# Recent Trends, Challenges, and Limitations of Explainable AI in Remote Sensing

Adrian Höhl*, Ivica Obadic*
Technical University of Munich (TUM)
*share the first-authorship

Miguel-Ángel Fernández-Torres
Universitat de València (UV)

Dario Oliveira
Getulio Vargas Foundation, Brazil

Xiao Xiang Zhu
Technical University of Munich (TUM)

## Abstract

*Training deep learning models on remote sensing imagery is an increasingly popular approach for addressing pressing challenges related to urbanization, extreme weather events, food security, deforestation, or poverty reduction. Although explainable AI is getting more frequently utilized to uncover the workings of these models, a comprehensive summary of how the fundamental challenges in remote sensing are tackled by explainable AI is still missing. By conducting a scoping review, we identify the current works and key trends in the field. Next, we relate them to recent developments and challenges in remote sensing and explainable AI. By doing so, we also point to novel strategies and promising research directions, such as the work on self-interpretable deep learning models and explanation evaluation.*

## 1. Introduction

By providing frequent views of the earth from above, earth observation (EO) data plays a key role in addressing the Sustainable Development Goals [18]. In recent years, the vast amounts of generated satellite imagery have empowered the development of novel deep-learning approaches designed to tackle these challenges. While these approaches have shown outstanding performance in many of these problems [9, 38, 65], they are often complex and lack the interpretability and explanation of their decisions. This represents a challenge for EO applications because understanding the model's functioning and visualizing the interpretations for analysis is crucial [50], as it allows practitioners to gain scientific insights, discover biases, assess trustworthiness and fairness for policy decisions, and to debug and improve a model. Hence, explainable AI (xAI) emerges as a promising approach to tackle the above-mentioned challenges with observational data [57].

Yet, the application of the popular xAI methods from computer vision to remote sensing (RS) is not straightforward as RS imagery introduces a unique modality in computer vision having different properties than natural images [51]. First, images are captured from above. This perspective comes with unique scales, resolutions, and shadows. Second, RS captures images in other electromagnetic spectra, apart from the usual Red-Green-Blue (RGB) channels. Third, usual RGB cameras are primarily passive, while RS can be active, which changes properties like the radar shadows, foreshortening, layover, elevation displacement, and speckle effects [33]. Several reviews in the literature already investigate the application of xAI in EO [20, 23, 30, 52, 58]. The pioneering work of [52] categorizes the identified studies in xAI according to the general challenges in bio- and geosciences. Similarly, challenges related to geographic data, xAI computation, and geosocial issues are pointed out in [58]. Further, [20] sheds light on the existing works and addresses the xAI usage from a regulatory and societal perspective, discussing the requirements and type of explanations needed in EO from policy and regulation. In related studies, [30] discusses the usage of xAI according to the stakeholder goals while [23] focuses on the xAI approaches used for poverty mapping. Although these reviews refer to novel approaches in the field, they are not based on a broad literature database required to discern the key trends and the state-of-the-art in the field. Additionally, they do not reflect on the limitations of the popular approaches to tackle interpretability challenges in RS nor relate how the recent developments of xAI are integrated into RS.

We address these issues by conducting a systematic literature search for approaches from xAI in RS in three commonly used literature databases, namely Scopus, Springer, and IEEE. The search query (see Listing 1) consists of two major parts: keywords related to xAI and keywords related to EO.

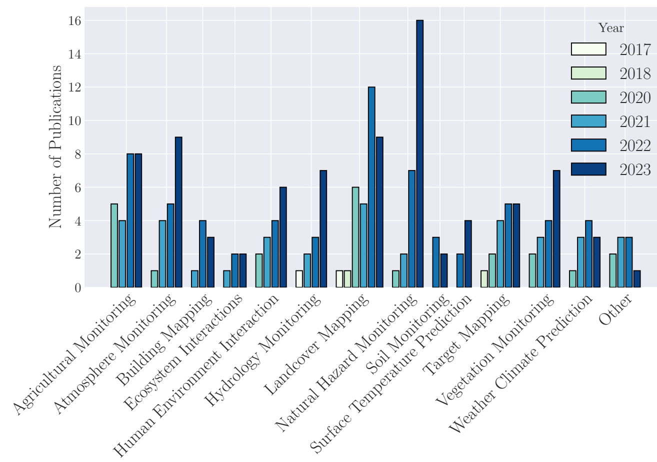
```
[earth observation OR remote sensing OR
   earth science OR ((satellite OR aerial
   OR airborne OR spaceborne OR radar) AND
   (image OR data)) OR LiDAR OR SAR OR UAV]
                  AND
[xai OR ((interpret* OR explain*) AND (deep
   learning OR dl OR machine learning OR ml
   OR artificial intelligence OR ai OR
   model))]
```

