





InstaGeo: Compute-Efficient Geospatial Machine Learning from Data to Deployment

Ibrahim Salihu Yusuf¹, Iffanice Houndayi¹, Rym Oualha¹, Mohamed Aziz Cherif¹, Kobby Panford-Quainoo ¹, and Arnu Pretorius¹

¹ InstaDeep

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

Software

- [Review](#) 
- [Repository](#) 
- [Archive](#) 

Editor: 

Submitted: 13 February 2026

Published: unpublished

License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Summary

InstaGeo is an open-source Python framework that helps users turn raw satellite imagery and in-situ observations into practical machine learning models and interactive web-map applications. It is built around three modular components: (1) automated dataset creation that converts raw satellite imagery such as Sentinel-1, Sentinel-2, and Harmonised Landsat Sentinel (HLS) into model-ready formats, (2) fine-tuning and task-specific distillation of geospatial foundation models (GFMs), and (3) browser-based inference with interactive visualisation.

By connecting these pieces, InstaGeo closes the gap between GFM research and real-world deployment. Manages the entire workflow—from geolocated field observations to production-ready predictions—so users can move from data to decisions without requiring users to integrate disparate tools or implement custom pipelines.

Statement of Need

Geospatial foundation models trained on open-access multispectral imagery have significantly improved performance on many Earth observation (EO) tasks. However, two key bottlenecks still limit the extent to which they are used in practice.

First, most published GFMs provide only model checkpoints and do not release the data pipelines needed to convert raw satellite imagery into model-ready inputs ([Cong et al., 2023](#)), ([Xiong et al., 2024](#)), ([Mendieta et al., 2023](#)), ([Hong et al., 2024](#)), ([Szwarcman et al., 2025](#)). Practitioners are left to independently implement STAC querying, temporal alignment, cloud masking, and label rasterisation. This process is time-consuming, error-prone, and often becomes the main barrier to using GFMs in real applications.

Second, the standard adaptation workflow typically fine-tunes the full encoder, regardless of how simple or complex the downstream task is. As a result, adapted models maintain their original size and computational cost, even when lighter models would be sufficient. This makes deployment on resource-constrained infrastructure unnecessarily expensive and can prevent models from being used operationally.

These issues create a divide. Domain experts in areas like agriculture or disaster response often have valuable in-situ observations but lack the machine learning expertise needed to effectively use GFMs. Conversely, ML practitioners may be comfortable with model adaptation but are less familiar with the practical challenges of working with geospatial data.

InstaGeo addresses both bottlenecks by bringing dataset construction, model compression, and deployment into a single, coherent, open-source workflow. It aims to enable collaboration between geo-experts and ML practitioners on deployable geospatial ML systems.

State of the Field

Several geospatial foundation models have already shown strong performance on standard benchmarks. SatMAE (Cong et al., 2023) introduced masked autoencoder pre-training for Sentinel-2 time series. DOFA (Xiong et al., 2024) and GFM (Mendieta et al., 2023) use teacher–student training schemes to reduce pre-training costs. SpectralGPT (Hong et al., 2024) focuses on preserving richer spectral information, and Prithvi-EO-2.0 (Szwarcman et al., 2025) achieves strong results on GEO-Bench (Lacoste et al., 2023) through large-scale HLS pre-training.

Despite these advances, existing projects generally do not provide open-source, end-to-end pipelines for turning raw satellite tiles into training chips. They also tend to produce full-sized models after fine-tuning, even when smaller models would be sufficient for downstream tasks.

TorchGeo (Stewart et al., 2025) is a widely used library that offers geospatial datasets, samplers, transforms, and pre-trained models. However, it does not cover the entire path from in-situ observations through dataset construction and model adaptation to deployed predictions. InstaGeo complements existing tools by focusing on the full process of adapting geospatial foundation models and deploying them in real-world systems.

Software Design

InstaGeo follows a modular design with three loosely coupled components: (i) a data pipeline that converts raw imagery into task-ready, multi-temporal chips, (ii) a model component for fine-tuning or distilling geospatial foundation models (GFMs) for downstream tasks, and (iii) an interactive web-map application for running inference and visualising predictions in the browser. This design is illustrated in Figure 1.

These components communicate through standardised artefacts—GeoTIFFs for imagery and PyTorch checkpoints for models—so they can be used independently or combined into a full pipeline.

