistently name species from multiple sources), the GBIF database was queried for observations (with images) for every species-range combination, yielding at least one observation for 2331 of the 3235 moth species and a total of 2,574,470 observations. This dataset was then filtered to exclude images of larvae, pupae, or large images where the moth takes up a small proportion of the image. The remaining dataset consisted of 1,873,998 observations of clearly visible adult moths. Lastly, the distribution of species frequencies was capped to remove the excessive amount of observation related to a few common and charismatic species: we chose to limit the maximum number of observations to 5,648 (from 41,117), leading to a retention of 80 % of the images and a final dataset size of 1,499,199 images.

The recently published CamTrap DP model (see: Bubnicki et al. 2023) is being adapted to accommodate the output format of the machine observed data from this pilot. A collaboration between INBO and AU has been established to publish datasets from the pilot to GBIF. The plan is to publish datasets from one partner at a time to GBIF during 2025 to allow each partner to adapt their license conditions.

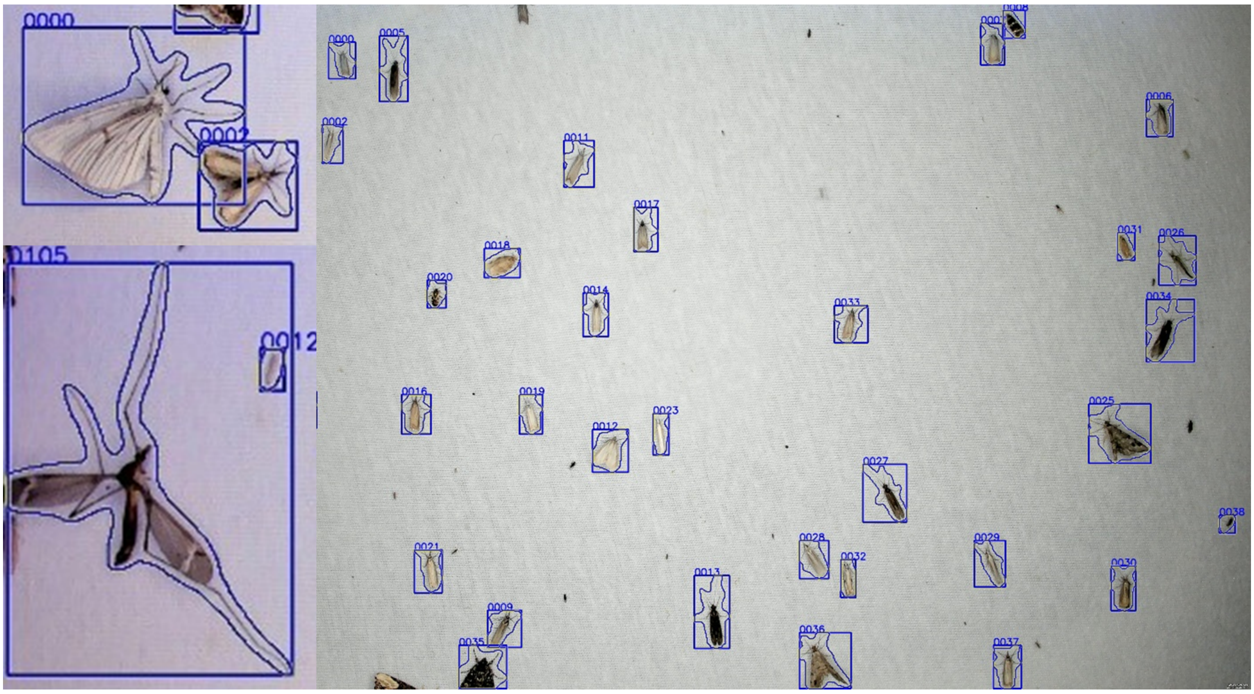


Figure 3: Example of an image recorded with the AMI trap. The new localisation model for AMI trap images draws a segmentation polygon as well as the rectangular bounding box around each detected animal. These are also shown in the main image as well as in two insets of zoomed in details from a similar image highlighting details for individual animals (Svenning et al., unpublished data).

2. Data management

2.1 Data description

For both modules, the species lists provided by the partners included GBIF IDs to ensure smooth integration with classification algorithms using these same GBIF IDs for training and prediction species identities.

The CamAlien system records images (4096 × 3000 pixels in tiff format as well as a compressed version of each image in jpg format) in two different user defined modes. During the IAS pilot, the recording is continuous when the system is powered on. It is also possible to run the system in a mode, where recording is only done when a remote-control button is pushed. Following a software update of the CamAlien system, the recording frequency (frame rate) is now adapted to the driving speed. This means that more images are recorded per second as driving speed increases. The maximum frame rate is five images are recorded per second, but when driving slowly, the frame rate is reduced to minimize the capture of redundant data generated by recording the same road section multiple times. Each image is stored in a buffer on the computer, which is part of the system. The location (latitude, longitude and altitude), speed during recording, exposure time, gain, date and time of recording are stored as part of the filename of each image. In addition, the user can add a tag stored in the filename to indicate the presence of invasive alien species by pressing a button on a remote control. The CamAlien system can inadvertently record images of people or registration plates of cars. During the recording process a software installed on the system directly processes each image and blocks out detected people, bikes and cars to comply with the General Data Protection Regulation of the European Union (GDPR).