Listing 1. Search query

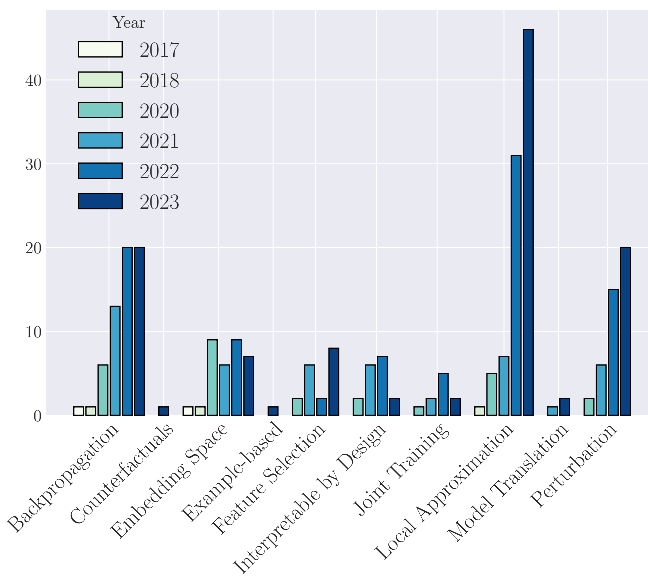
We considered papers discussing RS in EO, namely RS sensors mounted to aerial vehicles and satellites, as well as their high-level and compound products. Publications published before 2017, reviews, and short conference papers were excluded. In summary, the search resulted in 1075 papers, which were filtered in three steps: (i) removing duplicates, conference abstracts, and reports, (ii) screening through the abstract, and (iii) screening through the full text, which left us with 964, 357 and 147 papers, respectively. These 147 papers were accompanied by 60 papers we had in our libraries. In the rest of this review, we summarize these works according to the EO task, the xAI family, and the evaluation procedure. In the next section, we reveal the key trends emerging from xAI in RS. Next, we use these trends to identify the interpretability challenges and the promising research directions based on the recent developments in these fields.

## 2. Recent Trends of xAI in RS

Figure 1 gives a broad overview of the development and evolution of the xAI in RS field in the last years. Concretely, Figure 1a illustrates varying trends in the EO tasks, with vegetation, atmosphere, and natural hazard monitoring recently getting more attention. Meanwhile, despite their prominence, the xAI applications in landcover mapping and agricultural monitoring have stagnated in the last year. This suggests that the initial applications of xAI focused on traditional EO tasks. Indeed, landcover mapping is one of the most established EO tasks, with a variety of benchmark datasets and models. In contrast, the latest studies extensively explore xAI for other nowadays critical problems that usually lack established datasets and models. This interconnects to the recently elevated usage of local approximation (e.g., SHAP [36]) and perturbation approaches (Figure 1b). Being model-agnostic, these approaches enable a simple framework for initial assessments of the black-box models used for the recently explored EO tasks. Particularly, SHAP is by far the most utilized xAI approach used in 38% of the publications, with extremely high usage being noted for the last two years, whereas previously it had been rarely applied [20]. Other frequently used xAI approaches are backpropagation methods like Class Activation Mapping (CAM) [64] that are often applied to landcover and target mapping studies.




(a) EO tasks (see Appendix 1) over time.




(b) xAI categories over time. Based on [49].

Figure 1. The number of publications in the EO tasks and xAI categories over time.

## 3. Current Challenges in Remote Sensing and Explainable AI

### 3.1. Remote Sensing properties

The highly utilized xAI approaches mentioned in the previous section typically rely on assumptions that might not adhere to the RS image properties. For example, local approximation and perturbation approaches typically assume feature independence [43] or the backpropagation methods are designed to operate on natural imagery. These assumptions are often violated in RS data due to the presence of scale, geographical relationships, and temporal dependencies. Figure 2 illustrates that recently, there have been an increasing number of works that adapt the existing xAI methods to address these specific challenges of RS. Class activation map approaches, particularly Grad-CAM, are the most frequently adapted approach in RS. Other recently adapted approaches are the feed-forward, attention-based, and prototype deep neural networks, as well as permutation feature importance (PFI) and Generalized Linear Models (GLMs).

