De Olho na Mata: monitoring Atlantic Forests with drones and few-shot learning

Alexandra Aguiar Pedro[1], Farzaneh Dadrass Javan[2], Sonja Georgievska[3], Eduardo Hortal Pereira Barreto[4], Ou Ku[5], Felipe de Oliveira[6], Patricia do Prado Oliveira[7], Caroline Gevaert[8]

[1] São Paulo Municipal Green and Environment Secretariat – São Paulo, Brazil – aapedro@prefeitura.sp.gov.br
[2] University of Twente, Faculty ITC – Enschede, the Netherlands – f.dadrassjavan@utwente.nl
[3] Netherlands eScience Center – Amsterdam, the Netherlands – s.georgievska@esciencecenter.nl
[4] São Paulo Municipal Green and Environment Secretariat – São Paulo, Brazil – ehortal@prefeitura.sp.gov.br
[5] Netherlands eScience Center – Amsterdam, the Netherlands – o.ku@esciencecenter.nl
[6] São Paulo Municipal Green and Environment Secretariat – São Paulo, Brazil – felipedeoliveira@prefeitura.sp.gov.br
[7] São Paulo Municipal Green and Environment Secretariat – São Paulo, Brazil – patriciapradooliveira@prefeitura.sp.gov.br
[8] University of Twente, Faculty ITC – Enschede, the Netherlands – c.m.gevaert@utwente.nl

Keywords: forest management, few-shot learning, classification, Explainable AI, unpiloted aerial systems, Atlantic Forest

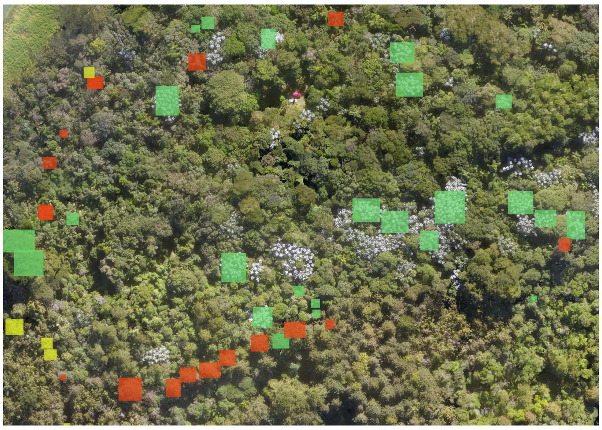
Abstract

The expansion of invasive species is a global challenge that leads to the loss of biodiversity habitat, and there are few tools to control it. In São Paulo, identification of invasive species is done through field inspections, in parts of Conservation Units and parks, making it difficult to map all tree individuals for adequate management and coping strategies. This manuscript presents a workflow that combines Unmanned Aerial Vehicles (UAVs), or drones, with Artificial Intelligence (AI) to accurately map invasive species in the Atlantic Forest. It describes best practices on how to conduct drone flights to map the forests, exponentially expanding the range of identification and efficiency in invasive tree species management. It also presents an AI workflow that uses few-shot learning and Explainable AI techniques (to guarantee transparency and understanding of the decisions made by the algorithms). Preliminary results indicate that the method obtains acceptable results in the range of 70 percent accuracy for *Archontophoenix cunninghamiana* (popular name: Seafórtia), an invasive Australian palm.

1. Introduction

The expansion of invasive species is a global challenge that leads to the loss of biodiversity habitat, and there are few tools to control it. This aligns with Sustainable Development Goal (SDG) Target 15.1, which focuses on ensuring the conservation, restoration, and sustainable use of terrestrial and inland freshwater ecosystems, and SDG Target 15.2, which emphasizes sustainable management of all types of forests, halting deforestation, restoring degraded forests, and significantly increasing afforestation and reforestation globally. At the Municipality of São Paulo, the remaining fragments of the Atlantic Forest can have the presence of native and exotic species (see Figure 1 for some examples). Here, a distinction is made between exotic species in general, and exotic species which are more concerning as they are invasive and therefore more threatening to the natural environment. The monitoring and identification of native and exotic species is done through field inspections, only in parts of Conservation Units (UCs) and Municipal Parks, making it difficult to map all individuals for adequate management and coping.