The AMI traps record images (4096 × 2160 pixels in jpg format) based on a user defined schedule. During operation, the camera records a snapshot image every 10 minutes. The system also runs a motion detection software and evaluates if motion has happened in the past two seconds and records an additional image if this was the case (Bjerge et al. 2021). The date and time of recording are stored as part of the filename of each image.

2.2 Ensuring the FAIR data principle

All image data collected as part of the IAS pilot is stored in the Electronic Research Data Archive (ERDA) at Aarhus University, Denmark and will be made publicly available through a Danish implementation of the dataverse.

All data created in the pilot study will be made openly available at the end of the pilot. During the first year, the project has worked directly with the European Alien Species Information Network (EASIN) to ensure image-based species observations will be presented through EASIN's data portal, and associated with appropriate measures of uncertainty. The intention is to submit species occurrences to the Global Biodiversity Information Facility (GBIF), and through tagging allow EASIN to present the data directly on their species occurrence portal.

3. Results and lessons learnt

3.1 Equipment orders, delivery and functionality

During the 2024 season, the CamAlien system has functioned mostly without problems. The software update to the CamAlien was implemented ahead of the field season in the beginning of the year. The main feature of the software update was the addition of the adaptive recording schedule, where the recording frame rate is controlled by the driving speed. The software update also allowed users to choose more flexibly whether to record uncompressed (TIFF) images, compressed (JPG) images or both. In connection with experiments with recording images from trains and other vehicles, in certain situations CamAlien was powered by external 12V batteries. Although this is logistically challenging, it turned out to be possible.

Several partners had experienced that the SSD hard drives in the CamAlien would stop working properly following hard shutdowns of the systems during the previous field season. Following substantial testing by The AI Lab using the Danish partner equipment, it was not possible to reproduce the problem and during 2024 following the software update, no partner has experienced this problem again.

A thorough evaluation of AMI trap issues has been carried out with the supplier UKCEH in the beginning of 2024. Still, some AMI traps have continued to cause problems during 2024. A few key issues affected multiple partners. First, the real-time clock batteries supporting the RockPI were depleted more quickly than anticipated. This does not seem to be a problem for AMI traps installed with Raspberry PI computers. Another recurrent problem was that RockPI computers would stop working. Several partners were supplied with new Raspberry PI computers and had to change their real-time clock batteries. Still, some AMI traps did not work as expected and a few traps completely failed to operate during the 2024 field season.

3.2 Data collection and data transfer

During the 2024 field season, data collection in the plant module went well. By implementing the adaptive recording, we managed to drastically reduce the amount of redundant data collected. All partners expressed enthusiasm about this new feature, which also made data handling much easier than during the previous field season. The CamAlien system produces approximately 2.57 GB of image data per kilometre driven with continuous recording. In 2024, we relied on the SFTP clients FileZilla for data upload and partners were asked to only record and upload the uncompressed TIFF files. This reduced the amount of data to upload substantially.

During the 2024 season, several partners established dialogues with train operators in their respective countries to explore options for using CamAlien to monitor invasive alien plant species along railway lines. In France, the CamAlien was deployed on SNCF trains, whereas in the Autonomous Province of Bolzano (Italy) on SAD trains. During these operations, it became clear that it is more difficult to identify plants from the side windows of trains because of the speed (trains travel at up to 140/160 km/h), whereas images from the front view could be available from other collection pictures tools, but would require a different data analysis. In addition, there is not enough space in the driver's cab to install the CamAlien.

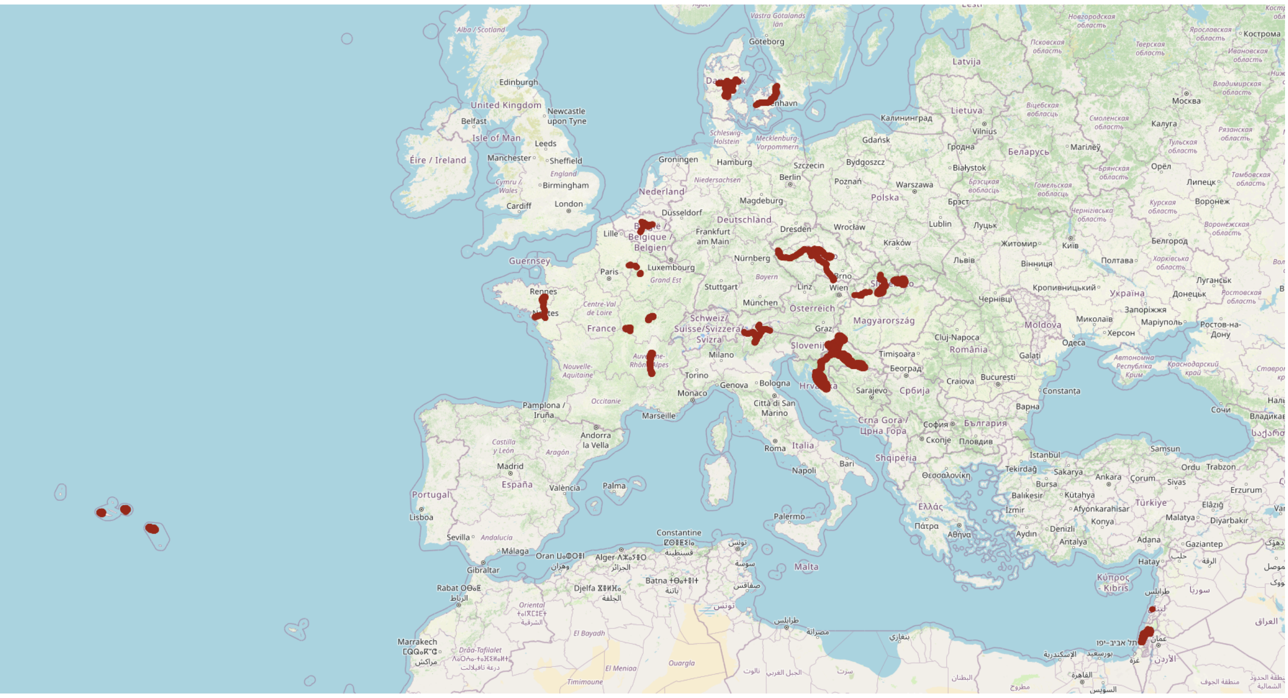


Figure 4: Map of all locations in the database of data recorded with CamAlien during the 2024 field season. Data was also collected in Bulgaria and will be added to the database.

However, the CamAlien image quality was still good and reflections from the side windows did not cause any problems. Unfortunately, it was not possible to get a reliable GPS signal inside the train with the current hardware configuration. For this reason, the images recorded in trains could not be geolocated at present. Through dialogue with The AI Lab – the providers of CamAlien – it seems that a stronger GPS antenna may solve the problem. We will explore this option during 2025.

In France and Belgium, the partners also experimented with taking photos with the CamAlien along rivers and canals. The camera was mounted on a car travelling along the road parallel to the river. From the French experience, It turned out that it was difficult to take a photo of both sides of the river and of the river itself (for terrestrial and aquatic plants) because of the width of the channels (from 5 to around 10 metres). It was also difficult to see aquatic plants because of the reflection of herbaceous plants and trees on the water. One idea would be to fit the camera with a polarisation filter. The speed of the car should be 30 km/h, as the path along the rivers is used by cyclists and walkers. This could be slow enough to monitor invasive alien species in an automated process.

For the insect module, the AMI traps produced much more data during the 2024 field season than was the case in 2023. Most partners managed to cover at least three months of the field seasons, although data gaps are common (Fig. 5). The AMI trap data was uploaded to the ERDA data storage facility at Aarhus University using the file handling software FileZilla. This system facilitates efficient transfer of data but does not allow users to add additional metadata related to the raw data. As such, it introduces a risk of loss of key information related to the raw data such as hardware metadata and location of the AMI trap during deployment. To allow efficient handling of data from multiple partners, it is also critical that the data follows a consistent structure. A customized upload module could ensure this more efficiently than the current approach using partner-specific upload sites (sharelinks) and the generic file transfer software Filezilla. In 2025, it is planned to trial a system of online communication between one AMI trap per partner and the data storage facility. In the future, this may further streamline the process of standardised data structures and efficient collection of deployment metadata.

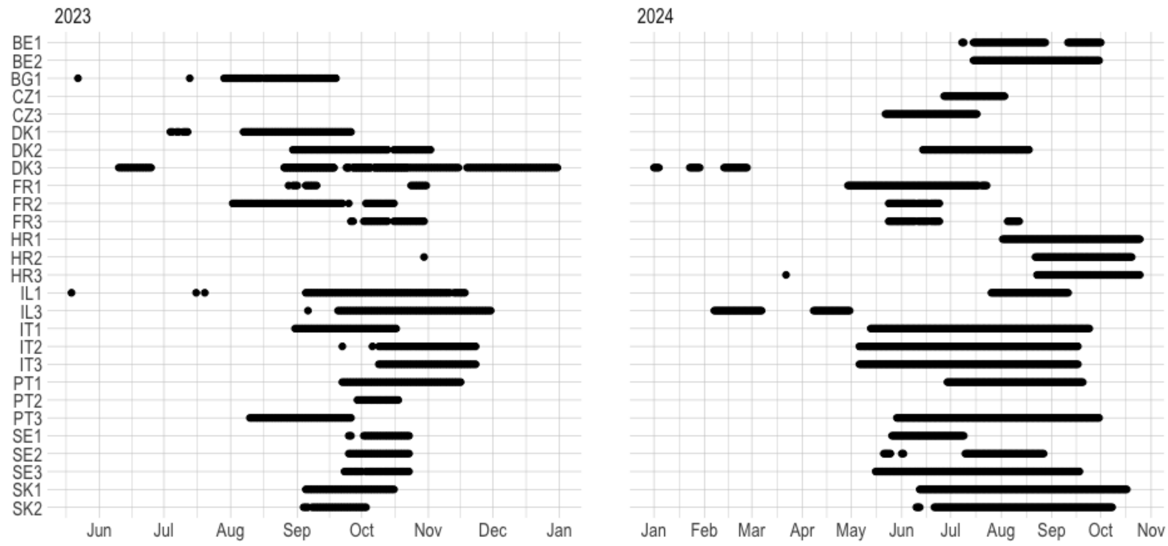


Figure 5: Recording activity for each AMI trap in 2023 and 2024. Data from a few traps deployed during 2024 have not yet been uploaded.