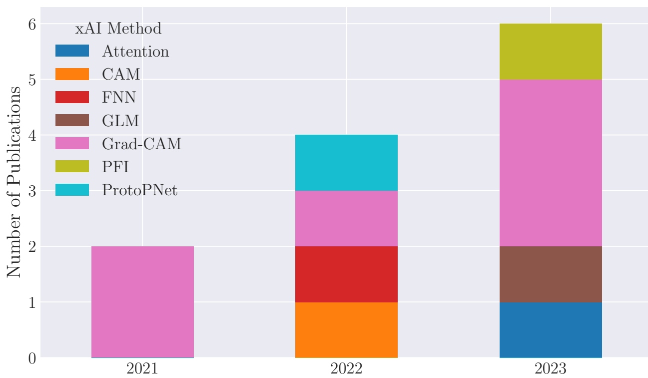



Figure 2. Number of adapted xAI methods per year.

The *scale* in RS describes the scope and resolution of input data [21]. While the scope determines the geographic extent, the resolution specifies to which degree information can be captured, defining the shape, granularity, and boundaries of the objects. Some concepts or objects, like landcover or mountains, have no distinct boundaries or areas. Instead, changes are more continuous and have irregular shapes. Hence, spatial resolution has semantic implications, since the presence or absence of information in the data influences the ability to distinguish and interpret specific features or objects within the scene. This remains a challenge for RS and xAI [58]: most methods have problems with the high granularity of features, highlight an increased spatial extent, or are very noisy. New methods are constantly being developed to overcome these disadvantages. Recent works leverage the lower-level feature representations to create much more fine-grained Grad-CAM saliency maps [22, 40]. Another novel method addressing the challenge of scale decouples the attribution from the spatial space, yielding a permutation method able to attribute across various neighborhoods [8]. The scale of RS observations also determines the spectral resolution. Every sensor captures reflectance data across distinct wavelengths, extracting different ground information. Until now, a comprehensive xAI evaluation taking into account the different spectral properties in RS images is missing. A 3D CAM variant adapted from medical imaging shows the importance of different spectral bands per pixel [13]. This provides insights into both spatial and spectral attribution, in contrast to the current CAM methods, which only visualize spatial attribution.

Another property is the *topology* of the data [58]. Geographic relations can be a hidden confounder not captured by the data. For example, social structures are strongly tied to the location or symbiotic relationships that exist in and between the biosphere, lithosphere, cryosphere, and hydrosphere. That makes analyzing and using xAI in RS difficult. One approach focuses on location-awareness [5], extending the ProtoPNet architecture [12] to account for the relative locations in the input, thus enabling local prototypes. A different strategy that facilitates the learning of spatial relations for monitoring the exposure to surface ozone is presented in [37]. The authors propose using a U-Net model whose input pixels represent high-level features of the neighboring geographical areas. Their approach uncovers the contribution of the neighboring areas by applying the Integrated Gradients [54] method per output pixel. Currently, there is also an increasing interest in the integration of topology data analysis [10] with AI [11, 24]. For example, [14] attempts to discover the topology of the data by mapping each pixel through a lens function, dividing them into subsets, and clustering within these subsets. Different groups are generated that classify the pixels into categories according to the number of pixels in each subset and their connectivity.

Furthermore, RS data often comprises *sequential data*, and most of the existing xAI methods do not account for temporal dependencies [28]. However, several possibilities designed for sequential data exist in the literature [56]. One of the most frequently used is the attention mechanism. The only novel approach identified in this review that explicitly tackles the problem of time dependencies is presented by [39]. They learn regression coefficients per feature, time point, and location. Additionally, the spatiotemporal dependencies are disentangled with a random effect component where the latent variables follow a temporal Markovian process. This approach is evaluated for air quality estimation, where the regression coefficients are used to reveal the temporal importance of various meteorological and landcover features.

### 3.2. Towards interpretable Deep Neural Networks

The initial research for interpreting Deep Neural Networks (DNNs) focused on developing post-hoc, backpropagation methods. Yet, these methods highlight only raw features that usually do not correspond to intuitive concepts and do not provide additional insights into how these features are used by the model [1]. In recent years, self-interpretable

DNNs emerged as a popular approach to address these limitations [25].