Remote sensing presents significant advantages over traditional field-based methods for forest monitoring, including scalability, multi-sensor capabilities, and cost and time efficiency (Cohen and Goward, 2004). Satellite imagery is effective for covering extensive areas. However, Unmanned Aerial Vehicles (UAVs), commonly known as drones, have recently become versatile tools for collecting high-resolution data, offering promising solutions to overcome the limitations of satellite-based data (Torresan et al, 2017). Both satellite and UAV imagery can be integrated with artificial intelligence techniques, such as machine learning, to automate information extraction, including identifying invasive species.



(a)

(b)

Figure 1. Example UAV image of the Atlantic Forest in São Paulo in Bororé Natural Park, marked by local experts: in green, native species, in yellow, exotic species and in red, exotic invasive species (a) and a close view of the invasive species Seafórtia (b).

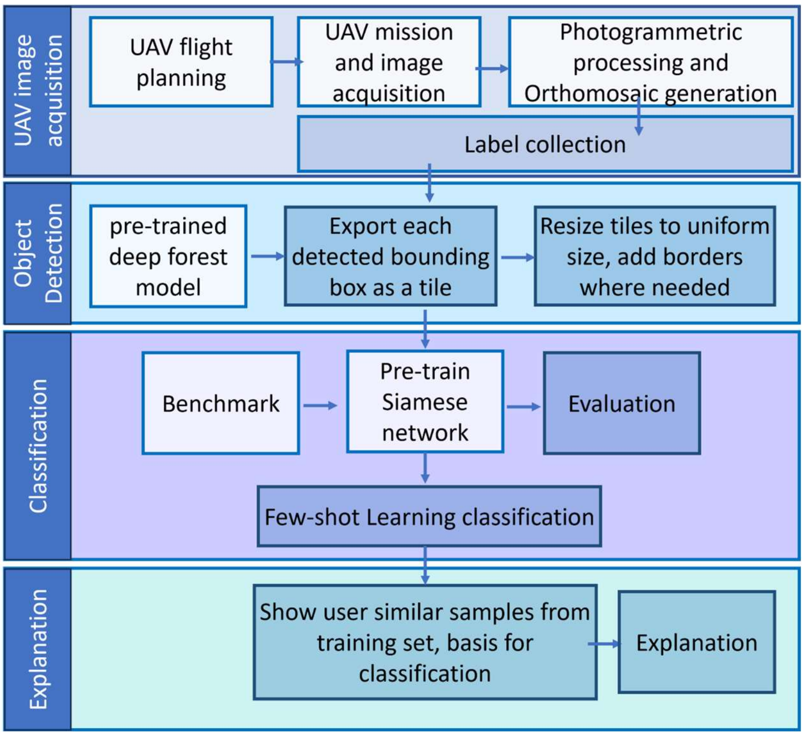


Figure 2. A schematic overview of the complete proposed workflow, including data collection with UAV and a custom classification algorithm for few-shot learning.

However, how AI algorithms can be incorporated into public policy decision-making to promote the Sustainable Development Goals in a transparent and reliable way is still a challenge, considering that the reasoning behind the workings of AI models is still difficult to understand. Furthermore, drone flights are still an expanding practice in developing countries' governments. As a result, there is little availability of images obtained by drones of Atlantic Forests, which hinders AI advancements, since a significant number of labelled images are needed to both train models and prove the accuracy of the results. This paper presents a novel workflow for municipal workers in Brazil to (1) use drones to collect imagery of forests in Brazil and (2) classify invasive and native species using a novel, Explainable, few-shot learning workflow.

