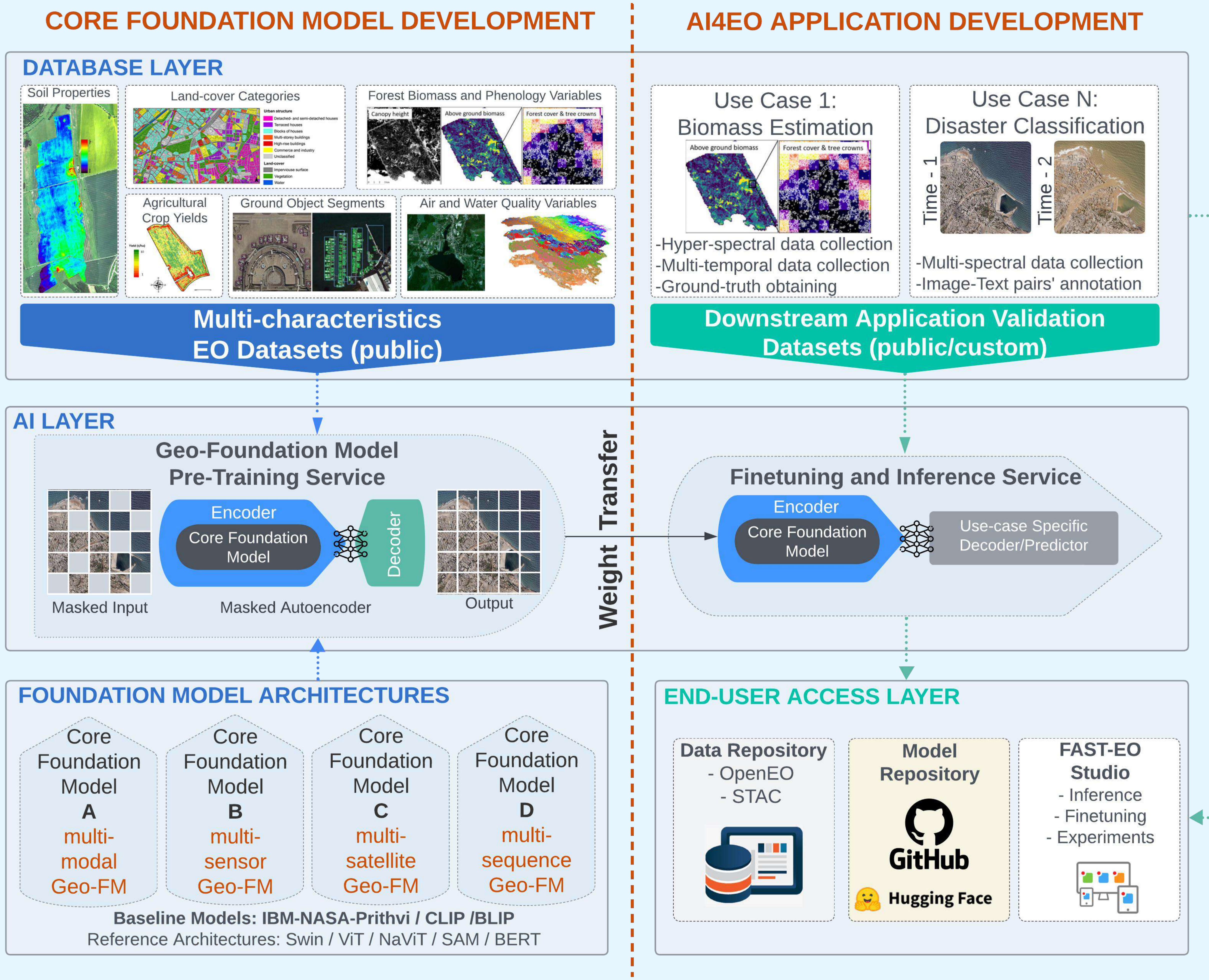
FAST-EO

Multi-Modal Foundation Models for Scalable Earth Observation and Earth Sciences

Image Credit: J. Jakubik, B. Blumenstiel

R. S. Kuzu, T. Brunschwiler, G. Cavallaro, J. Nalepa, 5, C. O. Dumitru, A. Zappacosta, D. E. Molina,
R. Kienzler, J. Jakubik, B. Blumenstiel, Paolo Fraccaro, Felix Yang, R. Sedona, S. Maurogiovanni,
E. Scheurer, A. Wijata, L. Tulczyjew, D. Marek, J. Sadel, S. Ofori-Ampofo , N. Dionelis, N. Longepe

FAST-EO
Project Overview



UC1: Weather & Climate Disaster Analysis

UC2: Detection of Methane Leaks

UC3: Observation of Changes in Forest Above-Ground Biomass

UC4: Estimation of Soil Properties

UC5: Detection of Semantic Land Cover Changes

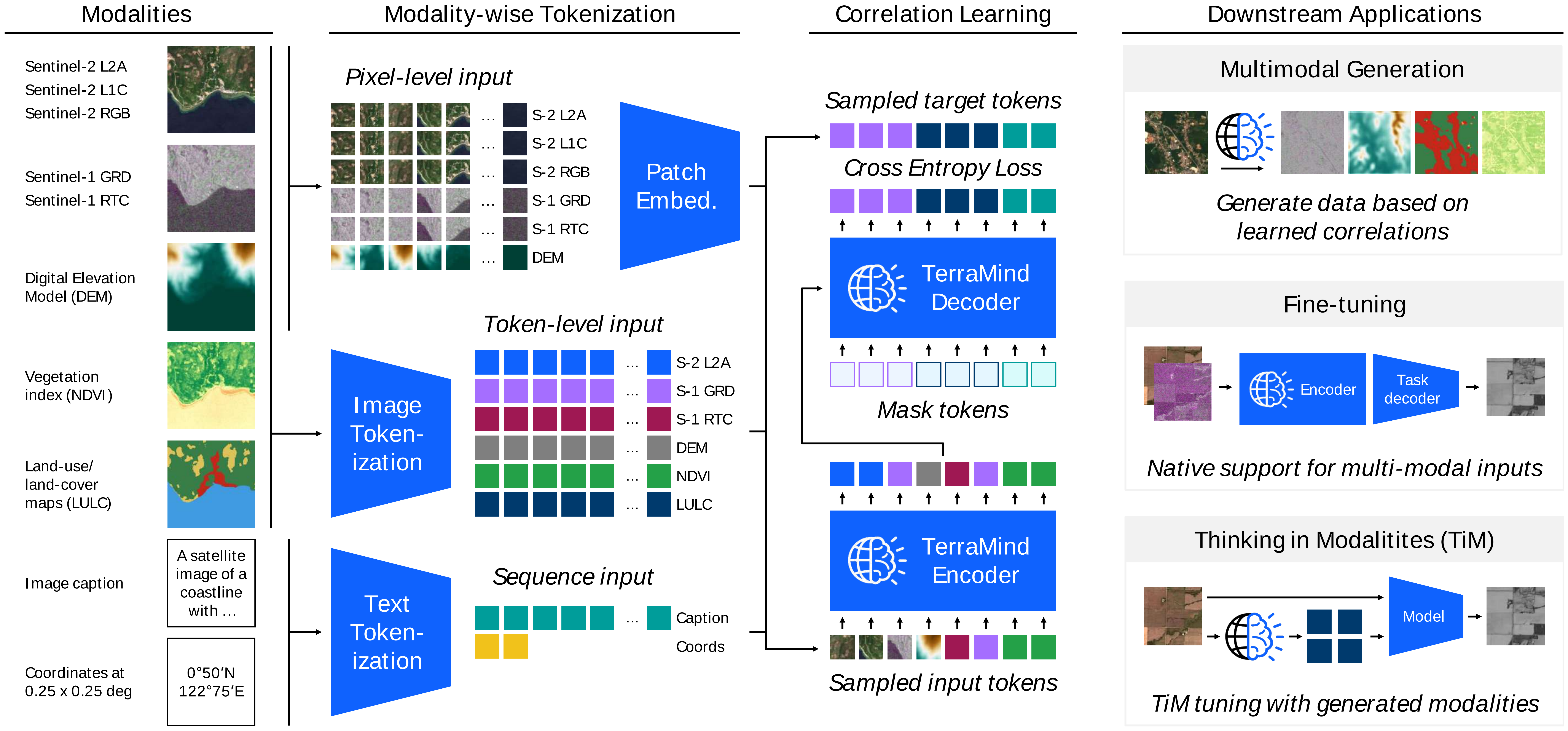
UC6: Monitoring Expansion of Mining Fields into Farmlands

TerraMind

TerraMind represents the first any-to-any generative, and large-scale **multimodal** model for Earth observation pre-trained on **500 billion tokens** from global geospatial data.