3.3 Image processing and annotations

3.3.1 Plant module

Each image recorded with CamAlien is divided into several hundred subsections or tiles of different sizes and each tile is classified with the Pl@ntNet algorithm. For each tile, the algorithm returns a confidence score for each detection of a plant species. We summarise these scores for a given CamAlien image by taking the maximum confidence score for each of the selected plant species in the project. However, translating the confidence scores to a probability of occurrence is not straightforward. This is particularly true because the Pl@ntNet classifier was not trained on images from CamAlien but typically on close-up images of a single plant or even just parts of a plant.

To establish relationships between confidence values from the classification models and probability of occurrence in the images, we selected images with scores along a gradient from 0.1 to 1 (the maximum possible value) and conducted expert review of the detections in 2023. The selection of images was done in a strategic manner to include images with varying probability of occurrence of the relevant plant species for the project.

During 2024, additional image review of CamAlien images was performed by the partners as it was clear after annotation in 2023 that annotation of images with scores from 0 to 0.1 was needed, particularly for some species that otherwise would have flat curves where a threshold could not be determined. Specifically, we selected up to 20 images per species and partner from images with scores from 0 to 0.1. As the number of target species varies greatly among partners, so does the number of images sampled. Annotation has been done voluntarily hence the annotated dataset is not perfectly stratified across partners and species. Nevertheless, it is possible to estimate the threshold where a detection is true in 50% of the cases across the entire dataset (Fig. 6).

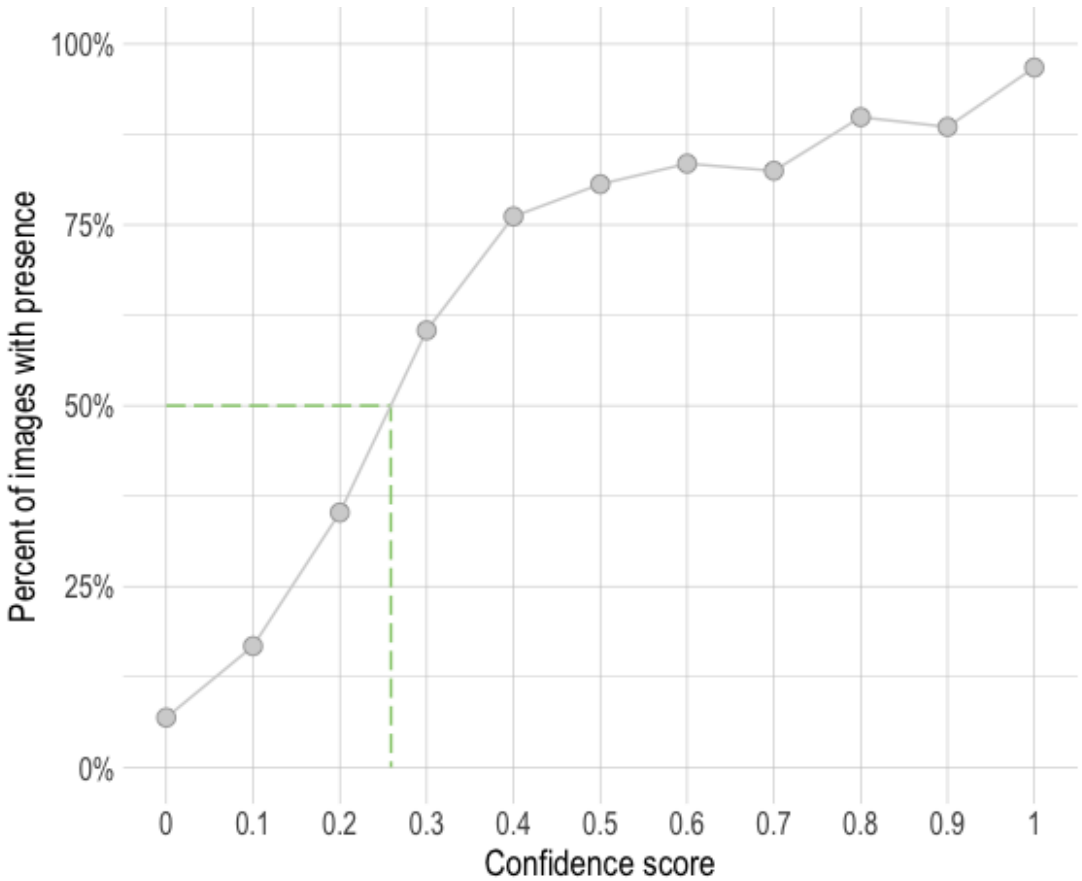


Figure 6: Threshold derived from annotated dataset. Relationship between the percent of the images with the focal species correctly detected by the Pl@ntNet algorithm and the confidence score from the model. Dashed green line shows the threshold confidence score across all focal species at a threshold of 50% (*Threshold 0.259)*. Sample size: 4047.