The self-interpretable networks proposed for tackling RS spatio-temporal inputs usually rely on the attention mechanism. For instance, the Earthformer [19] applies self-attention to input tensors decomposed into local cuboids, and then attends to global cuboids which summarize the overall status of the system. Another example is given in [55], where a temporal transformer encoder followed by a spatial one is used to process all patches composing the time series. Similar approaches suitable for RS spatio-temporal data can be found in [4, 15, 35, 45]. Further exploration of the explanations provided by the attention mechanisms in these new transformer architectures is still missing in the EO literature and could be a promising direction for latent space analysis-based xAI. It should be noted that attention mechanisms are not solely used in transformers but also in convolutional and recurrent DNNs [17, 27, 29, 41, 48, 53, 59, 60].

Prototype networks and Bagnets are other self-interpretable DNNs that have recently been utilized in EO studies. Prototype networks, which are becoming increasingly popular in computer vision [3, 12, 44], enforce a reasoning process that classifies input examples based on their similarity to prototypical images. In addition to the approach in [5] described above, a prototype network for flood mapping is presented in [63]. The prototypes in this case are associated with cluster centers in the latent space of a U-Net encoder. This enables interpretability in terms of linguistic *IF* ... *THEN* rules. A recent work introduces an interpretable convolutional DNN architecture, the BagNet [7], which, in contrast to the traditional convolutional architectures, processes small image patches at a time. The activations for each class in each receptive field are combined to provide a prediction at the image level. The magnitude of the receptive fields' activations corresponds to the importance of the features, and an attribution map can be directly compiled from the activations. Examples of this methodology applied to vehicle detection are given in [31, 32]. A similar approach integrates DNNs in Generalized Additive Models (GAMs) for landslide susceptibility prediction, where the DNNs approximates the GAM's spline functions. This combination enhances its interpretability and at the same time leverages the capabilities of DNNs [16]. Another study combines a DNN with linear regression, equipping the model with an interpretable-by-design component [62].

### 3.3. Evaluation of xAI

Evaluating explanation quality and its trustworthiness is an essential methodological challenge in xAI that has been extensively tested in several studies [2, 26]. Our results agree with [20] in a sense that most studies provide only anecdotal evidence through qualitative evaluation (see Figure 3).

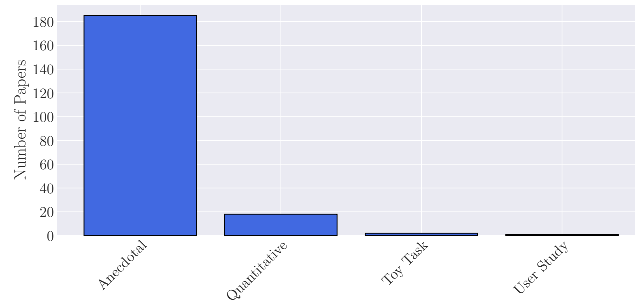



Figure 3. The number of times different evaluation types are considered in the literature, based on their corresponding number of papers.

However, cherry-picking and qualitative evaluation of explanations represent a challenge to humans. Because human perception is mainly visual, humans are biased toward certain types of xAI explanations. For example, humans introduce cognitive distortion by drawing more attention to negative examples and looking for simple but complete explanations [6]. Hence, it is hard to quantify the results objectively through anecdotal evidence. Also, the literature does not define what a sufficient explanation or interpretation is. Nevertheless, the qualification of the interpretability of xAI approaches is essential [34]. Likewise, no standardized and objective evaluation for xAI methods has been established, posing a challenge for potential partitioners outside the xAI domain. Even though frameworks for xAI have been developed recently [42, 46], they are tailored towards a specific type of method or problem [46]. Others focus more on the life-cycle of the systems than the evaluation [42]. Furthermore, we found only one survey that evaluates the usefulness of the explanations to experts [47]. However, with the emphasis on human-centered AI in the current research landscape, user studies are becoming significant to quantify the benefit of the explanations and the understanding of the end-users [61]. Because they have been largely unexplored in the context of xAI in RS, they pose a promising research direction. In summary, the evaluation of xAI lacks a standardized methodology, potentially limiting non-experts applying the methods. This circumstance might have contributed to the large number of anecdotal evidence we encountered in our review.