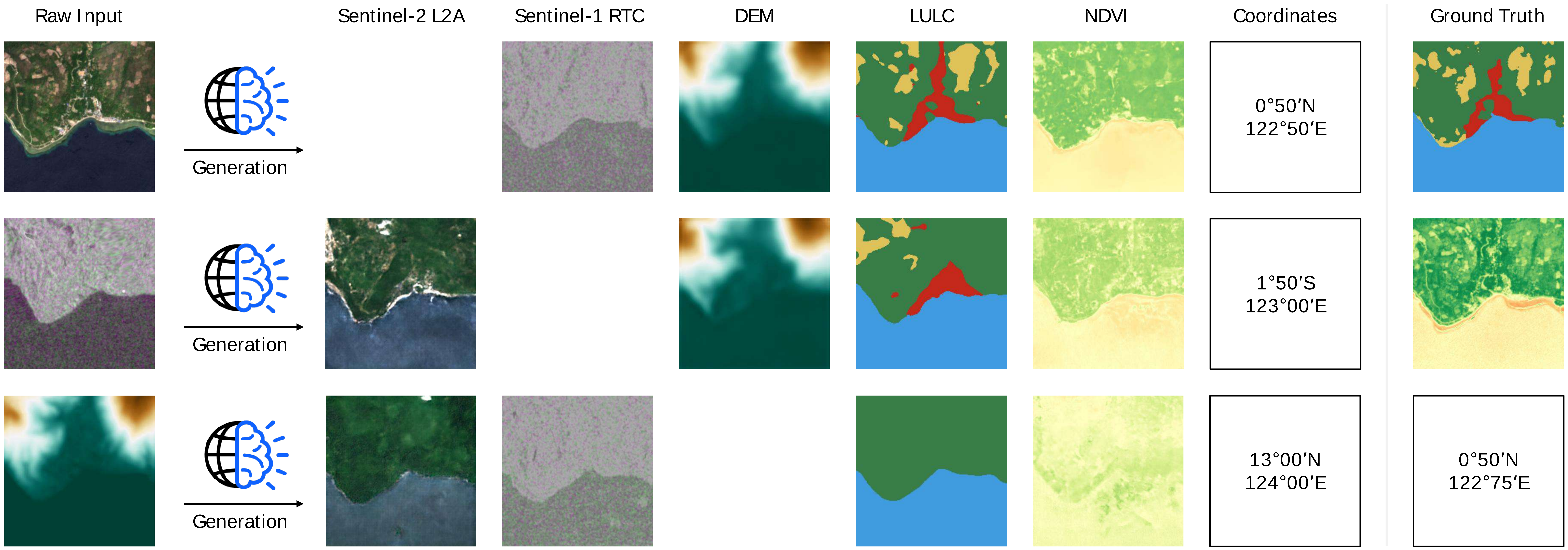
The model digests inputs at **pixel-level, token-level, and as sequences**, simultaneously.

TerraMind outperforms other deep learning models for Earth observation in downstream applications and unlocks **any-to-any generation** and **Thinking-in-Modalities (TiM)** finetuning and inference.

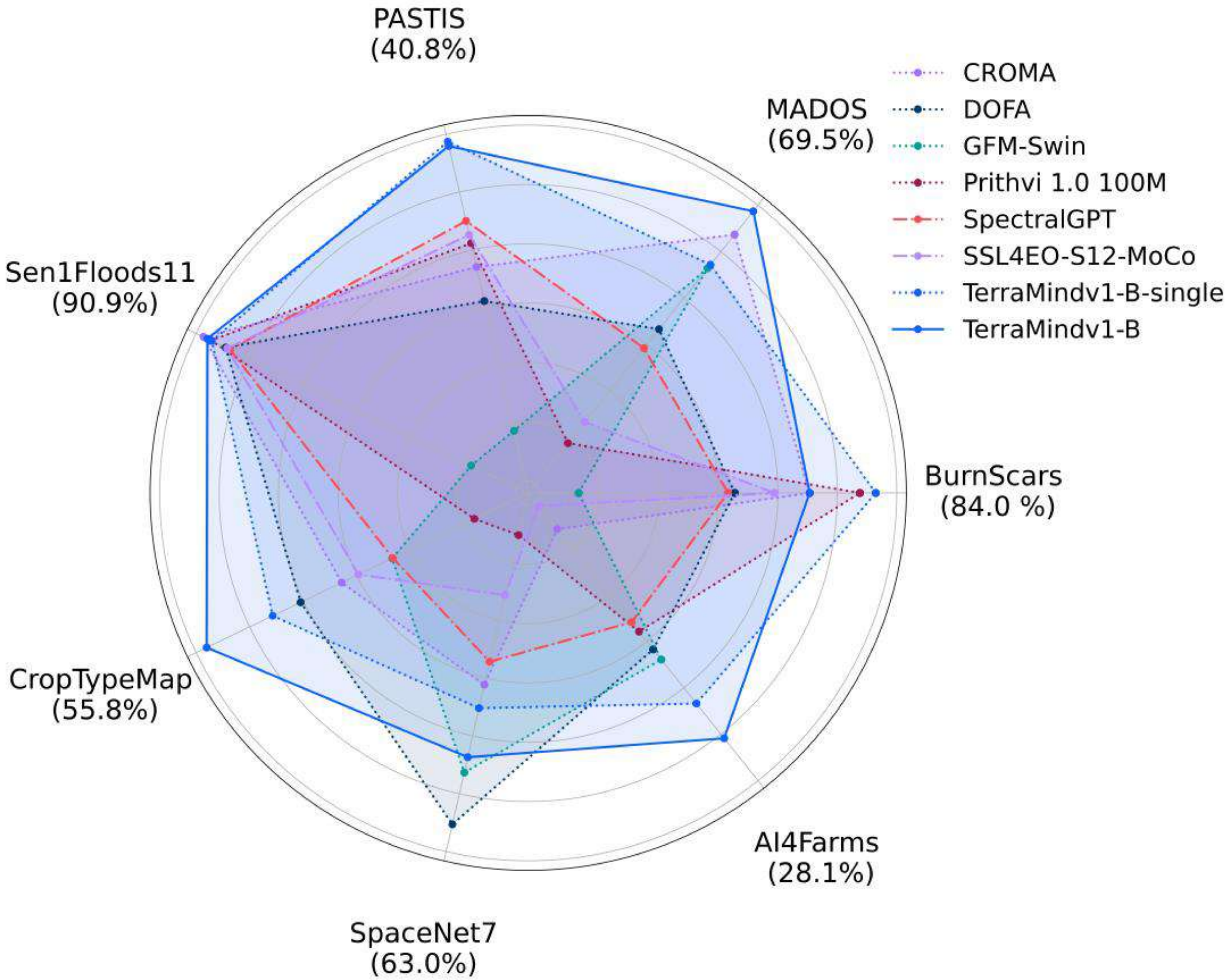


Jakubik, J., Yang, F., Blumenstiel, B., Scheurer, E., Sedona, R., Maurogiovanni, S., ... & Longépé, N. (2025). TerraMind: Large-Scale Generative Multimodality for Earth Observation. *arXiv preprint arXiv:2504.11171*.

TerraMind – foundation model with cross-modal understanding



Evaluation on PANGAEA bench



Hands-on session and Deep dive

TerraMind is evaluated on PANGAEA bench with a diverse set of modalties and downstream tasks – with a frozen encoder.

It outperforms all other evaluated geospatial foundation models and even fully fine-tuned UNet and ViT models.

TerraMind benefits from multi-modal inputs and the new Thinking-in-Modalities approach for improved performance results.

Marsocci, V., Jia, Y., Bellier, G. L., Kerekes, D., Zeng, L., Hafner, S., ... & Nascetti, A. (2024). PANGAEA: A global and inclusive benchmark for geospatial foundation models. *arXiv preprint arXiv:2412.04204*.

MS CLIP – Zero-shot applications via contrastive learning

Vision Language Models enable interactive applications based on natural language.

CLIP is the most promindent model with zero-shot classification and text-to-image retrieval capabilities.

Retrieved images based on image-text similarity

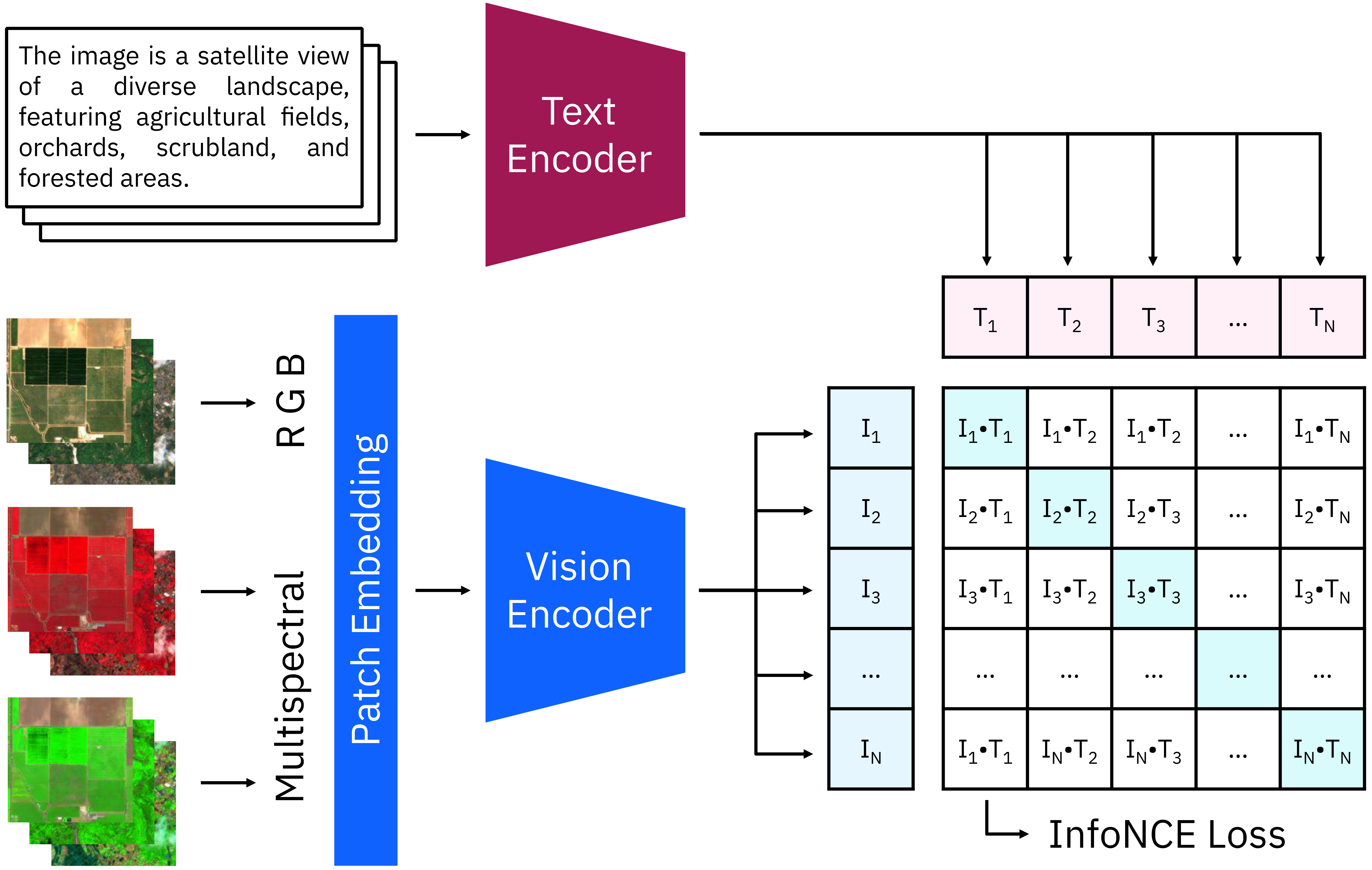
Classification with *solar farm* vs. *others*

MS CLIP

CLIP[1] is trained with Contrastive Learning on 400M image-text pairs.

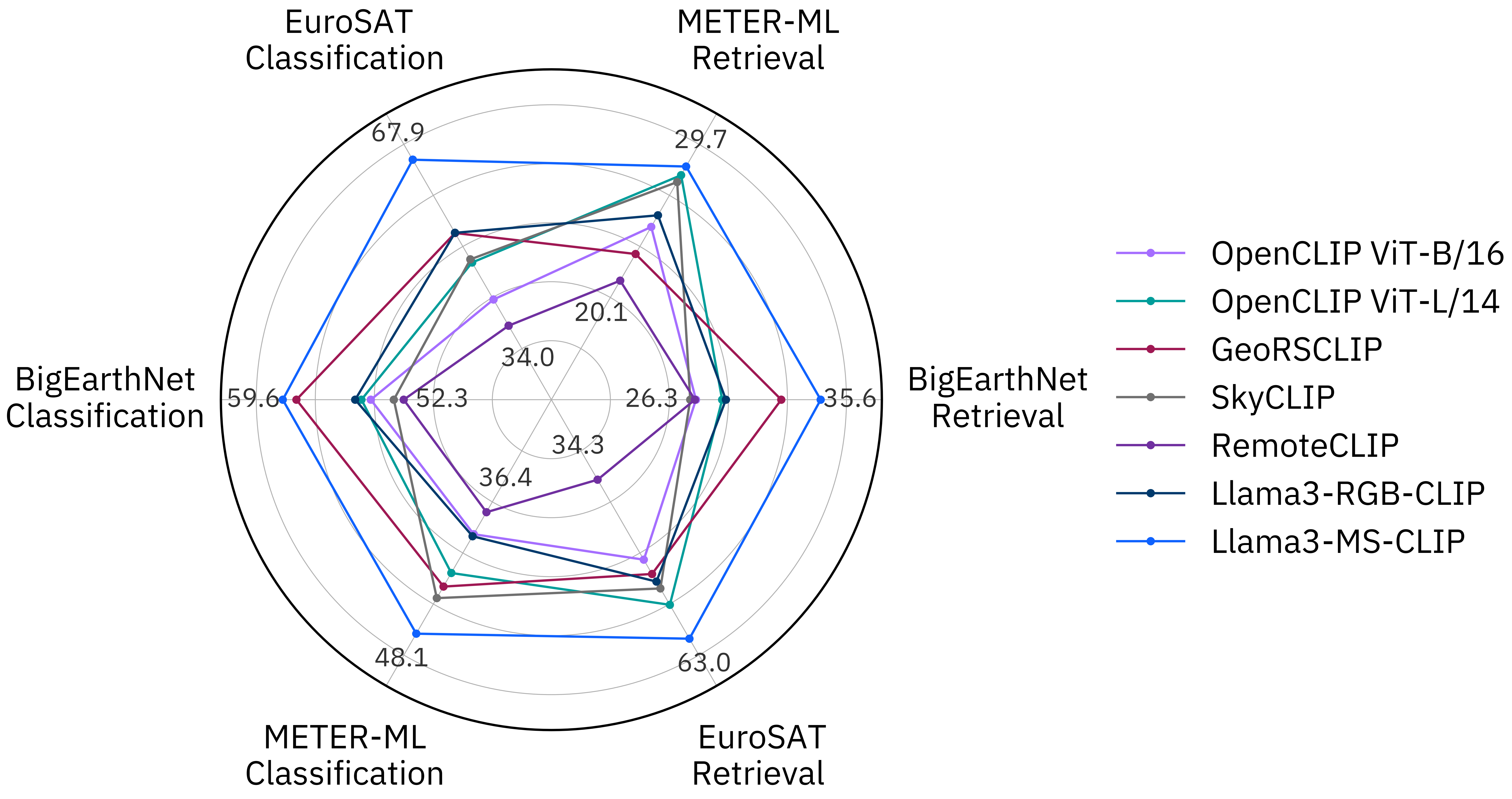
But the model does not generalize well on domain specific tasks.

Continous pre-training with additional channels for EO domain adoption.



1 Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... & Sutskever, I. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning.*

MS CLIP – Zero-shot evaluation



MS CLIP outperforms all baselines and EO-specific VLMs.

MS CLIP improves classification accuracy by +6.77% on average and retrieval by +4.63% mAP compared to the second-best model.

Benefit of multi-spectral data as our RGB-CLIP only performs on pair with other EO VLMs.

Marimo, C. T., Blumenstiel, B., Nitsche, M., Jakubik, J., & Brunschwiler, T. (2025). Beyond the Visible: Multispectral Vision-Language Learning for Earth Observation. *arXiv preprint arXiv:2503.15969*.