It is possible that the TIFF images capture details of importance for species recognition that are lost if the image is compressed to jpg format. On the other hand, a small compression of the original file can greatly reduce the amount of storage space needed for the dataset. To understand if image compression resulted in a reduced ability to correctly locate and identify invasive alien plants in the images, we conducted an experiment on the entire set of expert labelled images. This dataset consisted of a total of 4047 images (2469 images with the evaluated plant species present and 1578 images where it was absent). For each image, where the occurrence of a particular invasive alien species was evaluated, we compressed the image to each of the following percentages of original image quality: 95%, 90%, 85%, 80%, 70% and 50%. We then ran all versions of each image through the classification pipeline and evaluated the maximum confidence score for the relevant plant species. This allowed us to evaluate the proportion of images which would be misclassified as a function of varying the confidence value threshold used to determine if the species would be considered detected or not in the image (Fig. 7).

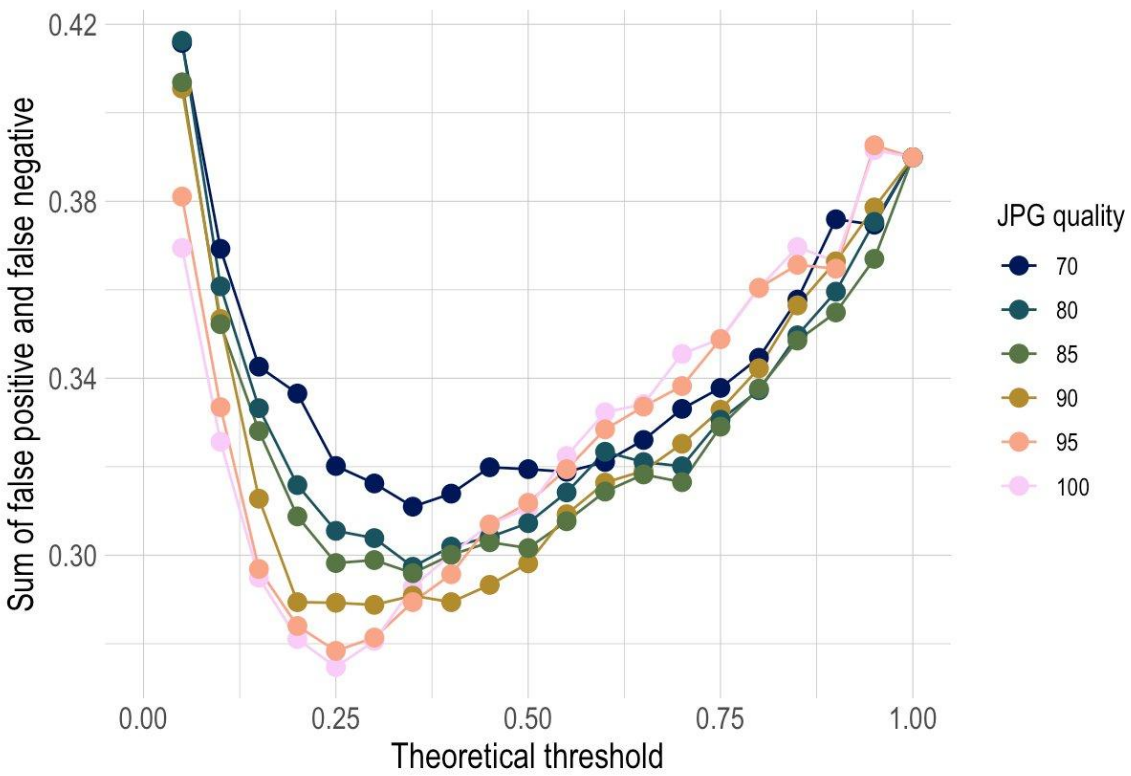


Figure 7: Proportion of the sum of false positives and false negatives across all images in the expert-reviewed dataset of images from CamAlien as a function of different confidence level thresholds. Each colour represents images in either original quality or a reduced image quality due to JPG compression. The figure illustrates that the false classifications are minimized when species receiving a maximum confidence level of 0.25 from the classification algorithm are assumed to be present in the images. The figure likely masks substantial species level variation.

3.3.2 Insect module

An overview of the AMI images recorded during the 2024 field season that have been uploaded to the ERDA facility and processed with the AMI pipeline is presented in Fig. 2. Data from some additional traps are still awaiting upload and could not be taken into account in these first analyses. The currently available dataset includes a total of 56,933 images recorded by 24 deployed AMI traps under the time-lapse based schedule (one image every ten minutes). All traps have been running a motion-detection based recording schedule as well. Past experience suggests that the number of images recorded under the motion detection-based recording schedule is 50-100 times higher than the time-lapse based recording schedule. This means that the total number of images probably exceeds 5 million images. All of these have been uploaded to the ERDA facility, but only the images from the time-lapse based schedule have been processed with the AMI pipeline so far (Fig. 2). Among the processed images, 597,676 arthropod detections have been made (Fig. 8). Following the same logic, it is possible that in the entire dataset from 2024, >50 million arthropod detections are likely. The largest number of detections (n=71,634) from a single trap was made from trap BE1 from Flanders (Belgium), while the largest average number of detections per image was made from trap FR1 from France (Table 1).

Table 1: Number of detections and images and average number of detections per images from AMI trap images recorded during the 2024 field season.

Trap ID	Detections	Images	Detections per image
BE1#2024	71634	5264	13.61
BE2#2024	40151	3027	13.26
CZ1#2024	8263	1269	6.51
CZ3#2024	9439	1146	8.24
DK2#2024	20484	1470	13.93
DK3#2024	7121	1634	4.36
FR1#2024	59585	1863	31.98
FR2#2024	4781	344	13.90
FR3#2024	5654	414	13.66
HR1#2024	64099	3120	20.54
HR2#2024	17075	1787	9.56
HR3#2024	12899	2262	5.70
IL1#2024	2089	1123	1.86
IL3#2024	13246	3270	4.05
IT1#2024	35270	2624	13.44
IT2#2024	10612	2343	4.53
IT3#2024	25398	2877	8.83
PT1#2024	18734	3086	6.07
PT3#2024	47363	4143	11.43
SE1#2024	5716	1642	3.48
SE2#2024	21097	1903	11.09
SE3#2024	51478	5518	9.33
SK1#2024	28741	2519	11.41
SK2#2024	16747	2285	7.33

The most abundant insect order detected in the processed images were Diptera with 285,448 (47.8%) of all detections. Lepidoptera made up 100,592 (16.8%) of all detections. This taxonomic group is the focus of the insect module, where species level identification is most feasible. Knowledge about invasive alien insect species is limited and so is the available training data for many taxa and the data collection in this project is facilitating better models for detection of invasive alien insect species in the future.

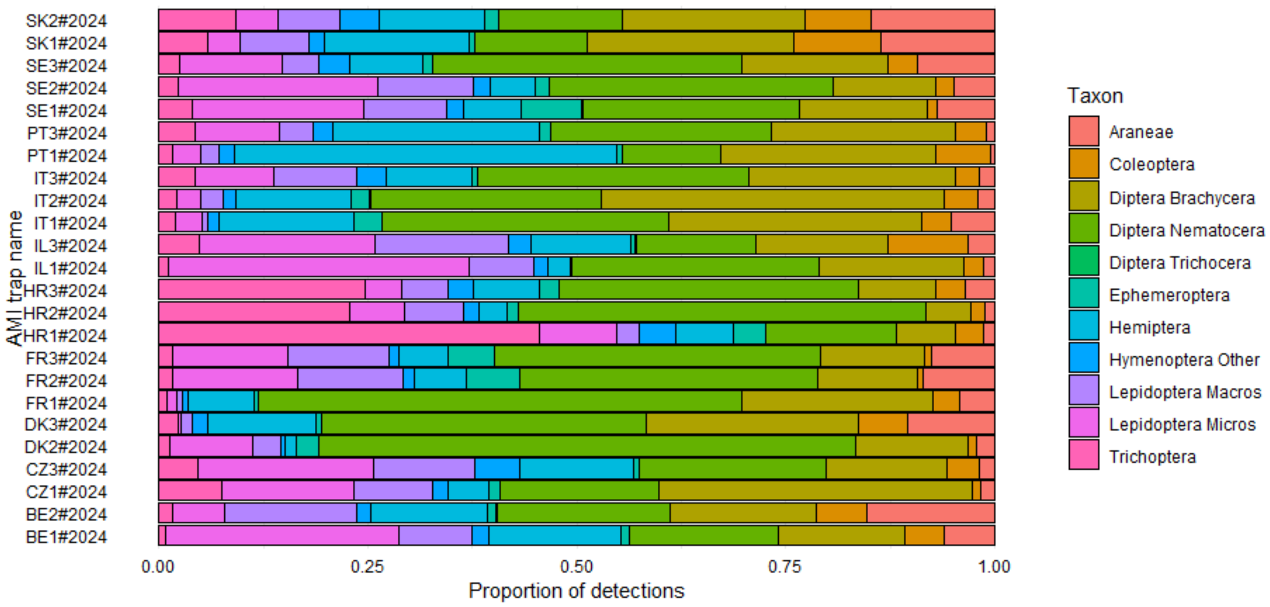


Figure 8: Proportion of detections in images from each AMI trap classified to each taxon by the broad taxon classifier. Detections classified as Diptera Tipulidae, Diptera Trichoceridae, Hymenoptera Vespidae, Opiliones, Neuroptera or as vegetation, were excluded from the graph as these classes were rare compared to the shown classes. The figure illustrates that macro moths (Lepidoptera Macros) make up a rather small proportion of the insects detected at the traps. The figure is based on a total of 597,676 detections.

During the 2023 field season, it became clear that the box tree moth (*Cydalima perspectalis*) is commonly observed in images recorded by the AMI traps and found in all partner-represented countries except for Israel. For the 2024 data, we have used our current species-level classification model for moths to examine how well our pipeline is able to identify this species. Across the nearly 600K detections, 989 were assigned to *C. perspectalis*.