2. Methodology

The methods consist of two main stages. Firstly, drone flights are conducted over a section of the Atlantic Forest to be monitored. Section 2.1 explains more about the study area for this area and Section 2.2 provides recommendations on Unpiloted Aerial System flights for forest monitoring. The second phase on using few-shot learning techniques to classify trees in the obtained imagery is described in Section 2.3. This second stage consists of: object detection, classification, and an explanation (Section 2.4). An overall overview of the workflow is provided in Figure 2.

2.1 Study area

The Atlantic Forest is a dense and tall formation, with diverse plant extracts and a high diversity of fauna and flora. However, due to urban expansion in São Paulo city, there are only a few remaining areas of this forest in different stages of conservation, most of them in some degradation process. Most of the remaining areas of the Atlantic Forest are on the edges of the city, where are located the five city’s natural parks (integral protection conservation units, which aim to preserve natural ecosystems of ecological relevance), such as the Bororé Natural Park, the area of the first case study of the project DE OLHO NA MATA. Despite being conservation areas, and thus protected from urban expansion, the remaining Atlantic Forest in these parks also suffer from the expansion of invasive exotic species over the native forest.

Invasive species compete with native species in natural environments, which can significantly alter habitats, causing the local extinction of native species and generating other ecosystem complications. To define actions that prevent the loss of biodiversity, the few experts from the Municipal Green and Environment Secretariat (SVMA), carry out field visits in the natural parks to elaborate biodiversity assessments of some existing plant formations and species, including the invasive ones. As natural parks are large areas, these assessments are carried out in only parts of the parks, that is, in a sample area. Therefore, there is not a mapping of individual invasive trees in the municipal parks, let alone outside parks or conservation units. Without knowing where the invasive species are in the municipality’s territory, it is not possible to adequately manage them and protect the native forest. In the Bororé Natural Park, the SVMA assessment found 141 individual trees of invasive exotic species. Based on this assessment, a section of this sampling with the highest concentration of identified species was selected to carry out the drone flight that generated the base image for the case study of this project, where individuals of the

https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-387-2024

Archontophoenix cunninghamiana (popular name: Seafórtia), an invasive Australian palm were identified for the first classification experiment.

2.2 Data Acquisition Considerations

While planning a UAV flight for vegetation monitoring, various elements should be accurately considered. These elements need to be carefully designed and decided upon, as the flight is primarily the main step in each study, considering factors such as cost, time, risk, and reliability. Any issues during flight or data acquisition can lead to inaccuracies in further processing steps, and a worst-case scenario may require repeating the entire mission.

In studying natural phenomena such as plants and forests, certain aspects of flight design hold special importance. For instance, environmental conditions like wind speed and direction, along with lighting conditions (due to the sun angle and any cloud coverage), must be precisely examined. Furthermore, in areas where tree height significantly varies concerning flight altitude, anticipating potential changes in image scale is crucial. Utilizing multiple flight designs with different altitudes can help maintain a constant relative height of the camera from the tree canopy or a consistent image scale. The platform should also be equipped with PPK post-processing GPS to provide accurate position of the imaging points which can later support photogrammetric processing while reducing the need for ground control points which is not easy to handle, especially in forests.

Moreover, obtaining a comprehensive overview of the targeted objects is essential considering the necessity of generating true orthophotos in such applications. Therefore, designing a grid network or at least incorporating crosslines into flight planning is recommended. Larger values for overlap and side lap (80% and 80%) are also suggested.

Ground Sampling Distance (GSD) also plays a critical role in earth observation, especially for classification applications (Anderson and Gaston, 2013). Extremely fine GSDs can result in large datasets and higher computational time and expenses. Conversely, coarser resolutions can impact the accuracy of object detection and final classification results.

2.3 Classification workflow

The classification workflow consists of (1) detecting tree crowns, (2) classifying the tree crown, and (3) providing an explanation. An overview flowchart of the methodology will be added to the camera-ready paper. The first step is object detection, which uses the pretrained Deep Forest model (Weinstein et al., 2019). This is an object detection model which was trained on aerial RGB imagery from almost 3000 hand-annotated trees and more than 400,000 self-generated trees from the National Ecological Observation Network. For this study, we utilized the model available through the repository (Weinstein et al., 2019), it was not further trained or fine-tuned with data from Bororé Natural Park. The second step is classification. A few-shot learning technique was selected, in this case a Siamese Neural Network.

Siamese networks are a type of neural network architecture designed for tasks requiring the comparison of two inputs, such as image recognition, face verification, and signature matching. Unlike traditional neural networks, Siamese networks consist of two identical subnetworks that share the same weights and parameters. These twin networks process two different inputs and produce comparable feature vectors. The core idea is to measure the similarity between these feature vectors using a distance metric, like Euclidean or cosine distance. During training, the network learns to distinguish between similar and dissimilar pairs through a contrastive or triplet loss function, optimizing to minimize the distance for similar pairs and maximize it for dissimilar ones. This unique architecture makes Siamese networks particularly effective for tasks involving similarity and ranking, where the goal is to determine whether two inputs belong to the same class or are related in some meaningful way (Bromley et al., 1993; Koch et al., 2015).

The Siamese Network utilized in this study was a small network using two 128 x 128 pixel input images with three bands, a densely connected layer, comparing both images with a feature vector with a length of 128, and a Euclidean distance function. All programming was done in Python using the Keras library.

Despite an exhaustive search, we team was unable to identify an online benchmark of image data for Atlantic Forest trees. It was therefore decided to utilize existing benchmarks of other trees and pre-trained models and few-shot learning strategies to perform the classification.

The Siamese Network was thus trained using the ReforesTree benchmark (Reiersen et al., 2022), a dataset of six tropical trees. Siamese Networks take two images and are trained to indicate whether the images are similar or not. The supervised training directs the loss function to learn characteristics from the images (in this case of trees) which are key to differentiating different types of tree species. To then move to a multi-class classification step, a test sample is run through the Siamese Network with many different labelled training samples. The result is a vector of similarity scores comparing this test image to each of these labelled samples. Samples were assigned to the class by taking the average similarity score between the test image and all of the labelled samples of a given class, and then determining which of the classes had the highest average similarity score. As a final step, the user has the option to request an explanation for each classified tree crown.

2.4 Explanation

The field of Earth Observation is starting to recognize the need for explainable machine learning workflows, also known by the term Explainable Artificial Intelligence or XAI (Gevaert 2023). This drive is one of six main research areas in Earth Observation (Tuia et al., 2021). The terms transparency, interpretability, and explainability are often used inconsistently (e.g. Roscher et al., 2020b). Transparency refers to the ease of accessing model parameters and decisions (Roscher et al., 2020; Tuia et al., 2021) and includes the broader ML system context, code availability, bias mitigation, and human rights considerations (European Commission, 2021; UNESCO, 2021). Researchers often use explainability and interpretability interchangeably to mean understanding the cause of a decision (e.g. Miller, 2019). Others distinguish between interpretability (understanding model predictions)

and explainability (linking predictions to domain knowledge) (e.g. Roscher et al., 2020; Tuia et al., 2021).
Adadi and Berrada (2018) identify four motivations for XAI: to justify, to control, to improve, and to discover. Justification addresses societal concerns about algorithmic transparency regarding a single decision. Control helps identify and debug errors. Improvement aims to enhance models, and discovery helps scientists uncover new knowledge. From a societal perspective, there is also an increasing call for explainable algorithms – for example for high-risk use cases in the recently adopted EU AI Act (European Commission, 2021) and the UNESCO recommendations for AI (UNESCO, 2021). Yet social scientists also identify a mismatch between the XAI methods developed by technical researchers and the way that human reasoning works. For example, Miller (2019) identifies three main gaps: human explanations are typically contrastive, selective by focusing on one or two key reasons rather than all possible reasons and are often part of social interactions to convey knowledge. Additionally, Miller notes that explaining the causes of predictions is more helpful for human understanding than simply providing statistical probabilities.

Gevaert (2023) provides a framework on how different types of XAI methods can be used by different persons and to different ends. In the context of this manuscript, explainability is important to control and improve the classification model used to identify the tree species in the UAV imagery. As such, we will add a post-hoc, model-specific, local explanation to the output classification. The explanation which will accompany the classification will be inspired by the limitations given by Miller (2019), having elements of contrastivity and be based on a few key reasonings.

The explanation method in this workflow visually presents the user with the most similar labelled samples from each class. So, from the list of similarity scores, it selects the most similar labelled sample per class and the corresponding similarity score. This helps the user understand why the classifier made their decision as well as get an idea of how well the classifier is working in a few-shot situation.

3. Results

3.1 Regarding data acquisition

Taking into account all the recommendations for flight planning, the primary components of the flight plan in this study are listed in Table 1 and a picture of the low-cost UAV, a DJI Mavric pro, used to collect the imagery is in Figure 3. Figure 4 also contains a sample of one of the different missions planned through flight planning.

Feature	Description
System	Mavic pro
Sensor	FC220, RGB
Focal length	5 mm
Resolution	4000, 3000
Flight height	120 m

Table 1. Flight mission planning and applied system elements.

Figure 3. A visualization of the low-cost DJI Mavric used to collect imagery for the current study (Source: https://www.drones.nl/drones/dji-mavic-2-zoom/specs).

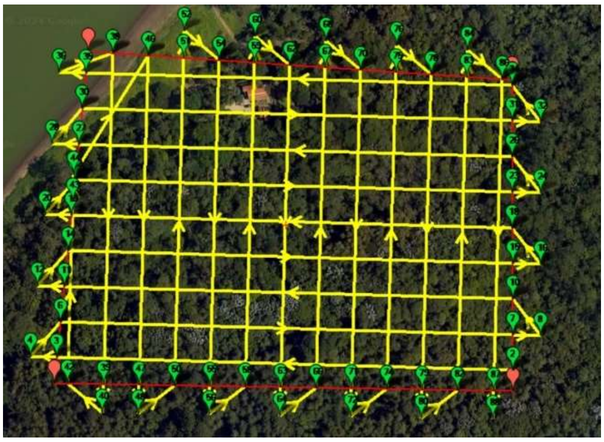


Figure 4. A schematic overview of the flight plan employed for this study. Yellow lines indicate the flight path.

(a)

(b)

Figure 5. The final generated orthomosaic covering a part (21.1 ha) of the Atlantic Forest in São Paulo (a) and a sample image with a snapshot showing the detail captured at full resolution (b).

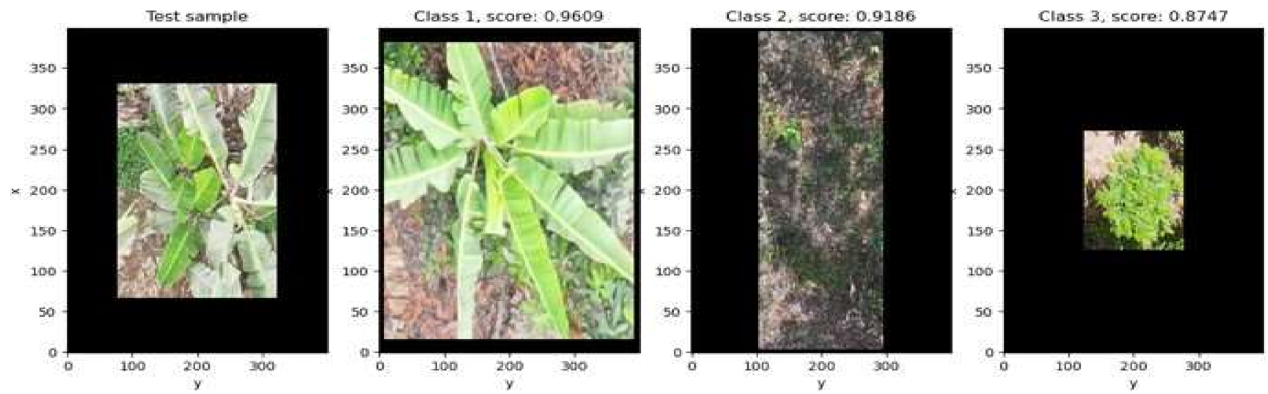


Figure 6. An example explanation with a unlabeled test sample (left) and the most similar training images per class. The images show the closest example of each of the three classes to the test sample, the scores indicate the similarity score between 0 and 1 that is the output of the Siamese Network where 1 is the most similar. Note that the test sample is visually much more similar to the selected sample of class 1 than class 2 and 3 – giving the user confidence that in this case the classifier is likely performing well.

To process the captured images Agisoft Metashape software was employed to produce a Digital Surface Model (DSM) and orthophoto which generally involves a few key processing steps. First, the images are aligned using key points extraction and through structure from motion (SFM) processing (Crandall et al., 2011). As the images are precisely georeferenced, no Ground Control Points (GCPs) are needed. Afterward, the dense point cloud is generated to construct the model geometry, followed by the creation of a 3D textured mesh. Finally, the point cloud is interpolated to generate a DSM, and with the created DSM, the orthomosaic of the region is produced. Figure 5 presents the final generated orthomosaic overlaid on satellite imagery with a close shot of a sample image acquired through the flight.

3.2 Regarding classification and explanations

Results of the study indicate that the classification accuracy obtains an average class accuracy of 0.71 (see Table 2). The first three classes, banana, cacao, and fruit trees, originate from the benchmark used to train the model (ReforesTree). Here we see that the accuracy of banana and fruit trees is very high. banana in particular was easier for the Siamese network to distinguish from the other species. Note that while training the model, no examples from the Seafórtia class were used in the Siamese network. It was only used for inference. Although there is still room for improvement in the classification results, note that the accuracy of the unseen Seafórtia class is similar to that of cacao. It is still thus within the range of accuracy as the classes which the Siamese network has been pretrained on.

Class	Accuracy
Banana	1.0
Cacao	0.57
Fruit trees	0.71
Seafórtia	0.57
Overall	**0.71**

Table 2: Results of the classification.

An example of an explanation provided by the Explainable AI component can be found in Figure 6. The explanation is a local explanation, which means that it explains the classification result of a single classification (as opposed to global models which explain the workings of the model over all of the testing samples). See how the test image which represents a sample of banana tree (on the left) compares to the three closest images from the classes banana (class 1), cacao (class 2), and fruit trees (class 3). The similarity scores at the top of the images also indicate how similar the trained Siamese Network considered the image to be to the test sample. The image considered closest by the algorithm (the second image from the left, with a similarity score of 0.9609) is also visually the most similar to the test image. This gives the user confidence that this image was correctly classified. In this way, the explanation can be used to give a user a visual indication of which training samples are the closest to the test image in question according to the algorithm. This helps give confidence in correct predictions and can perhaps in the future also be combined with active learning techniques – where a user may identify wrongly labelled samples or get a better understanding of the reasons why a class is difficult to classify, and obtain new labels more efficiently.

4. Conclusions

Satisfying results in the classification of the Australian palm *Archontophoenix cunninghamiana* were obtained, which means that the proposed workflow was able to identify individual trees with similarities with what we as human beings sought to identify. The innovation with Explainable AI is that when we provide the algorithm with a new tree to classify, it is possible to see which images it considers as a basis for previous comparison and this immediately allows it to evaluate its performance. Such visualizations can help identify mislabeled data or give the user an indication of which classes require additional labelled sampling if the user repeatedly encounters confusion between two classes.

In future work, as more labeled data from the study area becomes available, we plan to fine-tune or re-train our models to better capture the unique characteristics of the local environment. Additionally, we recognize the importance of innovative modifications to the networks and will explore such enhancements in subsequent iterations. Thus, further research will focus on: further improving the accuracy of the classification model, training the model to recognize more tree species, and training employees of SVMA to utilize the tool.

5. Acknowledgements

This publication is part of the project "Bridging the gap between artificial intelligence and society: Developing responsible and viable solutions for geospatial data" (with project no. 18091) of the research program Talent Program Veni 2020, which is (partly) financed by the Dutch Research Council (NWO).

References

Adadi, A. and Berrada, M., 2018. Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, pp.52138-52160. https://doi.org/10.1109/ACCESS.2018.2870052.

Anderson, K., Gaston, K. J., 2013. Lightweight unmanned aerial vehicles will revolutionize spatial ecology. *Frontiers in Ecology and the Environment*, 11(3), 138–146. https://doi.org/10.1890/120150.

Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., & Shah, R. (1993). Signature Verification Using a "Siamese" Time Delay Neural Network. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(4), 669-688. https://doi.org/10.1142/S0218001493000339.

Cohen, W.B. and Goward, S.N., 2004. Landsat's role in ecological applications of remote sensing. *Bioscience*, 54(6), pp.535-545. https://doi.org/10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2.

Crandall, D., Owens, A., Snavely, N., & Huttenlocher, D. (2011, June). Discrete-continuous optimization for large-scale structure from motion. In CVPR 2011 (pp. 3001-3008). IEEE. https://doi.org/10.1109/CVPR.2011.5995626.

European Commission Directorate-General for Communications Networks, C., and T., 2021. Proposal for a Regulation of the European Parliament and of the council Laying Down Harmonised Rules on ARtificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

Gevaert, C.M., 2022. Explainable AI for earth observation: A review including societal and regulatory perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 112, p.102869. https://doi.org/10.1016/j.jag.2022.102869.

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese Neural Networks for One-shot Image Recognition. *Proceedings of the 32nd International Conference on Machine Learning (ICML)*. https://doi.org/10.1007/978-1-0716-0826-5_3.

Miller, T., 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, pp.1-38. https://doi.org/10.1016/j.artint.2018.07.007.

Reiersen, G., Dao, D., Luitjens, B., Klemmer, K., Amara, K., Steinegger, A., Zhang, C., Zhu, X., 2022. ReforesTree: A Dataset for Estimating Tropical Forest Carbon Stock with Deep Learning and Aerial Imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(11), 12119-12125. http://ojs.aaai.org/index.php/AAAI/article/view/21471.

Roscher, R., Bohn, B., Duarte, M.F. and Garcke, J., 2020. Explainable machine learning for scientific insights and discoveries. *IEEE Access*, 8, pp.42200-42216. https://doi.org/10.1109/ACCESS.2020.2976199.

Torresan, C., Berton, A., Carotenuto, F., Di Gennaro, S.F., Gioli, B., Matese, A., Miglietta, F., Vagnoli, C., Zaldei, A. and Wallace, L., 2017. Forestry applications of UAVs in Europe: A review. *International Journal of Remote Sensing*, 38(8-10), pp.2427-2447. https://doi.org/10.1080/01431161.2016.1252477.

Tuia, D., Roscher, R., Wegner, J.D., Jacobs, N., Zhu, X. and Camps-Valls, G., 2021. Toward a collective agenda on ai for earth science data analysis. *IEEE Geoscience and Remote Sensing Magazine*, 9(2), pp.88-104. https://doi.org/10.1109/MGRS.2020.3043504.

UNESCO, D., 2021. Preliminary report on the first draft of the recommendation on the ethics of artificial intelligence. *United Nations Educational, Scientific and Cultural Organization*. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000380455

Weinstein, B. G., Marconi, S., Bohlman, S., Zare, A., White, E., 2019. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*, 11(11), 1309. https://doi.org/10.3390/rs11111309.