DOI: 10.34185/1991-7848.itmm.2026.01.063

ШІ-ПІДХОДИ ПОЯСНЕННЯ СУПУТНИКОВИХ ЗНІМКІВ

Бубнов М.С.[1] [ORCID], Гнатушенко В.В.[2] [ORCID], Гнатушенко Вік.В.[3] [ORCID]

[1]Український державний університет науки і технологій, аспірант, України

[2]Національний технічний університет «Дніпровська політехніка»

д.т.н., професор, України

[3]Український державний університет науки і технологій

д.т.н., професор, України

Анотація. *У роботі розглянуто сучасні підходи пояснюваного штучного інтелекту (Explainable Artificial Intelligence, XAI), що застосовуються для інтерпретації та обробки супутникових і аерокосмічних знімків у задачах дистанційного зондування Землі. Проаналізовано основні класи XAI-методів, зокрема атрибуцію ознак, дистиляцію моделей, внутрішньо інтерпретовані підходи та контрастивні пояснення, принципи їх роботи, переваги, обмеження та обчислювальні особливості. Наведено приклади практичного використання XAI для моніторингу природних катастроф, агромоніторингу, оцінки соціально-економічних індикаторів, класифікації землекористування та аналізу мультимодальних даних. Зроблено висновок, що XAI є важливим інструментом для підвищення надійності, прозорості та прийнятності результатів моделей штучного інтелекту у критично важливих прикладних застосуваннях.*

Ключові слова: *пояснюваний штучний інтелект, XAI, дистанційне зондування Землі, аерокосмічні знімки, інтерпретація моделей, атрибуція ознак, дистиляція моделей, супутникові дані.*

Обробка та аналіз супутникових і аерокосмічних знімків є однією з ключових задач сучасного дистанційного зондування Землі. Швидке зростання обсягів даних, доступних з оптичних, радарних, мульти- та гіперспектральних сенсорів, зумовлює активне використання методів штучного інтелекту, зокрема глибокого навчання. Разом з тим складність таких моделей призводить до зниження прозорості процесу прийняття рішень, що є критичним у задачах з високим рівнем відповідальності. Пояснюваний штучний інтелект (Explainable AI, XAI) пропонує набір підходів, спрямованих на підвищення зрозумілості та інтерпретованості моделей. У контексті супутникових знімків XAI дозволяє не лише перевіряти надійність прогнозів, а

й інтегрувати експертні знання, виявляти помилки в даних та забезпечувати регуляторну й етичну прозорість.

Застосування XAI у задачах дистанційного зондування має низку суттєвих переваг. По-перше, пояснювальні методи дозволяють виявляти ситуації, коли модель ґрунтує свої рішення на нерелевантних ознаках або артефактах, таких як хмарність чи смуги сенсора. По-друге, XAI сприяє підвищенню довіри користувачів та полегшує комунікацію результатів між розробниками моделей і доменними експертами. По-третє, у контексті прийняття управлінських або гуманітарних рішень XAI забезпечує можливість обґрунтування отриманих висновків [4].

Водночас XAI-підходи мають і низку обмежень. Пояснення часто є нечіткими та не завжди відображають причинно-наслідкові зв’язки. Результати можуть бути чутливими до архітектури моделі, параметрів навчання та попередньої обробки даних. Крім того, багато методів характеризуються високою обчислювальною складністю, що ускладнює їх застосування для великих зображень або гіперспектральних даних. Важливою проблемою залишається відсутність стандартизованих метрик оцінки якості пояснень [3].

Розглянемо класифікацію XAI-підходів для аналізу та пояснення супутникових знімків, а також приклади практичного застосування XAI.

Атрибуція ознак (Feature Attribution). Методи атрибуції ознак спрямовані на визначення вхідних параметрів, які найбільше впливають на прогноз моделі. Для досягнення цієї мети, дані методи використовують навчену модель машинного навчання. Залежно від того, чи пояснення генеруються шляхом перевірки внутрішніх механізмів моделі, чи шляхом аналізу змін у вихідних даних моделі після модифікації вхідних ознак, методи цієї категорії додатково поділяються на методи на основі зворотного поширення та методи на основі збурень.

Методи зворотного поширення, такі як Gradient, Integrated Gradients, Deconvolution, Class Activation Mapping (CAM, Grad-CAM) та Layer-wise Relevance Propagation, використовують внутрішню структуру глибоких

нейронних мереж для оцінки важливості пікселів або регіонів зображення. Їхньою перевагою є відносна простота та наочність у вигляді теплових карт, проте результати можуть бути складними для інтерпретації у мультиспектральних та мультикласових задачах.

Методи на основі збурень (Occlusion Sensitivity, PDP, ALE) оцінюють важливість ознак шляхом аналізу зміни прогнозу при систематичній модифікації вхідних даних. Ці методи відрізняються тим, як саме ознаки збурюються. Серед інших, типи збурення включають розмиття, усереднення, перетасування або додавання шуму. Дані методи інтуїтивні та часто дають стабільніші пояснення, однак є обчислювально затратними та чутливими до параметрів збурення [2, 6].

Дистиляція моделей (Distillation). Методи дистиляції будують інтерпретовану сурогатну модель, яка апроксимує поведінку складної моделі. Відтворюючи прогнози складної моделі, сурогатна модель пропонує гіпотези про відповідні ознаки та кореляції, отримані складною моделлю, не надаючи додаткового розуміння її внутрішнього механізму прийняття рішень. До цієї групи належать локальні апроксимації (LIME, SHAP) та методи трансляції моделей у правила, дерева або графи. Основною перевагою таких підходів є кількісна оцінка внеску ознак, однак їх застосування до гіперспектральних і мультимодальних даних потребує додаткових адаптацій [5].

Внутрішньо інтерпретовані методи (Intrinsic). Внутрішньо інтерпретовані підходи орієнтовані на створення моделей, які є зрозумілими за своєю структурою. До них належать інтерпретовані за проектом моделі (дерева рішень, лінійні та адитивні моделі), а також підходи на основі простору вкладень і концептів. Методи на кшталт TCAV дозволяють пояснювати рішення через семантично зрозумілі концепти, що є особливо важливим у соціально значущих застосуваннях. Перевагами цих методів є те, що вони створюють семантично інтерпретовані пояснення. Можуть використовуватись у політико-соціальних застосунках (наприклад, пояснення прогнозів добробуту за супутниковими даними). Головним недоліком є те, що такі методи

потребують наборів прикладів концептів. Також концепти інколи важко формалізувати для мультиполюсних/гетерогенних сцен.

Контрастивні пояснення (Contrastive Examples). Контрастивні методи надають альтернативні приклади, що дозволяють пояснити рішення моделі шляхом порівняння. Контрфактичні пояснення відповідають на питання про мінімальні зміни, необхідні для зміни прогнозу, тоді як пояснення на основі прикладів демонструють схожі історичні випадки з навчальної вибірки. Такі підходи є інтуїтивно зрозумілими та наближеними до людського способу міркування [1].

Приклади практичного застосування XAI. XAI-методи широко застосовуються для моніторингу природних катастроф, де теплові карти дозволяють визначити критичні регіони зображення, що впливають на класифікацію пожеж або повеней. В агромоніторингу пояснення важливих спектральних смуг сприяють кращому розумінню впливу біофізичних факторів на врожайність. У задачах оцінки соціально-економічних індикаторів концепт-базовані підходи допомагають встановити зв’язок між візуальними ознаками інфраструктури та прогнозованими показниками добробуту.

Висновки. Пояснюваний штучний інтелект є перспективним напрямом розвитку методів аналізу супутникових знімків. Його використання сприяє підвищенню довіри до моделей, полегшує інтеграцію експертних знань і дозволяє виявляти помилки у даних та моделях. Разом з тим актуальними залишаються проблеми нечіткості інтерпретацій, високої обчислювальної вартості та відсутності єдиних стандартів оцінки. Для практичних застосувань доцільно комбінувати різні XAI-підходи, проводити кількісну валідацію пояснень і документувати параметри використаних методів.

ЛІТЕРАТУРА / REFERENCE

1. Höhl, A., Obadic, I., Fernández Torres, M. Á., Najjar, H., Oliveira, D., Akata, Z., Dengel, A., & Zhu, X. X. Opening the Black-Box: A Systematic Review on Explainable AI in Remote Sensing. arXiv preprint arXiv:2402.13791 (2024). URL: https://arxiv.org/abs/2402.13791

2. Kakogeorgiou, I. & Karantzalos, K. Evaluating explainable artificial intelligence methods for multi-label deep learning classification tasks in remote sensing. Journal of Arid Environments (Elsevier). URL: https://www.sciencedirect.com/science/article/pii/S0303243421002270

3. (P08) Recent Trends, Challenges, and Limitations of Explainable AI in Remote Sensing (scoping review). XAI4CV workshop paper. URL: https://xai4cv.github.io/assets/papers2024/P08.pdf
4. Klotz, J., Burgert, T., & Demir, B. On the Effectiveness of Methods and Metrics for Explainable AI in Remote Sensing Image Scene Classification. arXiv:2507.05916 (2025). URL: https://arxiv.org/abs/2507.05916
5. Interpretable Deep Learning Framework for Land Use and Land Cover Classification in Remote Sensing using SHAP. International Journal of Information Technology and Computer Engineering (2025). URL: https://ijitce.org/index.php/ijitce/article/view/1346/1165
6. Hnatushenko Vik., Honcharov O. Land cover mapping with Sentinel-2 imagery using deep learning semantic segmentation models (2024). Proceedings of the X International Scientific Conference "Information Technology and Implementation", p.1-18. URL: CEUR-WS.org/Vol-3909

EXPLAINABLE AI APPROACHES FOR SATELLITE IMAGE INTERPRETATION

Mykola Bubnov, Volodymyr Hnatushenko, Viktoriia Hnatushenko

Abstract. *This paper explores contemporary Explainable Artificial Intelligence (XAI) methods applied to the interpretation and analysis of satellite and aerospace imagery within remote sensing tasks. We present a comprehensive overview of primary categories of XAI techniques – including feature attribution, model distillation, intrinsically interpretable models, and contrastive explanations – and discuss their operational principles, strengths, limitations, and computational characteristics. The practical relevance of XAI for tasks such as natural disaster monitoring, agricultural assessment, socio-economic indicator estimation, and land use/land cover classification is highlighted with real-world examples. Emphasis is placed on how explainability enhances model transparency, reliability, and integration with domain expert knowledge. Challenges such as ambiguous interpretations, high computational costs, and the lack of standardized evaluation metrics for explanations are also discussed. The review underscores XAI's growing importance in bridging the gap between black-box AI performance and human understanding in Earth observation applications.*

Keywords: *Explainable Artificial Intelligence, Remote Sensing, Satellite Imagery, Interpretable Models, Feature Attribution, Model Distillation, Satellite Data.*