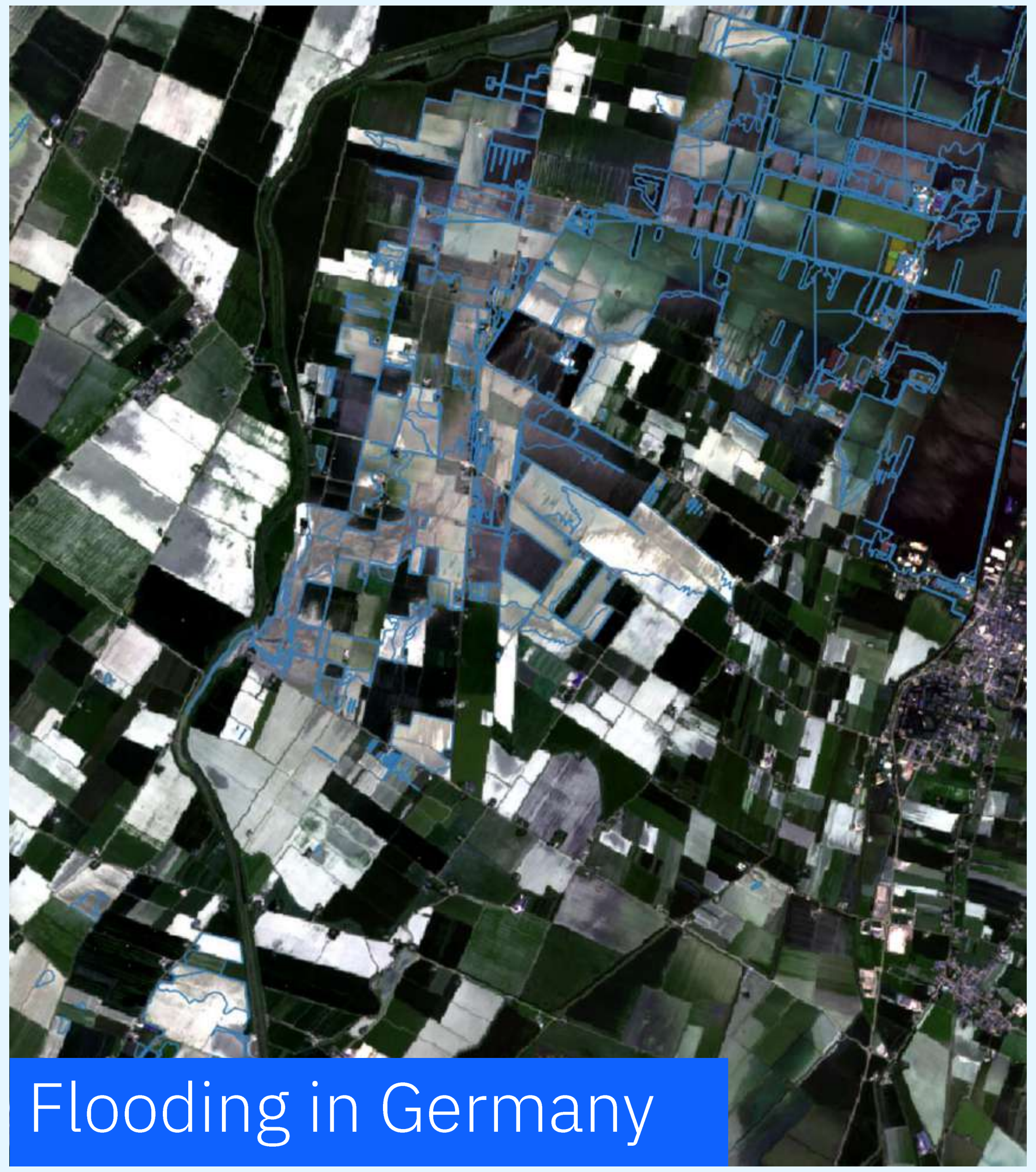
Downstream Applications

Climate impact analysis

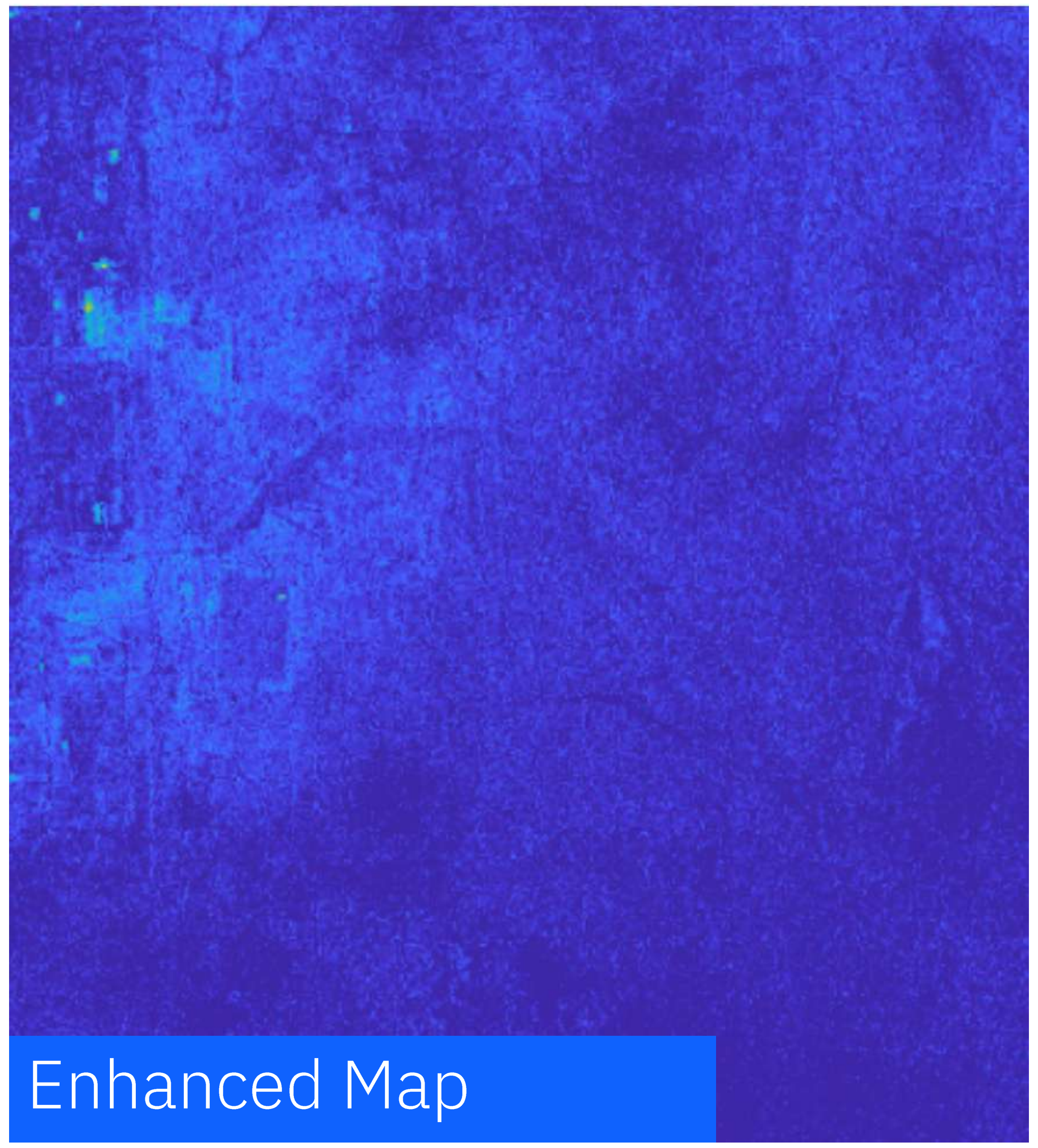
Building a large-scale, multi-modal, multi-temporal dataset for predicting various disaster types. Release in Q2 – stay tuned!

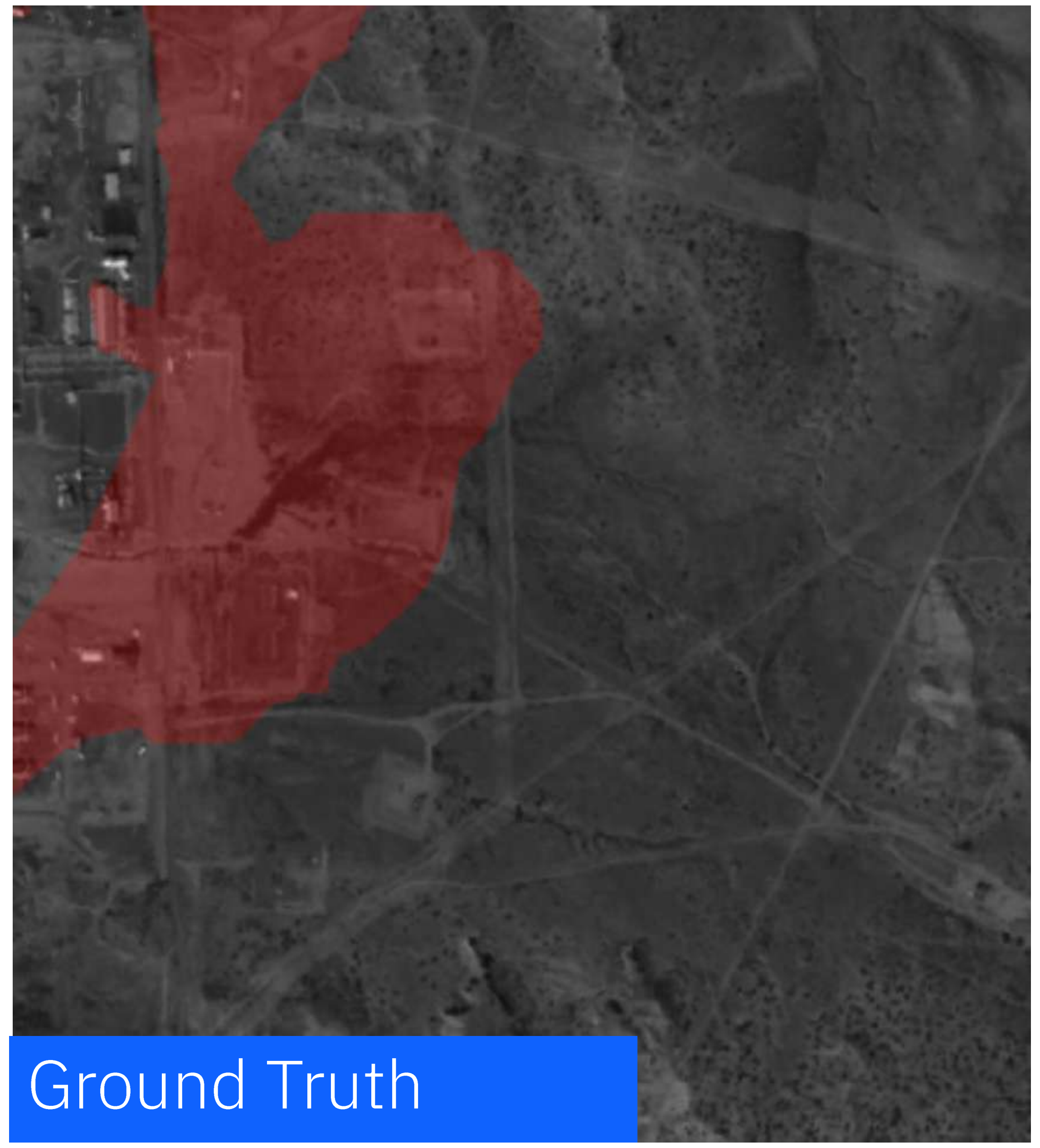
Detection of Methane Leaks

Fine-tuning TerraMind to detect methane leaks in airborne and satellite imagery, achieving sensitivity of 86.9%, and specificity of 82.3%, thereby surpassing the benchmarks.