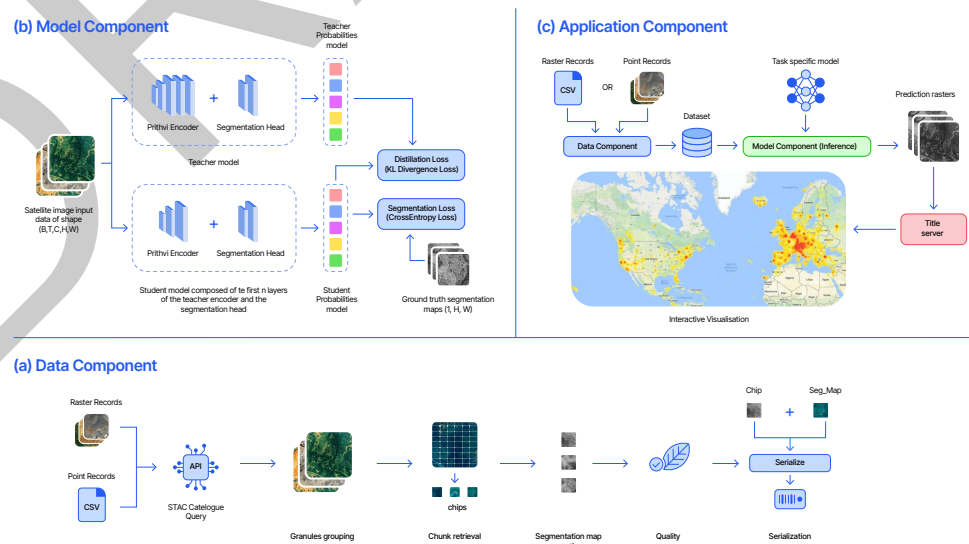


Figure 1: Overview of InstaGeo. (a) The data pipeline builds cloud-masked multi-temporal chips and labels. (b) Teacher–student distillation trains a lightweight model. (c) The application serves predictions as web-map tiles.

64 **Modularity over monolithic integration.** Users can adopt only the parts they need: geo-experts
65 can build datasets with the data pipeline, while ML practitioners can plug in existing datasets
66 and focus on model adaptation and evaluation.

67 **Data pipeline and *chip_creator*.** The *chip_creator* module turns geolocated labels into paired
68 image/label chips by querying STAC catalogues, selecting suitable acquisitions (e.g., low cloud
69 cover) and exporting multi-temporal tensors (T, C, H, W) plus segmentation maps as GeoTIFFs.
70 It supports HLS and Sentinel-1/2 and is extensible to other STAC sources.

71 **Task-specific distillation over universal fine-tuning.** InstaGeo supports teacher–student
72 distillation: a student with the first N encoder layers is trained with task loss plus KL
73 divergence to a frozen teacher, reducing inference cost with minimal accuracy loss.

74 **STAC-first data access and browser-based deployment.** The imagery is accessed through
75 STAC APIs (cloud-hosted COGs), reducing local storage but requiring network access during
76 the construction of the dataset. For deployment, predictions are served as map tiles (TiTiler)
77 and explored in a browser (see Figure 2).

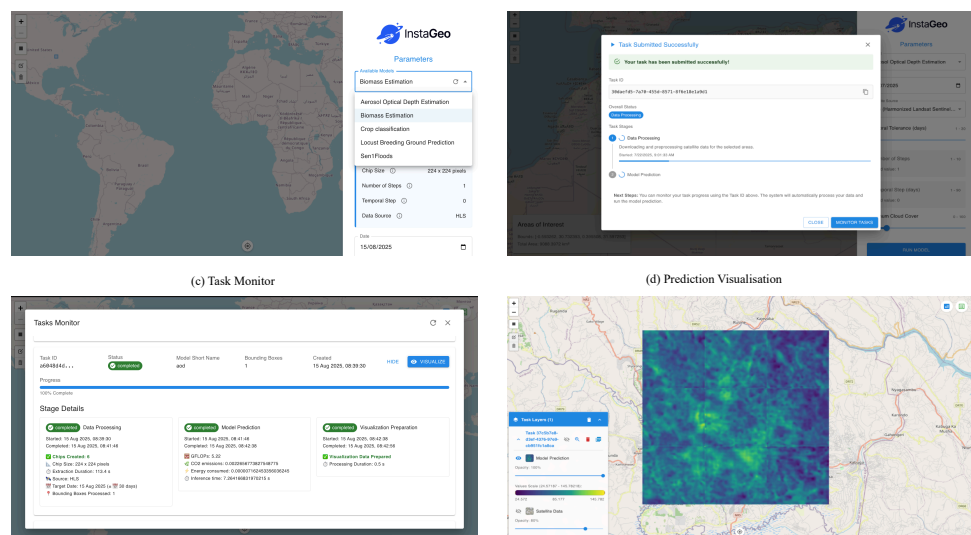


Figure 2: Application component interface. Users draw bounding boxes to define regions of interest and select model to run inference, monitor task progress, and view predictions as interactive map overlays.

78 Research Impact

79 InstaGeo has been validated across three published benchmarks: flood mapping, multi-temporal
80 crop classification, and desert locust breeding ground prediction. A key requirement for the
81 broad adoption of geospatial foundation models is a data pipeline that can accurately reproduce
82 the exact chip tensors used during training, ensuring that the performance of the original
83 model is preserved at inference time.

84 Using InstaGeo's data pipeline, we reconstructed a replica dataset from scratch that
85 matches the spatial, temporal, and spectral specifications of each original study, and
86 then re-trained the corresponding models on the replica. Across all tasks, replica models
87 reproduced the original performance within ± 2 percentage points mean Intersection over
88 Union (pp mIoU) (see Table 1)—differences minimal enough to attribute to floating point
89 precision—demonstrating that InstaGeo effectively replicates complex EO pipelines with
90 minimal performance degradation.

91 Due to the ease of use of the data pipeline, we curated a larger crop segmentation dataset,

92 achieving a new state-of-the-art mIoU of 60.65%, an improvement of 12 pp over the previous
93 baseline.

94 Task-specific distillation reduced model size by up to $8\times$ while retaining comparable
95 accuracy—for example, compressing a 389M-parameter encoder to 46M parameters with less
96 than 1 pp mIoU loss on locust prediction.

97 End-to-end, the framework reduces the data-to-deployment cycle to under nine hours on
98 standard hardware, enabling rapid iteration for time-sensitive applications such as emergency
99 flood response. Full benchmark results and reproduction scripts are available in the project
100 [repository](#).

Table 1: InstaGeo reproduces unpublished data pipelines. For each task, we report the performance of (i) the Baseline performance reported in the corresponding study, (ii) the InstaGeo-Baseline model trained on the authors' data but using InstaGeo's model component, and (iii) the InstaGeo-Replica model trained on a dataset reconstructed entirely with InstaGeo. The flood mapping replica has a version derived from HLS and another derived from Sentinel-2.

Task	Model	GFM	mIoU (std)	Acc	mF1 (std)	ROC- AUC (std)
Flood Mapping	Baseline	Prithvi-V1-100M	88.3 (0.3)	-	97.3 (0.1)	-
	InstaGeo-Baseline	Prithvi-V1-100M	88.53	97.24	93.71	99.16
	InstaGeo-Replica (HLS)	Prithvi-V1-100M	85.40	96.39	91.78	97.15
	InstaGeo-Replica (S2)	Prithvi-V1-100M	87.80	97.07	93.26	97.61
Multi-Temporal Crop Segmentation (US)	Baseline	Prithvi-V1-100M	42.70	60.7	-	-
	InstaGeo-Baseline	Prithvi-V1-100M	48.07	65.77	64.34	95.79
	InstaGeo-Replica	Prithvi-V1-100M	47.87	66.10	64.19	95.82
Locust Breeding Ground Prediction	Baseline	Prithvi-V1-100M	-	83.03	81.53	-
	InstaGeo-Baseline	Prithvi-V1-100M	71.51	83.39	83.39	86.74
	InstaGeo-Replica	Prithvi-V1-100M	73.30	84.60	84.60	88.66

101 AI usage disclosure

102 No generative AI tools were used to generate the scientific claims, experimental results, or
103 evaluations described in this paper. Generative AI tools have been used for minor language
104 editing and formatting, and all text was reviewed by the authors for accuracy and clarity.

105 References

106 Cong, Y., Khanna, S., Meng, C., Liu, P., Rozi, E., He, Y., Burke, M., Lobell, D. B., & Ermon,
107 S. (2023). SatMAE: Pre-training transformers for temporal and multi-spectral satellite

- 108 imagery. *arXiv Preprint arXiv:2207.08051*. <https://arxiv.org/abs/2207.08051>
- 109 Hong, D., Zhang, B., Li, X., Li, Y., Li, C., Yao, J., Yokoya, N., Li, H., Ghamisi, P., Jia, X.,
110 Plaza, A., Gamba, P., Benediktsson, J. A., & Chanussot, J. (2024). SpectralGPT: Spectral
111 remote sensing foundation model. *IEEE Transactions on Pattern Analysis and Machine*
112 *Intelligence*, 46(8), 5227–5244. <https://doi.org/10.1109/TPAMI.2024.3362475>
- 113 Lacoste, A., Lehmann, N., Rodriguez, P., Sherwin, E. D., Kerner, H., Lütjens, B., Irvin, J. A.,
114 Dao, D., Alemohammad, H., Drouin, A., Gunturkun, M., Huang, G., Vazquez, D., Newman,
115 D., Bengio, Y., Ermon, S., & Zhu, X. X. (2023). Geo-bench: Toward foundation models
116 for earth monitoring. *arXiv Preprint arXiv:2306.03831*. <https://arxiv.org/abs/2306.03831>
- 117 Mendieta, M., Han, B., Shi, X., Zhu, Y., & Chen, C. (2023). