Goals and Stakeholder Involvement in XAI for Remote Sensing: A Structured Literature Review

Carolin Leluschko(✉) and Christoph Tholen

German Research Center for Artificial Intelligence RG Marine Perception, Oldenburg, Germany
{carolin.leluschko,christoph.tholen}@dfki.de

Abstract. A currently upcoming direction in the research of explainable artificial intelligence (XAI) is focusing on the involvement of stakeholders to achieve human-centered explanations. This work conducts a structured literature review to asses the current state of stakeholder involvement when applying XAI methods to remotely sensed image data. Additionally it is assessed, which goals are pursued for integrating explainability. The results show that there is no intentional stakeholder involvement. The majority of work is focused on improving the models performance and gaining insights into the models internal properties, which mostly benefits developers. Closing, future research directions, that emerged from the results of this work, are highlighted.

Keywords: Explainable AI · Remote Sensing · Aerial Imagery

1 Introduction

The use of artificial intelligence is getting more prominent in a variety of domains which is why the need for interpreting and understanding model predictions is of utmost importance. A potential future application of artificial intelligence (AI) methods is the assessment of marine litter using airborne based remote sensing. This is the goal, the PlasticObs+ project is aiming for. In this project, different AI systems are developed utilizing aerial images with varying resolutions from different sensors [30]. The assessment of plastic litter is of importance for different stakeholders, for instance local governments, NGO's or members of the society. In this project, the resulting information is provided using a geographic information system (GIS) which builds the interface to the stakeholder. The AI systems are developed using neural network which makes the overall system a black box, not providing any information about the decision making process to the stakeholders. However, explanations are needed to aid the stakeholders' informed decision making. Here, explainable AI (XAI) [9] could be a valuable asset to the application.

The field of XAI has emerged in recent years with intense research being conducted to open the "black-box" nature of the state-of-the-art machine learning

M. Bramer and F. Stahl (Eds.): SGAI 2023, LNAI 14381, pp. 519–525, 2023.
https://doi.org/10.1007/978-3-031-47994-6_47

systems. While early research focused on the development of methods for delivering explanations of a models behaviour and its output, the focus is now shifting towards human-centered XAI [19]. Not all kinds of explanations are relevant and useful for all stakeholders of an AI system, therefore the need for individual explanations is arising across domains. In this work, a structured literature review is conducted to assess the current state of integrating XAI methods into the domain of remotely sensed aerial images with a focus on goals and stakeholder involvement. In particular the following two research questions will be investigated: (1) are stakeholders involved in applying XAI to remote sensing data, and (2) what goals are pursued when applying XAI to aerial imagery?

Different tools and algorithms that make the machine learning models more interpretable and explainable resulted from the intense research in the area of XAI. Many authors have categorized such algorithms and tools [2,7,34] for instance into model agnostic or model specific methods, their output formats or whether they provide global or local explanations.

Arrieta et al. [3] categorize and describe different general goals that are targeted with XAI methods. In addition to those goals, the authors gave an overview of stakeholders involved in AI systems which are *domain experts/users of the model, regulatory entities, managers, developers, and people affected by model decisions.* Considering different stakeholders during the development of XAI methods and when integrating explainability frameworks into deployed AI systems is a topic which is focused on in the research area of human-centered XAI [19]. The different stakeholders of an AI system have different backgrounds and agendas, which is why one single explanation may not be beneficial or effective for every person interacting with the system. In addition to that, the evaluation of XAI methods is still an ongoing open research question. Hoffmann et al. [11] discussed different user-centered evaluation measurements that can be applied to get a quality measure of XAI methods.

This work is focusing on XAI methods in the application domain of remote sensing. In order to gain more information and insights about the earth's surface, remotely sensed data from different platforms, e.g. ground, aerial, or satellites, is used [28]. In addition to different platforms, different sensors are used to receive diverse characteristics about the regarded area. These sensors include, amongst others, hyperspectral sensors, visual imaging sensors and LiDAR sensors [31]. Remote sensing applications have been used in various fields, including marine pollution monitoring [38], plastic waste assessment [35], or the mapping of earthquake-induced building damage [23]. With the vast progress and intense research on artificial intelligence, applying machine learning to remote sensing applications became more and more common [21].

2 Methodology

To answer the research questions posed, databases relevant to computer science were searched using specific keyphrases applied to the content of the full paper. The literature search was conducted in the databases Association for Computing Machinery Digital Library (ACM) (31 results), Science Direct (133 results), IEEE Xplore (30 results) and Web of Science (55 results). In addition to the simple keyphrase search in the databases mentioned above, a forward and backward search was performed. The keyword search was conducted using a combination of *("Explainable AI" OR XAI) AND ("remote sensing" OR "aerial image").* Literature published and indexed before 12th of July 2023 was considered.

The resulting papers from the database search were examined for relevance by reviewing the content of the paper. Papers that covered the topics of Explainable AI in combination with remote sensing data were shortlisted. The remaining literature was critically assessed to determine whether or not they are relevant to answer the research questions. In order to do that, two main criteria regarding the content are applied. The potentially included literature has to be an actual application of one or multiple methods of XAI to remote sensing data. In addition, the data that the methods are applied to, need to be aerial image data in the domain of remote sensing. In this study, there are no restrictions regarding the origin, for instance satellite or UAV images, of the data.

The final selection of literature was investigated under certain criteria targeting the posed research question of this literature review. It was investigated if stakeholders were considered when applying XAI methods to the application. Building on that, it was assessed which groups of stakeholder were addressed or even intentionally targeted. Derived from a subset of the goals of applying explainability methods in [3], the intentions the authors in the reviewed literature had, when using XAI methods in their work, were assessed, categorized into *informativeness, trustworthiness, model performance and causality.* Abbreviated from [3], informativeness regards acquiring information about the internal operations of the system, trustworthiness targets the confidence in the models output and causality refers to finding relations among the dataset. Model performance is an additional considered goal which means that the authors intend to improve the models performance by applying XAI methods. Apart from the intentions regarding the benefit of adding explainability methods, the selected literature was searched regarding the evaluation of the applied XAI methods. The analysis was conducted by one reviewer.

3 Results and Discussion

The first selection of literature lead to 52 shortlisted publications which covered the fields of remote sensing and XAI. From this first selection, all papers which are not based on imagery as well as papers which did not apply XAI methods to the imagery data were excluded. The remaining literature which was further analyzed consisted of 25 publications. It can be observed that 72% of the publications address the goal of informativeness, while 44% fall into the category of trustworthiness. The goals of model performance and causality were equally present with 24% and 20% respectively.

The results show that the majority of reviewed literature fit into the category of *informativeness* as the pursued goal for utilizing explainability methods. This consisted, in many cases, in analyzing which data attributes contributed the most to a models output [36], finding a feature importance metric [12] and gaining information about individual model layers [10]. In the case of informativeness, there was no specific stakeholder standing out to be the one benefiting from the findings the most. The evaluation of whether or not the integration of XAI methods helped achieving the goal of informativeness happened in a more qualitative way where the assessed information were presented as additional findings.

In case of *trustworthiness*, the literature in this category tried to achieve a (mostly visual) explanation on what the model is focusing on in the input data to assess if this is in line with the authors expectations of what is important [5]. Another finding in the category of trustworthiness is, that there was no evaluation of whether or not the shown explanations actually helped improving the trust in the model for the various stakeholders.

Another prominent goal for applying explainable AI was found to be the improvement of *model performance*. This goal is closely related to the goal of informativeness in terms of finding the most relevant features or those parts of a model, that are less important for a models decision [1]. In the case of model performance, however, this information was then used to improve the models performance by, for instance, training a scarcer model [4]. The evaluation, if utilizing explainability tools helped to improve model performance, was performed using common metrics for model performance such as accuracy or inference and training time. This category, in most cases, took the developers into account as they are usually the ones evaluating the models performance.

The literature categorized into the goal of *causality* applied XAI to gain more insight into relations among the input features, the model was trained with. In case of remotely sensed images, that information consisted, for instance, in finding different combinations of spectral bands to understand which information is the most influential for certain decisions [25].

During the reviewing process of the selected literature, there was neither any mentioning of specifically targeted stakeholders nor involvement of external stakeholders. The shown results can therefore not be split into the categories of stakeholders introduced in Sect. 2 but rather into the categories of no stakeholders or exclusively developers which were, to some extend, involved when applying the XAI methods. This lead to 32 % of the publications being categorized into involving the developer as stakeholder. This was the first important finding, that there is no active involvement of different stakeholders when applying XAI methods to remotely sensed images. The applications can therefore not be considered human-centered. The results of both, stakeholders and goal analysis are summarized in Table 1.

After reviewing and categorizing the selected literature, there are a few things that are worth mentioning. First, there are some similarities between the different goals, which, in combination with the fact that the authors, in most cases, did not specifically state an intended goal, made the process of categorizing the literature into pursued goals less straight forward. In addition, none of the reviewed literature involved any external stakeholders which is why the question of stakeholder involvement can only indirectly be answered. As it is usually the

Table 1. Resulting literature categorized into stakeholders and goals.

	Category	References
Stakeholder	Developer	[1,6,8,10,12,23,26,37]
	None	[4,5,15,17,18,20,24,25,29,33,36] [13,14,16,22,27,32]
Goal	Informativeness	[1,4,6,12,15,17,24,33,36,37] [8,10,14,16,18,20,23,29]
	Trustworthiness	[5,6,8,13,14,16,17,22,23,27,29]
	Model Performance	[1,4,10,12,26,37]
	Causality	[18,25,29,32,36]

model developers that are participating in the scientific publications, this category of stakeholders was the only one that was, to some extend, involved in applying the XAI methods. This category was, however, only considered if the developers did actually benefit from applying XAI e.g. by improving the model performance, or if they contributed to an evaluation of the methods.

4 Conclusion and Future Work

The conducted literature review on goals and stakeholder involvement in XAI for remote sensing imagery revealed that stakeholders are not yet an actively considered part when applying and evaluating XAI methods. The only stakeholders that are indirectly involved to this state are the model developers. Another finding was, that the goals, the authors had in appyling XAI methods were mostly not clearly stated or evaluated, especially when the general goal of improving trust in the model was among the motivations of their work. Future research should take up on that in involving stakeholders to 1) define goals that are pursued when providing model explanations and to 2) evaluate the goals and the XAI methods from the different stakeholders perspectives. In doing this, the application domain of remote sensing could benefit even more from the opportunities that come with machine learning by providing user-centric explanations.

Acknowledgements. Funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) based on a resolution of the German Bundestag (Grant No. 67KI21014A).

References

1. Abdollahi, A., Pradhan, B.: Urban vegetation mapping from aerial imagery using explainable AI (XAI). Sensors **21**(14), 4738 (2021)
2. Angelov, P.P., et al.: Explainable artificial intelligence: an analytical review. WIREs Data Min. Knowl. Discov. **11**(5), e1424 (2021)

3. Arrieta, A.B., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion **58**, 82–115 (2019)
4. Burgueño, A.M., et al.: Scalable approach for high-resolution land cover: a case study in the Mediterranean Basin. J. Big Data **10**(1), 91 (2023)
5. Carneiro, G.A., et al.: Segmentation as a preprocessing tool for automatic grapevine classification. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 6053–6056 (2022). ISSN: 2153–7003
6. Chen, L., et al.: Towards transparent deep learning for surface water detection from SAR imagery. Int. J. Appl. Earth Obs. Geoinf. **118**, 103287 (2023)
7. Das, A., Rad, P.: Opportunities and challenges in explainable artificial intelligence (XAI): a survey (2020)
8. Feng, J., et al.: Bidirectional flow decision tree for reliable remote sensing image scene classification. Remote Sens. **14**(16), 3943 (2022)
9. Gohel, P., et al.: Explainable AI: current status and future directions (2021)
10. Guo, X., et al.: Network pruning for remote sensing images classification based on interpretable CNNs. IEEE Trans. Geosci. Remote Sens. **60**, 1–15 (2022)
11. Hoffman, R.R., et al.: Metrics for explainable AI: challenges and prospects (2019)
12. Hosseiny, B., et al.: Urban land use and land cover classification with interpretable machine learning - A case study using Sentinel-2 and auxiliary data. Remote Sens. Appl.: Soc. Environ. **28**, 100843 (2022)
13. Huang, X., et al.: Better visual interpretation for remote sensing scene classification. IEEE Geosci. Remote Sens. Lett. **19**, 1–5 (2022)
14. Ishikawa, S.N., et al.: Example-based explainable AI and its application for remote sensing image classification. Int. J. Appl. Earth Obs. Geoinf. **118**, 103215 (2023)
15. Jeon, M., et al.: Recursive visual explanations mediation scheme based on dropattention model with multiple episodes pool. IEEE Access **11**, 4306–4321 (2023)
16. Kakogeorgiou, I., Karantzalos, K.: Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. Int. J. Appl. Earth Obs. Geoinf. **103**, 102520 (2021)
17. Kawauchi, H., Fuse, T.: SHAP-based interpretable object detection method for satellite imagery. Remote Sens. **14**(9), 1970 (2022)
18. Levering, A., et al.: Liveability from above: understanding quality of life with overhead imagery and deep neural networks. In: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 2094–2097 (2021). ISSN: 2153–7003
19. Liao, Q.V., Varshney, K.R.: Human-centered explainable AI (XAI): from algorithms to user experiences (2022). arXiv:2110.10790
20. Luo, R., et al.: Glassboxing deep learning to enhance aircraft detection from SAR imagery. Remote Sens. **13**(18), 3650 (2021)
21. Ma, L., et al.: Deep learning in remote sensing applications: a meta-analysis and review. ISPRS J. Photogramm. Remote. Sens. **152**, 166–177 (2019)
22. Marvasti-Zadeh, S.M., et al.: Crown-CAM: interpretable visual explanations for tree crown detection in aerial images. IEEE Geosci. Remote Sens. Lett. **20**, 1–5 (2023)
23. Matin, S.S., Pradhan, B.: Earthquake-induced building-damage mapping using explainable AI (XAI). Sensors **21**(13), 4489 (2021)
24. Moradi, L., et al.: On the use of XAI for CNN model interpretation: a remote sensing case study. In: 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1–5 (2022)
25. Saeidi, V., et al.: Water depth estimation from Sentinel-2 imagery using advanced machine learning methods and explainable artificial intelligence. Geomat. Nat. Haz. Risk **14**(1), 2225691 (2023)

26. Seydi, S.T., et al.: BDD-Net+: a building damage detection framework based on modified coat-net. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. **16**, 4232–4247 (2023)
27. Su, S., et al.: Explainable analysis of deep learning methods for sar image classification. In: IGARSS 2022–2022 IEEE International Geoscience and Remote Sensing Symposium, pp. 2570–2573 (2022). ISSN: 2153–7003
28. Sugumaran, R., et al.: Processing remote-sensing data in cloud computing environments (2015)
29. Temenos, A., et al.: Interpretable deep learning framework for land use and land cover classification in remote sensing using SHAP. IEEE Geosci. Remote Sens. Lett. **20**, 1–5 (2023)
30. Tholen, C., et al.: Machine learning on multisensor data from airborne remote sensing to monitor plastic litter in oceans and rivers (plasticobs+). In: OCEANS 2023 Limerick. OCEANS MTS/IEEE Conference (OCEANS-2023), 5–8 June, Limerick, Ireland, pp. 1–7. IEEE (2023)
31. Toth, C., Jóźków, G.: Remote sensing platforms and sensors: a survey. ISPRS J. Photogramm. Remote. Sens. **115**, 22–36 (2016)
32. Valdés, J.J., Pou, A.: A machine learning - explainable AI approach to tropospheric dynamics analysis using Water Vapor meteosat images. In: 2021 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 1–8 (2021)
33. Vasu, B., Savakis, A.: Resilience and plasticity of deep network interpretations for aerial imagery. IEEE Access **8**, 127491–127506 (2020)
34. Vilone, G., Longo, L.: Classification of explainable artificial intelligence methods through their output formats. Mach. Learn. Knowl. Extr. **3**(3), 615–661 (2021)
35. Wolf, M., et al.: Machine learning for aquatic plastic litter detection, classification and quantification (aplastic-q). Environ. Res. Lett. (ERL) **15**(11), 1–14 (2020)
36. Woo Kim, Y., et al.: Validity evaluation of a machine-learning model for chlorophyll a retrieval using Sentinel-2 from inland and coastal waters. Ecol. Ind. **137**, 108737 (2022)
37. Zaryabi, H., et al.: Unboxing the black box of attention mechanisms in remote sensing big data using XAI. Remote Sens. **14**(24), 6254 (2022)
38. Zielinski, O., et al.: Detecting marine hazardous substances and organisms: sensors for pollutants, toxins, and pathogens. Ocean Sci. **5**(3), 329–349 (2009)