We manually reviewed all predictions made to the species class *C. perspectalis.* From the verified images, it was clear that the species was detected by 15 of the 24 traps for which we have data. We found a false positive rate of 9.7% but only 0.3% were actually moths and the rest of the false predictions were mostly tiny insects that could not easily be identified even to the order level. A simple size criterion could easily filter these false detections out in the future. By further processing of the motion-detection images, we expect to be able to include more species of invasive alien insect species in the training data for future species-level classification models to accurately and efficiently detect such species when recorded at the traps.

4. Perspectives for long-term transnational monitoring

Invasive alien species are, by definition, a problem that needs to be addressed at the transnational scale (IPBES 2023). The data from each module can feed directly into a transnational database and image processing pipeline for generation of transnationally standardised and automated monitoring data.

The image recognition tools developed in this pilot are highly scalable and should be of great value at the transnational scale. The central data processing and management makes it straightforward to add partners to the project. This was evidenced by the addition of Belgium (VL O) to the pilot in 2024 and a smooth integration in the data collection, processing and management. The monitoring goals and taxonomic focus may vary among partners. For example, through the collaboration with Pl@ntNet and the application of a general-purpose plant identification model, we can easily meet the needs of the community for distribution of plant species of interest. One important perspective is that the tools can be used to monitor the spread of species already introduced to the country. Another important perspective is that models of species found in one country can be used for early detection of species as they may become introduced in other countries. This is possible, because the image recognition model does not rely on locally collected training data but lends its strength from a global collection of curated images of >50,000 plant species. Similarly, the insect classification models can cover continental scale and be trained on globally available image data. Finally, the economic burden of maintaining the digital infrastructure for databases and image-recognition will be much smaller for each active partner, if shared at a Pan-European scale. At the moment, this is made possible through the funding for the pilot. In the future, the system could be maintained through modest funding to the data management infrastructure as a component of a transnational unit e.g. under the European Biodiversity Observation Coordination Centre (EBOCC). In such a scenario, each partner will only need to maintain their equipment and take care of data collection and interpretation of results. By centralising the data management, it is also possible to build dashboards for interaction with and visualisation of the outcomes of the data analysis. This part will be further elaborated in the final stage of this pilot.

In the long-term, the image-based approaches offer highly cost-effective means to map and monitor the distribution of invasive alien species and can substantially shorten the time from an observation is made until it enters databases and decision support systems on which management decisions are taken. A dual approach is planned, where detections of invasive alien species are automatically published to the Global Biodiversity Information Facility and EASIN and also ingested in a bespoke database for efficient visualisation and synthesis of the data.

There are still some challenges on the road to implementation in transnational monitoring. One element relates to the delay caused by transfer of data from camera systems to the image recognition models and the time involved with the processing of the image data. In the final stage of the pilot, we will explore options for moving the image recognition models onto the camera systems to further reduce the time delay from observation to actionable information. Also, this pilot focuses entirely on the transport infrastructures. For managers, these are important areas for establishment of invasive alien species, but to fully map the distribution area of populations of invasive alien species, complementary approaches should also be considered e.g. drone surveys for plants. However, this will be outside the scope of the pilot project.

The pilot delivers new ways of standardised and cost-efficient monitoring of species with automated image recognition. Reliable data on the natural environment is fundamental to support science on species conservation in human-modified landscapes. The pilot results can further be used to evaluate if adapted systems could be used to monitor additional taxa of invasive alien species with image-based tools (e.g., mammals often observed as road kills, such as raccoon *Procyon lotor*; arboreal nests of the Asian hornet *Vespa velutina* in fall; fruit flies or other day-active insect species using attractants).

A number of research infrastructures and international projects can support the implementation of these tools in long-term transnational monitoring. Here, we list a few:

- The EASIN information network established by the JRC provides collated information from EU member states as well as regional initiatives such as NOBANIS and ESENIAS. None of these sites present information about monitoring invasive alien species.
- Pl@ntNet is an identification system that helps with the identification of plants through images. It is a research and a citizen science project, initially supported by Agropolis Foundation, and developed since 2009. The Pl@ntNet system works by comparing visual patterns transmitted by users via photos of entire plants or plant organs (e.g. flowers, fruits, leaves) that they seek to determine. These images are analysed and compared to an image database produced collaboratively. Pl@ntNet is governed by a consortium composed of four research organisations: CIRAD, Inria, INRAE, IRD and the Agropolis Foundation. The management of the consortium is part of the InriaSOFT program, a program to perpetuate the digital achievements of Inria and its partners.
- The AMI traps represent one of the recommended tools from the STING expert group appointed by the European Commission Joint Research Centre. Among other tasks, STING develops recommendations on novel technology for pollinator monitoring, which should be tested at European scale for moth monitoring (Potts et al. 2024).
- The Cost Action InsectAI launched in November 2023 will be engaged in the insect module regarding image analysis. In the longer term, this network can also contribute to the further development of improvements and tests of hardware and software and standards for deployments and model training as well as through the collection of additional training data.
- Horizon Europe projects MAMBO (2022-2026), OneSTOP (2025-2028) and BMD (2025-2028) will provide additional technical expertise on image recognition, invasive alien species

and early warning systems and will specifically contribute to the continued development of deep learning models.

The challenge over the coming years is to transform this pilot study into long term monitoring and to integrate it into a network of national and transnational biodiversity monitoring schemes.