## 4. Conclusion

A scoping review has been conducted to reveal the emerging trends of xAI in RS. Our results show a recently elevated usage of xAI in critical applications related to natural hazards or atmosphere monitoring, as well as a stagnation of standard tasks like landcover mapping. This is accompanied by increased utilization of local approximation and perturbation approaches, together with a consistent application of backpropagation approaches for explaining DNNs. Yet, these popular approaches usually do not capture the RS data properties like scale, topology, and time series, so we have highlighted novel works tackling these challenges. Finally, we emphasize the promising research directions of developing interpretable DNNs along with quantitative and user-study evaluation procedures to verify the reliability of the extracted explanations.

## Acknowledgements

The work of A. Höhl is funded by the project ML4Earth by the German Federal Ministry for Economic Affairs and Climate Action (grant number 50EE2201C). The work of I. Obadic is funded by the Munich Center for Machine Learning. M.Á. Fernández-Torres acknowledges the support from the European Research Council (ERC) under the ERC Synergy Grant USMILE (grant agreement 855187) and the European Union's Horizon 2020 research and innovation program within the project 'XAIDA: Extreme Events - Artificial Intelligence for Detection and Attribution,' under grant agreement No 101003469. The work of A. Höhl, I. Obadic, M.Á. Fernández-Torres and X. Zhu is supported by the German Federal Ministry of Education and Research (BMBF) in the framework of the international future AI lab "AI4EO – Artificial Intelligence for Earth Observation: Reasoning, Uncertainties, Ethics and Beyond" (grant number: 01DD20001). The authors are responsible for the content of this publication.

## References

[1] Reduan Achtibat, Maximilian Dreyer, Ilona Eisenbraun, Sebastian Bosse, Thomas Wiegand, Wojciech Samek, and Sebastian Lapuschkin. From" where" to" what": Towards human-understandable explanations through concept relevance propagation. *arXiv preprint arXiv:2206.03208*, 2022. 3

[2] Julius Adebayo, Justin Gilmer, Michael Muelly, Ian Goodfellow, Moritz Hardt, and Been Kim. Sanity Checks for Saliency Maps. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2018. 4

[3] Plamen Angelov and Eduardo Soares. Towards explainable deep neural networks (xdnn). *Neural Networks*, 130:185–194, 2020. 4

[4] Anurag Arnab, Mostafa Dehghani, Georg Heigold, Chen Sun, Mario Lučić, and Cordelia Schmid. ViViT: A Video Vision Transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6836–6846, 2021. 4

[5] Elizabeth A. Barnes, Randal J. Barnes, Zane K. Martin, and Jamin K. Rader. This Looks Like That There: Interpretable Neural Networks for Image Tasks When Location Matters. *Artificial Intelligence for the Earth Systems*, 1(3):e220001, 2022. 3, 4

[6] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 78–91, New York, NY, USA, 2022. Association for Computing Machinery. 4

[7] Wieland Brendel and Matthias Bethge. Approximating CNNs with Bag-of-local-Features models works surprisingly well on ImageNet. *International Conference on Learning Representations*, 2019. 4

[8] Alexander Brenning. Spatial machine-learning model diagnostics: A model-agnostic distance-based approach. *International Journal of Geographical Information Science*, 37 (3):584–606, 2023. 3

[9] Gustau Camps-Valls, Devis Tuia, Xiao Xiang Zhu, and Markus Reichstein. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*. 2021. 1

[10] Gunnar Carlsson. Topology and data. *Bulletin of the American Mathematical Society*, 46(2):255–308, 2009. 3

[11] Gunnar Carlsson and Rickard Brüel Gabrielsson. Topological Approaches to Deep Learning. In *Topological Data Analysis*, pages 119–146. Springer International Publishing, Cham, 2020. 3

[12] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. This looks like that: deep learning for interpretable image recognition. *Advances in neural information processing systems*, 32, 2019. 3, 4

[13] G. De Lucia, M. Lapegna, and D. Romano. Towards explainable AI for hyperspectral image classification in Edge Computing environments. *Computers and Electrical Engineering*, 103, 2022. 3

[14] Ludovic Duponchel. When remote sensing meets topological data analysis. *Journal of Spectral Imaging*, page a1, 2018. 3

[15] Haoqi Fan, Bo Xiong, Karttikeya Mangalam, Yanghao Li, Zhicheng Yan, Jitendra Malik, and Christoph Feichtenhofer. Multiscale Vision Transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6824–6835, 2021. 4