Flooding in Germany


Wild fire on Corfu


Enhanced Map


Ground Truth

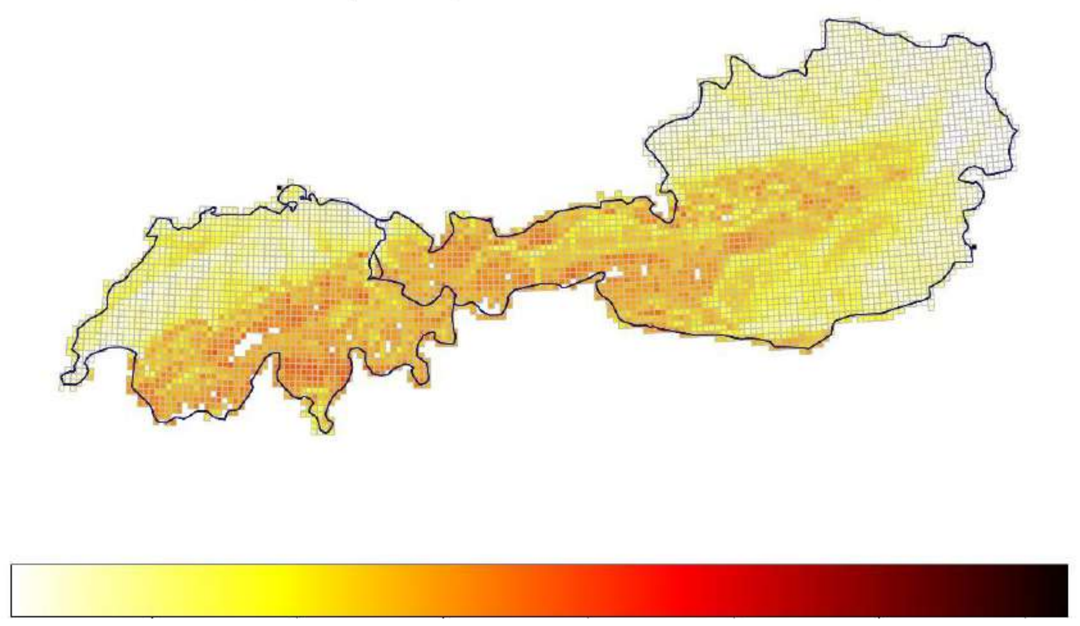
Downstream Applications

Forest Biomass Change Monitoring

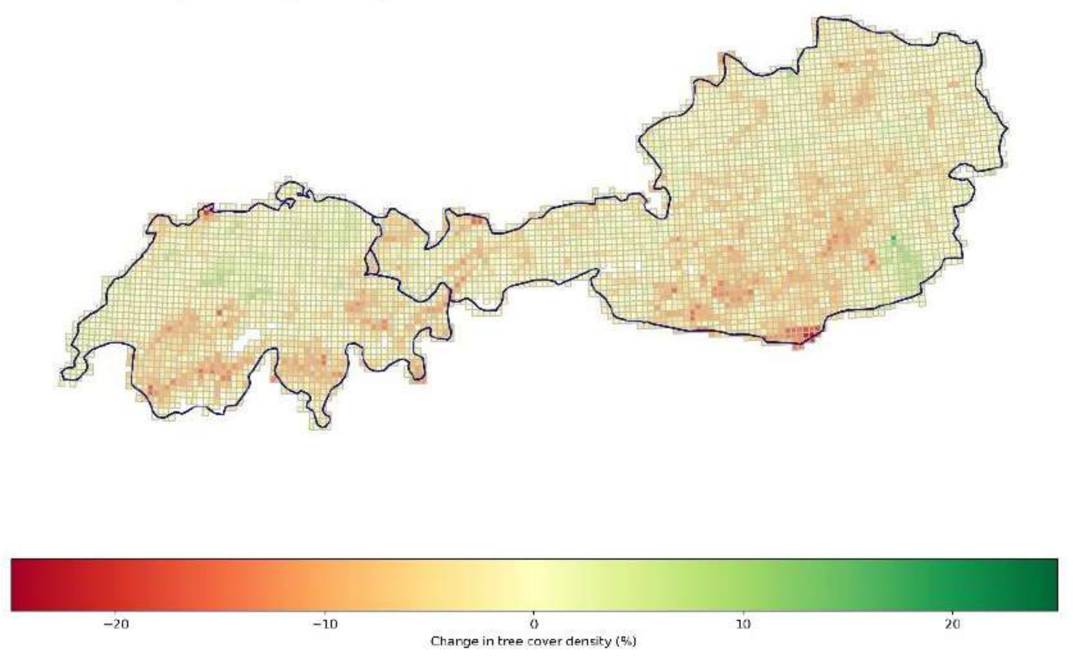
Investigating how terrain topography affects the detection of forest biomass changes and seeking improvements for mountainous regions using the TerraMind model.

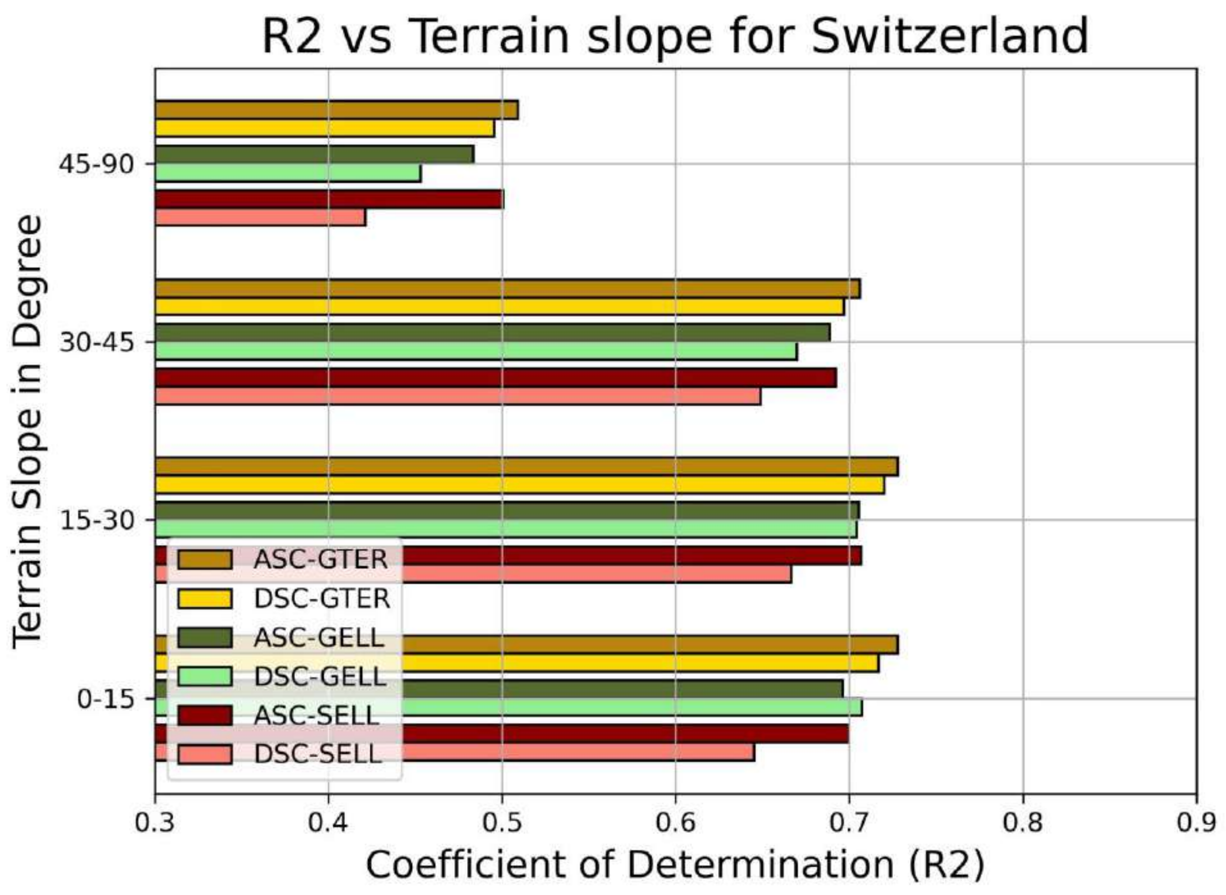
.

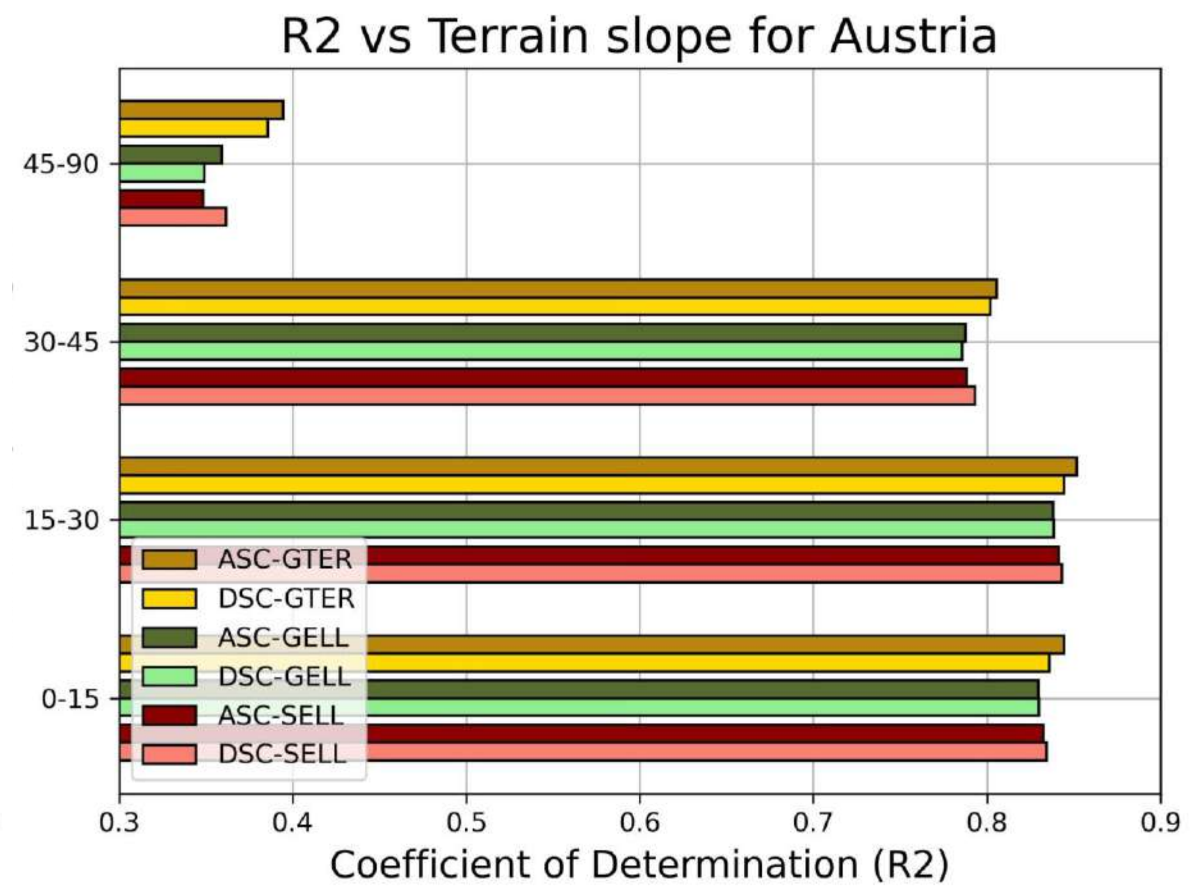
Terrain slope map in Austria and Switzerland



Tree cover density change map in Austria and Switzerland - between 2015-2018

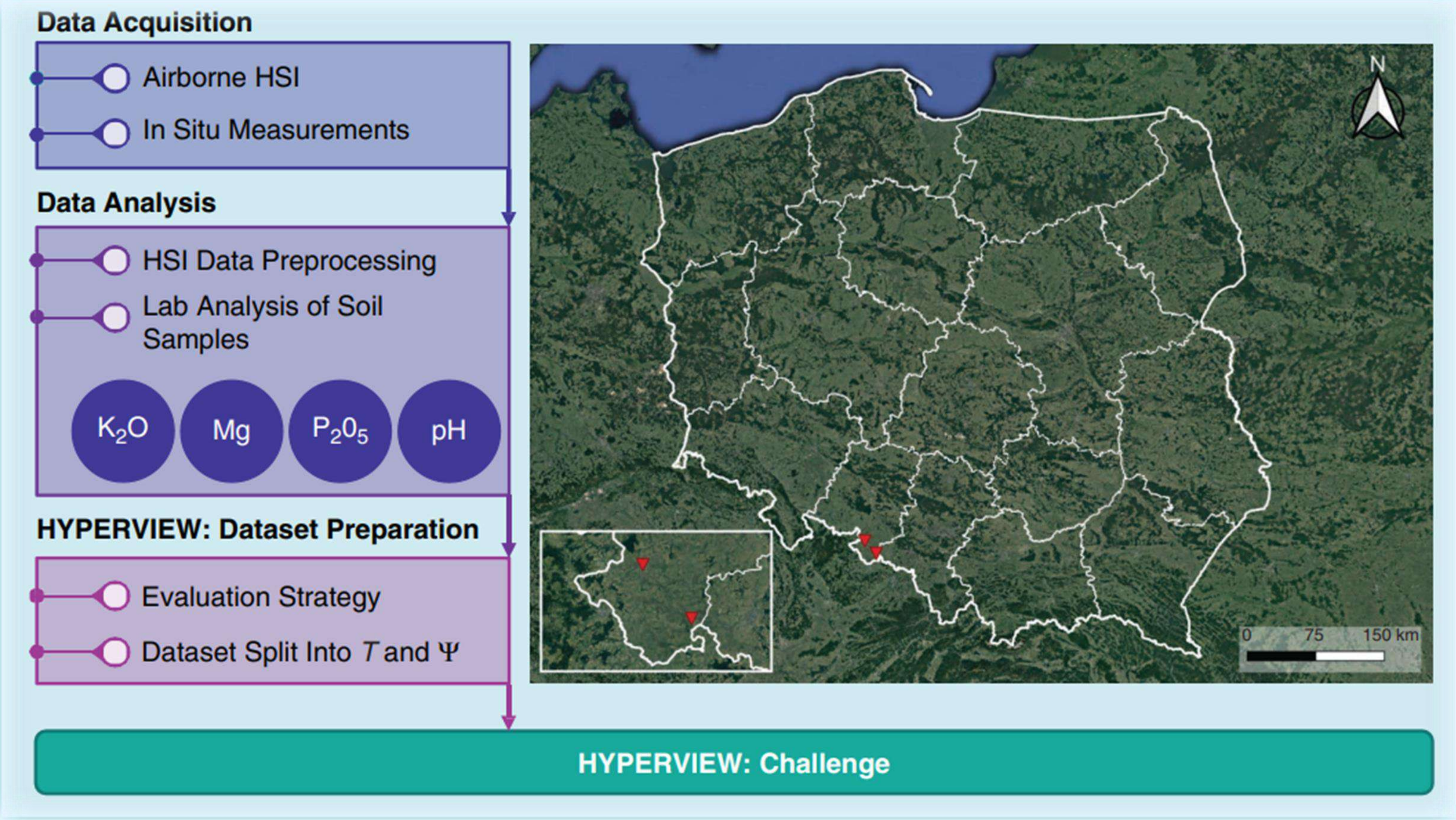




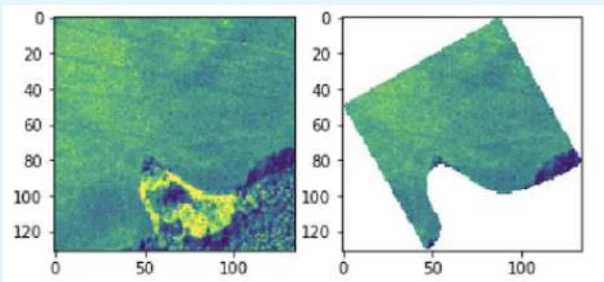


Estimation of Soil Properties

Fine-tuning the TerraMind model to regress soil properties, achieving a HYPERVIEW score of 0.8537 and securing 4th place in the benchmark.



J. Nalepa et al., "Estimating Soil Parameters From Hyperspectral Images: A benchmark dataset and the outcome of the HYPERVIEW challenge," in IEEE Geoscience and Remote Sensing Magazine, vol. 12, no. 3, pp. 35-63, Sept. 2024,

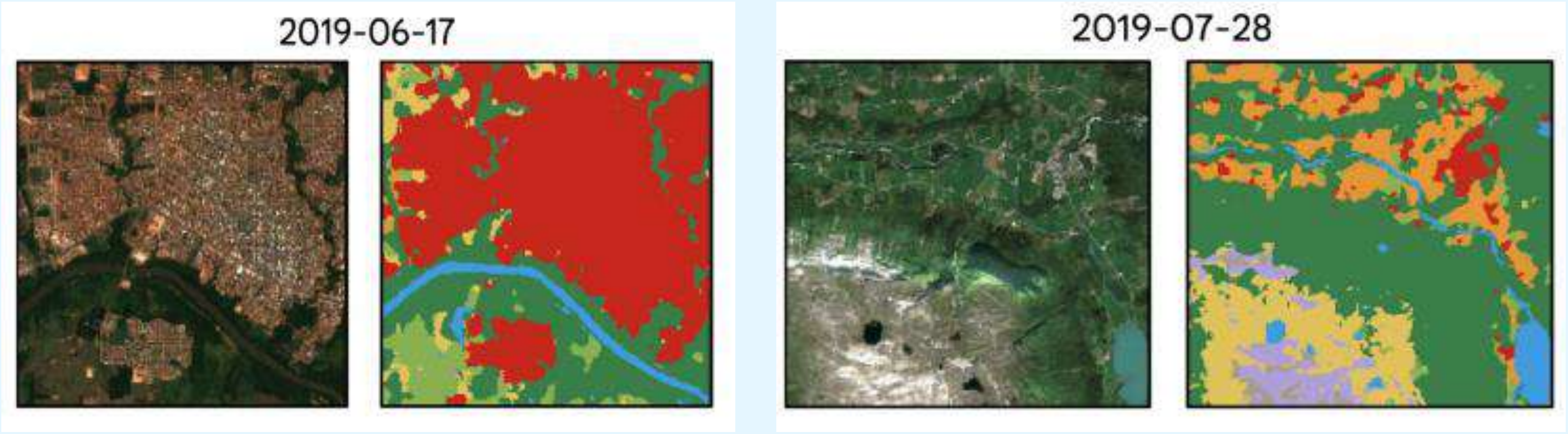


Downstream Applications

Detection of Semantic Land Cover Changes

Curating the 335,125-patch Sen4Map Sentinel-2 time-series dataset and fine-tuning Geo-FMs for semantic land-cover change detection—establishing robust benchmarks with Random Forest, pixel-based Transformers, ViTs, and Video ViTs.

.

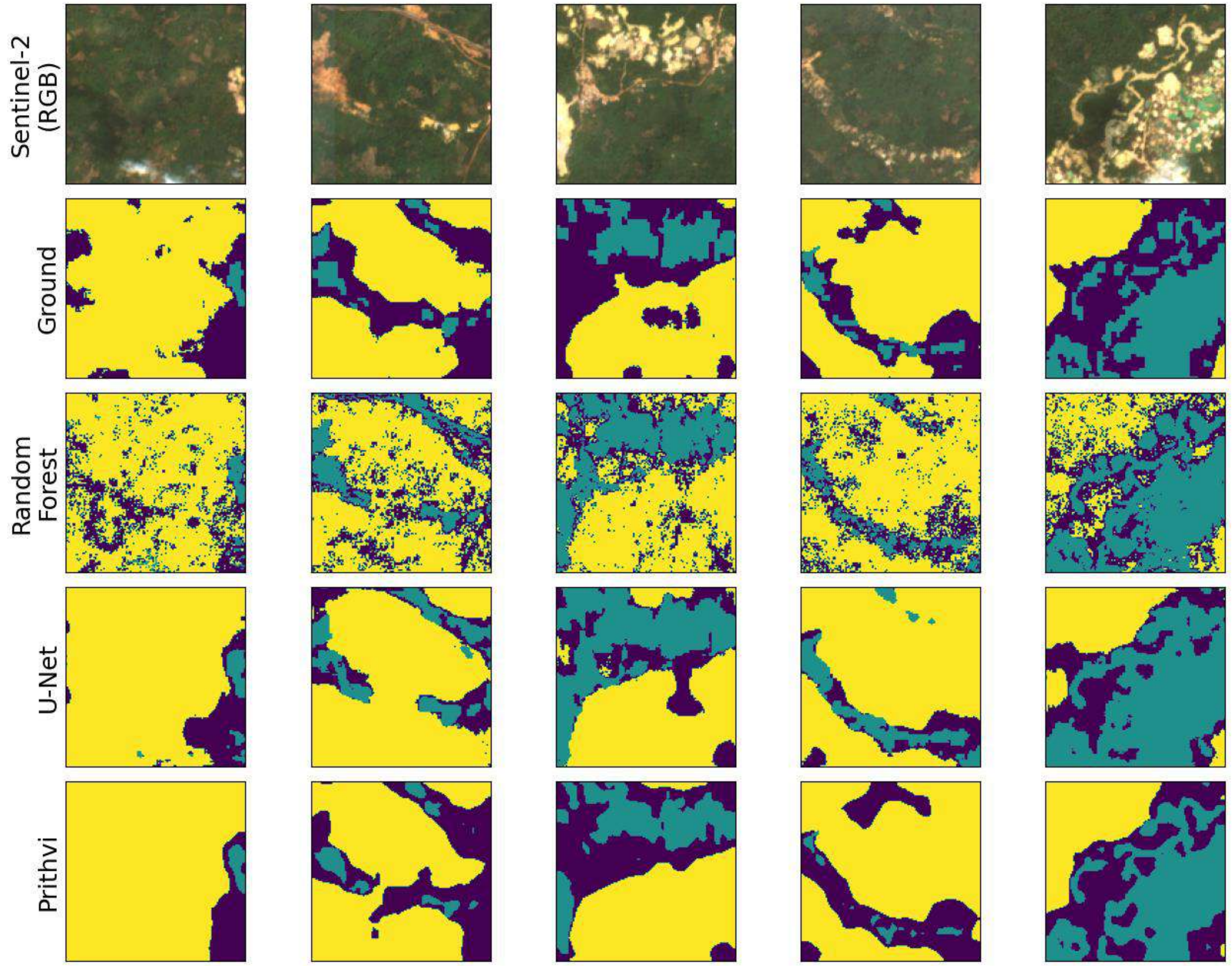


Classes	Random Forest	Transformer (pixel-based)	ViT	VViT	Prithvi-EO 1.0-100M	Prithvi-EO 2.0-300M	Prithvi-EO 2.0-600M
Artificial land	0.49	0.57	0.53	0.59	0.59	0.63	**0.64**
Bareland	0.20	0.24	0.20	0.25	0.27	0.34	**0.39**
Broadleaves	0.69	0.73	0.69	0.75	0.75	0.76	**0.77**
Conifers	0.76	0.80	0.78	0.81	0.81	0.83	**0.84**
Cropland	0.80	0.83	0.78	0.83	0.84	**0.85**	**0.85**
Grassland	0.69	0.73	0.68	0.73	0.74	0.75	**0.76**
Shrubland	0.29	0.42	0.31	0.43	0.43	**0.53**	0.52
Water	0.61	0.63	0.60	0.65	0.65	**0.68**	0.67
Wetlands	0.60	0.67	0.61	0.70	0.72	0.74	**0.75**
W.A. F-score	0.67	0.72	0.67	0.72	0.74	**0.76**	**0.76**
Overall Accuracy	0.68	0.73	0.68	0.73	0.74	**0.77**	**0.77**

FAST-EO / ESA Contract No. 4000143501/23/I-DT

Monitoring Expansion of Mining Fields into Farms

Artisanal gold-mining segmentation using Sentinel-1, Sentinel-2, and DEM data—achieving a mean IoU of 0.76 surpassing the benchmarks. Stay tuned for the improved results with TerraMind 1.0 soon!

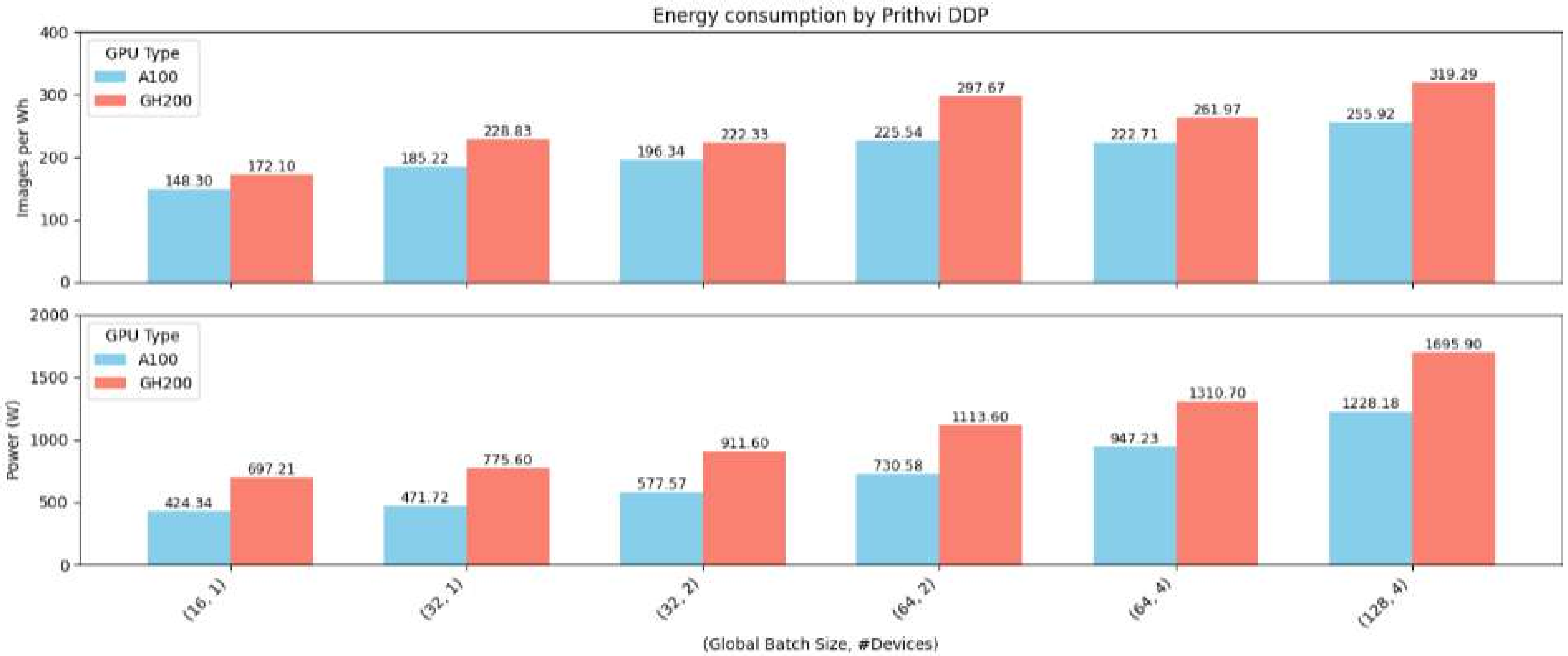


S. Ofori-Ampofo, A. Zappacosta, R. S. Kuzu, P. Schauer, M. Willberg and X. X. Zhu, "SmallMinesDS: A Multi-Modal Dataset for Mapping Artisanal and Small-Scale Gold Mines," in *IEEE Geoscience and Remote Sensing Letters*, doi: 10.1109/LGRS.2025.3566356

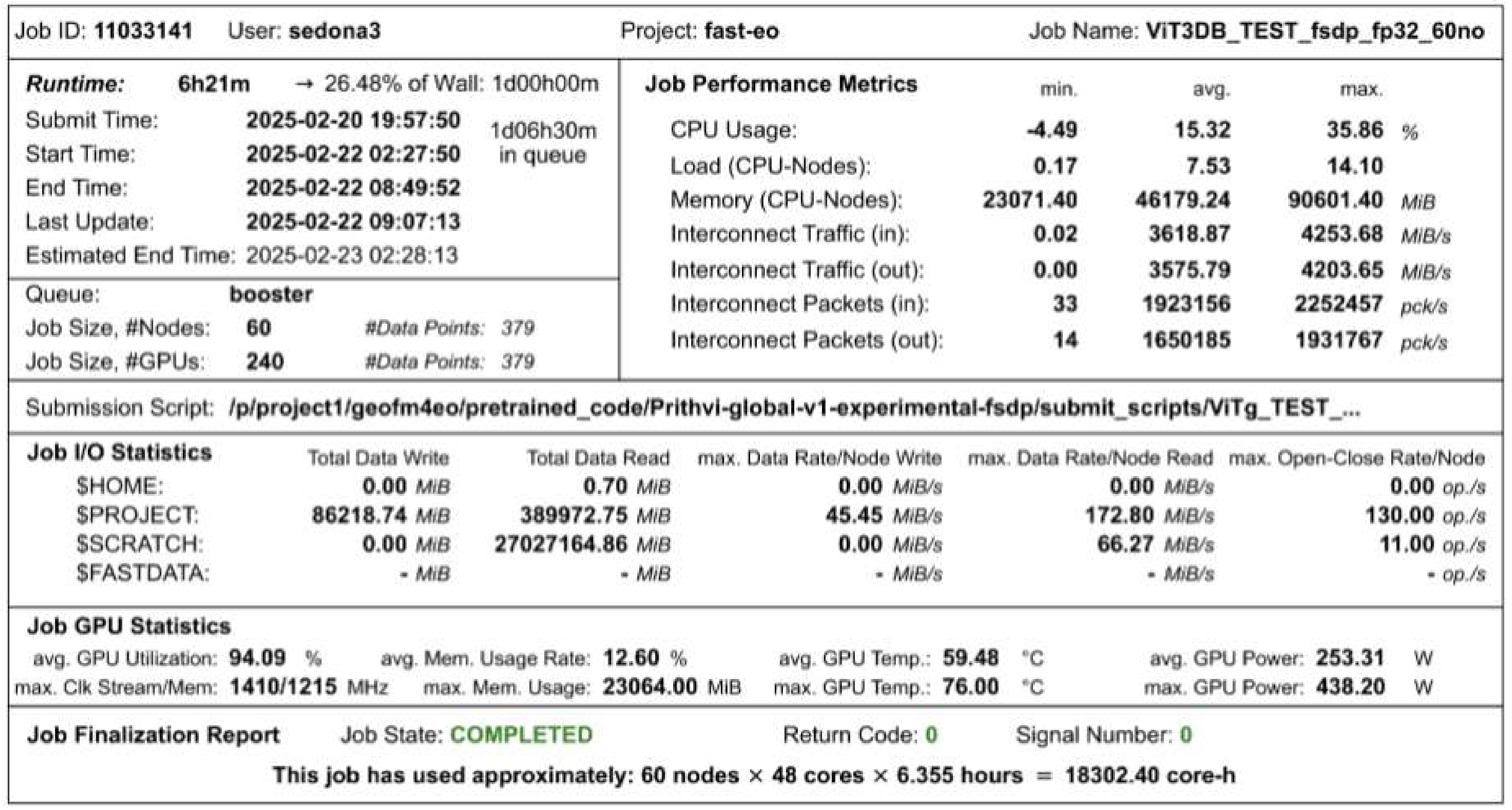
Monitoring Energy Efficiency and Resource Utilization

- HPC Utilization:
  - Standardized Metrics and Tools
    - LLview (memory, FLOPs), tools for estimation of carbon footprint (can only be order of magnitude estimation)
- Experimental set-up to measure resource utilization for pre-training, fine-tuning, prediction

Energy efficiency of Prithvi-EO, same set-up for TerraMind



LLview reports on resource utilization



Job ID: **11033141** User: **sedona3** Project: **fast-eo** Job Name: **ViT3DB_TEST_fsdp_fp32_60no**

Runtime: **6h21m** → 26.48% of Wall: 1d00h00m
Submit Time: **2025-02-20 19:57:50** 1d06h30m in queue
Start Time: **2025-02-22 02:27:50**
End Time: **2025-02-22 08:49:52**
Last Update: **2025-02-22 09:07:13**
Estimated End Time: 2025-02-23 02:28:13

Queue: **booster**
Job Size, #Nodes: **60** *#Data Points: 379*
Job Size, #GPUs: **240** *#Data Points: 379*

Job Performance Metrics	min.	avg.	max.	
CPU Usage:	-4.49	15.32	35.86	%
Load (CPU-Nodes):	0.17	7.53	14.10	
Memory (CPU-Nodes):	23071.40	46179.24	90601.40	*MiB*
Interconnect Traffic (in):	0.02	3618.87	4253.68	*MiB/s*
Interconnect Traffic (out):	0.00	3575.79	4203.65	*MiB/s*
Interconnect Packets (in):	33	1923156	2252457	*pck/s*
Interconnect Packets (out):	14	1650185	1931767	*pck/s*

Submission Script: **/p/project1/geofm4eo/pretrained_code/Prithvi-global-v1-experimental-fsdp/submit_scripts/ViTg_TEST_...**

Job I/O Statistics	Total Data Write	Total Data Read	max. Data Rate/Node Write	max. Data Rate/Node Read	max. Open-Close Rate/Node
$HOME:	0.00 *MiB*	0.70 *MiB*	0.00 *MiB/s*	0.00 *MiB/s*	0.00 *op./s*
$PROJECT:	86218.74 *MiB*	389972.75 *MiB*	45.45 *MiB/s*	172.80 *MiB/s*	130.00 *op./s*
$SCRATCH:	0.00 *MiB*	27027164.86 *MiB*	0.00 *MiB/s*	66.27 *MiB/s*	11.00 *op./s*
$FASTDATA:	- *MiB*	- *MiB*	- *MiB/s*	- *MiB/s*	- *op./s*

Job GPU Statistics
avg. GPU Utilization: **94.09** % avg. Mem. Usage Rate: **12.60** % avg. GPU Temp.: **59.48** °C avg. GPU Power: **253.31** W
max. Clk Stream/Mem: **1410/1215** MHz max. Mem. Usage: **23064.00** MiB max. GPU Temp.: **76.00** °C max. GPU Power: **438.20** W

Job Finalization Report Job State: COMPLETED Return Code: 0 Signal Number: 0
This job has used approximately: 60 nodes × 48 cores × 6.355 hours = 18302.40 core-h

Next Steps for FAST-EO

- ✓ Develop and Open-Source TerraMesh+
- ✓ Integrate Advanced SAR Capabilities
- ✓ Embed Trust and Governance Tools
- ✓ Optimize for Edge and Cloud Deployment
- ✓ Demonstrate End-to-End Operational Workflows

THANK YOU