Next steps

To facilitate the implementation of transnational monitoring of invasive alien plant and insect species with image-based methods, this pilot aims to estimate of the costs, refine the protocols, analysis pipelines, data infrastructure and work with the visualisation and synthesis of data into Essential Biodiversity Variables and the flow of actionable information to end users.

Specifically, a well-performing and stable infrastructure for users to collect, process and manage data is needed. The core functionality of this system is already in place, but further system development is needed to consolidate the IT infrastructure back-end and front-end to make it more user friendly and less dependent on specialised expertise. Processing of data from partners will continue and steps will be taken to estimate and minimise the costs of image analysis and data management for future users. The coordinators will also support partners in maintaining and operating their equipment and data uploads.

Secondly, reducing the costs and time-lag involved in translating field observations into actionable knowledge for decision makers is critical. This will partly be achieved by fitting a subset of the AMI traps with new mini-computers for online communication and deep learning models, so they can send occurrence data of invasive alien species in real-time to centralised servers for publishing to GBIF and EASIN and EBV datasets to the Geo-BON portal. To achieve this, we will finalise the revision of data publishing models with GBIF for CamAlien and AMI. Workshops during 2024 with GBIF and data scientists behind the Camera Trap Data Package (CamTrap DP) have revealed that further adaptations of existing standards are needed. This dialogue will be continued to reach consensus and best practice guidelines.

Thirdly, knowledge transfer around distributed, scalable IAS monitoring approaches is needed for the implementation of transnational monitoring. We will connect to partners' national networks of managers, researchers and authorities as well as EU initiatives (such as the upcoming EBOCC) and projects (such as the Horizon Europe projects MAMBO, GUARDEN, B-CUBED, Biodiversity Meets Data and OneSTOP) to identify ways the pilot project can add value and complement ongoing activities. The coordinator is already in close dialogue with the coordinators of the mentioned Horizon Europe projects. Dialogue meetings will be held with all of these projects. In addition, two in person workshops focussed on coordination and capacity building with EPAs will be organised and participation in relevant conferences and events will be organised.

References

Basille, M., Body, G., Eggermont, H., Mandon, C. and Vihervaara, P. (2023). Shared goals and priorities for biodiversity monitoring within Biodiversa+. *Biodiversa+ report*. 22pp. Available at: https://www.biodiversa.eu/wp-content/uploads/2023/06/D2.5-Priorities.pdf

Bjerge K, Nielsen JB, Sepstrup MV, Helsing-Nielsen F, Høye TT. (2021) An Automated Light Trap to Monitor Moths (Lepidoptera) Using Computer Vision-Based Tracking and Deep Learning. *Sensors*. 2021; 21(2):343. https://doi.org/10.3390/s21020343

Bubnicki, J. W., Norton, B., Baskauf, S. J., Bruce, T., Cagnacci, F., Casaer, J., Churski, M., Cromsigt, J. P., Farra, S. D., Fiderer, C., Forrester, T. D., Hendry, H., Heurich, M., Hofmeester, T. R., Jansen, P. A., Kays, R., Kuijper, D. P., Liefting, Y., Linnell, J. D., … Desmet, P. (2023). CAMTRAP DP: An open standard for the fair exchange and archiving of Camera Trap Data. *Remote Sensing in Ecology and Conservation*. https://doi.org/10.1002/rse2.374

Dyrmann M, Mortensen AK, Linneberg L, Høye TT, Bjerge K. (2021) Camera Assisted Roadside Monitoring for Invasive Alien Plant Species Using Deep Learning. *Sensors*. 21(18):6126. https://doi.org/10.3390/s21186126

IPBES Invasive Alien Species global assessment: IPBES (2023). Summary for Policymakers of the Thematic Assessment Report on Invasive Alien Species and their Control of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Roy, H. E., Pauchard, A., Stoett, P., Renard Truong, T., Bacher, S., Galil, B. S., Hulme, P. E., Ikeda, T., Sankaran, K. V., McGeoch, M. A., Meyerson, L. A., Nuñez, M. A., Ordonez, A., Rahlao, S. J., Schwindt, E., Seebens, H., Sheppard, A. W., and Vandvik, V. (eds.). IPBES secretariat, Bonn, Germany. https://doi.org/10.5281/zenodo.7430692

Potts, S., Bartomeus, I., Biesmeijer, K., Breeze, T., Casino, A., Dauber, J., Dieker, P., Hochkirch, A., Høye, T., Isaac, N., Kleijn, D., Laikre, L., Mandelik, Y., Montagna, M., Montero Castaño, A., Öckinger, E., Oteman, B., Pardo Valle, A., ...Zhang, J., (2024). Refined proposal for an EU pollinator monitoring scheme, Publications Office of the European Union. European Commission: Joint Research Centre https://data.europa.eu/doi/10.2760/2005545

Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *International Conference on Machine Learning*. https://doi.org/https://doi.org/10.48550/arXiv.1905.11946

Michelle Silva del Pozo, Guillaume Body, Gaby Rerig, Mathieu Basille. (2023). Guide on harmonising biodiversity monitoring protocols across scales. Biodiversa+ report. 60 pp. Available at: https://www.biodiversa.eu/wp-content/uploads/2023/10/Biodiversa_Harmonising-monitoring-protocols.pdf