[16] Haoran Fang, Yun Shao, Chou Xie, Bangsen Tian, Chaoyong Shen, Yu Zhu, Yihong Guo, Ying Yang, Guanwen Chen, and Ming Zhang. A new approach to spatial landslide susceptibility prediction in karst mining areas based on explainable artificial intelligence. *Sustainability (Switzerland)*, 15 (4), 2023. 4

[17] Sijia Feng, Kefeng Ji, Fulai Wang, Linbin Zhang, Xiaojie Ma, and Gangyao Kuang. PAN: Part attention network integrating electromagnetic characteristics for interpretable SAR vehicle target recognition. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–1, 2023. 4

[18] Bruno Ferreira, Muriel Iten, and Rui G Silva. Monitoring sustainable development by means of earth observation data and machine learning: A review. *Environmental Sciences Europe*, 32(1):1–17, 2020. 1

[19] Zhihan Gao, Xingjian Shi, Hao Wang, Yi Zhu, Yuyang Wang, Mu Li, and Dit-Yan Yeung. Earthformer: Explor-

ing Space-Time Transformers for Earth System Forecasting, 2023. 4

[20] Caroline M. Gevaert. Explainable AI for earth observation: A review including societal and regulatory perspectives. *International Journal of Applied Earth Observation and Geoinformation*, 112:102869, 2022. 1, 2, 4

[21] Michael F. Goodchild. Scale in GIS: An overview. *Geomorphology*, 130(1):5–9, 2011. 3

[22] Xianpeng Guo, Biao Hou, Chen Yang, Siteng Ma, Bo Ren, Shuang Wang, and Licheng Jiao. Visual explanations with detailed spatial information for remote sensing image classification via channel saliency. *International Journal of Applied Earth Observation and Geoinformation*, 118:103244, 2023. 3

[23] Ola Hall, Mattias Ohlsson, and Thorsteinn Rögnvaldsson. A review of explainable AI in the satellite data, deep machine learning, and human poverty domain. *Patterns*, 3(10): 100600, 2022. 1

[24] Felix Hensel, Michael Moor, and Bastian Rieck. A Survey of Topological Machine Learning Methods. *Frontiers in Artificial Intelligence*, 4, 2021. 3

[25] Andreas Holzinger, Randy Goebel, Ruth Fong, Taesup Moon, Klaus-Robert Müller, and Wojciech Samek. xxai-beyond explainable artificial intelligence. In *International Workshop on Extending Explainable AI Beyond Deep Models and Classifiers*, pages 3–10. Springer, 2020. 4

[26] Sara Hooker, Dumitru Erhan, Pieter-Jan Kindermans, and Been Kim. A benchmark for interpretability methods in deep neural networks. *Advances in neural information processing systems*, 32, 2019. 4

[27] Dino Ienco, Yawogan Jean Eudes Gbodjo, Raffaele Gaetano, and Roberto Interdonato. Weakly supervised learning for land cover mapping of satellite image time series via attention-based CNN. *IEEE access : practical innovations, open solutions*, 8:179547–179560, 2020. 4

[28] Aya Abdelsalam Ismail, Mohamed Gunady, Héctor Corrada Bravo, and Soheil Feizi. Benchmarking deep learning interpretability in time series predictions. *Advances in Neural Information Processing Systems*, 2020-Decem(NeurIPS):1–12, 2020. 3

[29] Taewoo Kim, Minsu Jeon, Changha Lee, Junsoo Kim, Geonwoo Ko, Joo-Young Kim, and Chan-Hyun Youn. Federated onboard-ground station computing with weakly supervised cascading pyramid attention network for satellite image analysis. *IEEE access : practical innovations, open solutions*, 10:117315–117333, 2022. 4

[30] Carolin Leluschko and Christoph Tholen. Goals and Stakeholder Involvement in XAI for Remote Sensing: A Structured Literature Review. In *Artificial Intelligence XL*, pages 519–525, Cham, 2023. Springer Nature Switzerland. 1

[31] Peng Li, Cunqian Feng, Xiaowei Hu, and Zixiang Tang. SAR-BagNet: An Ante-hoc Interpretable Recognition Model Based on Deep Network for SAR Image. *Remote Sensing*, 14(9):2150, 2022. 4

[32] Peng Li, Xiaowei Hu, Cunqian Feng, Xiaozhen Shi, Yiduo Guo, and Weike Feng. SAR-AD-BagNet: An Interpretable Model for SAR Image Recognition based on Adversarial Defense. *IEEE Geoscience and Remote Sensing Letters*, pages 1–1, 2022. 4

[33] Thomas Lillesand, Ralph W Kiefer, and Jonathan Chipman. *Remote sensing and image interpretation*. John Wiley & Sons, 2015. 1

[34] Zachary C Lipton. The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57, 2018. 4

[35] Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video Swin Transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3202–3211, 2022. 4

[36] Scott M Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. 30, 2017. 2

[37] Nana Luo, Zhou Zang, Chuan Yin, Mingyuan Liu, Yize Jiang, Chen Zuo, Wenji Zhao, Wenzhong Shi, and Xing Yan. Explainable and spatial dependence deep learning model for satellite-based O3 monitoring in China. *Atmospheric Environment*, page 119370, 2022. 3

[38] Lei Ma, Yu Liu, Xueliang Zhang, Yuanxin Ye, Gaofei Yin, and Brian Alan Johnson. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, 152:166–177, 2019. 1

[39] Paolo Maranzano, Philipp Otto, and Alessandro Fassò. Adaptive LASSO estimation for functional hidden dynamic geostatistical models. *Stochastic Environmental Research and Risk Assessment*, 37(9):3615–3637, 2023. 3

[40] Seyed Mojtaba Marvasti-Zadeh, Devin Goodsman, Nilanjan Ray, and Nadir Erbilgin. Crown-CAM: Interpretable visual explanations for tree crown detection in aerial images. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 3

[41] N. Méger, H. Courteille, A. Benoit, A. Atto, and D. Ienco. Explaining a Deep SpatioTemporal Land Cover Classifier with Attention and Redescription Mining. pages 673–680, 2022. 4

[42] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems*, 11(3-4):24:1–24:45, 2021. 4

[43] Christoph Molnar. *Interpretable Machine Learning*. 2 edition, 2022. 3

[44] Meike Nauta, Ron Van Bree, and Christin Seifert. Neural prototype trees for interpretable fine-grained image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14933–14943, 2021. 4

[45] Daniel Neimark, Omri Bar, Maya Zohar, and Dotan Asselmann. Video Transformer Network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3163–3172, 2021. 4

[46] Frederik Pahde, Maximilian Dreyer, Wojciech Samek, and Sebastian Lapuschkin. Reveal to Revise: An Explainable AI Life Cycle for Iterative Bias Correction of Deep Models. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 596–606, Cham, 2023. Springer Nature Switzerland. 4

[47] Dilli Paudel, Allard de Wit, Hendrik Boogaard, Diego Marcos, Sjoukje Osinga, and Ioannis N. Athanasiadis. Interpretability of deep learning models for crop yield forecasting. *Computers and Electronics in Agriculture*, 206:107663, 2023. 4

[48] J. Pyo, K.H. Cho, K. Kim, S.-S. Baek, G. Nam, and S. Park. Cyanobacteria cell prediction using interpretable deep learning model with observed, numerical, and sensing data assemblage. *Water Research*, 203, 2021. 4

[49] Gabrielle Ras, Ning Xie, Marcel van Gerven, and Derek Doran. Explainable Deep Learning: A Field Guide for the Uninitiated. *Journal of Artificial Intelligence Research*, 73: 329–396, 2022. 2

[50] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204, 2019. 1

[51] Esther Rolf, Konstantin Klemmer, Caleb Robinson, and Hannah Kerner. Mission Critical – Satellite Data is a Distinct Modality in Machine Learning, 2024. 1

[52] R. Roscher, B. Bohn, M.F. Duarte, and J. Garcke. Explain it to me-facing remote sensing challenges in the bio-and geosciences with explainable machine learning. pages 817–824, 2020. 1

[53] Seyd Teymoor Seydi, Mahdi Hasanlou, Jocelyn Chanussot, and Pedram Ghamisi. Leveraging involution and convolution in an explainable building damage detection framework. *European Journal of Remote Sensing*, 0(0):2252166, 2023. 4

[54] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic Attribution for Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3319–3328. PMLR, 2017. 3

[55] Michail Tarasiou, Erik Chavez, and Stefanos Zafeiriou. ViTs for SITS: Vision Transformers for Satellite Image Time Series, 2023. 4

[56] Andreas Theissler, Francesco Spinnato, Udo Schlegel, and Riccardo Guidotti. Explainable AI for Time Series Classification: A Review, Taxonomy and Research Directions. *IEEE Access*, 10:100700–100724, 2022. 3

[57] Devis Tuia, Ribana Roscher, Jan Dirk Wegner, Nathan Jacobs, Xiaoxiang Zhu, and Gustau Camps-Valls. Toward a Collective Agenda on AI for Earth Science Data Analysis. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):88–104, 2021. 1

[58] Jin Xing and Renee Sieber. The challenges of integrating explainable artificial intelligence into GeoAI. *Transactions in GIS*, 27(3):626–645, 2023. 1, 3

[59] Wei Xiong, Zhenyu Xiong, Yaqi Cui, Linzhou Huang, and Ruining Yang. An Interpretable Fusion Siamese Network for Multi-modality Remote Sensing Ship Image Retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2022. 4

[60] J. Xu, J. Yang, X. Xiong, H. Li, J. Huang, K.C. Ting, Y. Ying, and T. Lin. Towards interpreting multi-temporal deep learning models in crop mapping. *Remote Sensing of Environment*, 264, 2021. 4

[61] Wei Xu, Marvin J. Dainoff, Liezhong Ge, and Zaifeng Gao. Transitioning to Human Interaction with AI Systems: New Challenges and Opportunities for HCI Professionals to Enable Human-Centered AI. *International Journal of Human–Computer Interaction*, 39(3):494–518, 2023. 4

[62] X. Yan, Z. Zang, Y. Jiang, W. Shi, Y. Guo, D. Li, C. Zhao, and L. Husi. A Spatial-Temporal Interpretable Deep Learning Model for improving interpretability and predictive accuracy of satellite-based PM2.5. *Environmental Pollution*, 273, 2021. 4

[63] Ziyang Zhang, Plamen Angelov, Eduardo Soares, Nicolas Longepe, and Pierre Philippe Mathieu. An interpretable deep semantic segmentation method for earth observation. In *2022 IEEE 11th International Conference on Intelligent Systems (IS)*, pages 1–8, 2022. 4

[64] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2921–2929, 2016. 2

[65] Xiao Xiang Zhu, Devis Tuia, Lichao Mou, Gui Song Xia, Liangpei Zhang, Feng Xu, and Friedrich Fraundorfer. Deep Learning in Remote Sensing: A Comprehensive Review and List of Resources. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):8–36, 2017. 1

# Recent Trends, Challenges, and Limitations of Explainable AI in Remote Sensing

Supplementary Material

## 1. EO tasks grouping

We have categorized related EO tasks into groups to provide a better overview. Here we provide a glossary of these groups and the tasks they include.

**Agricultural Monitoring:** All crop-related tasks, like crop yield prediction, crop type classification, irrigation scheme classification, and crop lodging detection.

**Atmosphere Monitoring:** The prediction of atmospheric phenomena, like air quality, aerosol optical depth, and dust storm indices.

**Building Mapping:** All tasks related to buildings and urban structures, like building footprint classification and building damage mapping.

**Ecosystem Interactions:** All interactions of the ecosystem with other systems, e.g., the atmosphere and the hydrosphere. This includes the ecosystem CO2 exchange or the sun-induced fluorescence prediction.

**Human Environment Interaction:** The monitoring of human structures and the environment, like human footprint estimation, socioeconomic status estimation, or well-being prediction.

**Hydrology Monitoring:** Tasks related to hydrology, like runoff forecasting, water quality, streamflow prediction, and water segmentation, but excluding floods.

**Landcover Mapping:** The most common EO task includes mainly landcover classification but also related tasks like slum mapping.

**Natural Hazard Monitoring:** Monitoring of natural hazards, like landslides, wildfires, floods, earthquakes, and volcanos.

**Soil Monitoring:** Monitoring soil properties, like soil texture, respiration, moisture, and salinity.

**Surface Temperature Prediction:** The prediction of the Earth's surface temperature.

**Target Mapping:** Tasks related to the mapping of specific targets, like vehicles and objects.

**Vegetation Monitoring:** Monitoring of vegetation, excluding crops, like vegetation regeneration, tree monitoring, tree classification, and tree mapping.

**Weather and Climate Prediction:** Forecasting of weather and climate variables, like precipitation, temperature, and drought.

**Other:** All the tasks which did not fit into the other groups. This includes change detection, urban mobility, sea ice classification, mosquito modeling, and satellite product quality tasks.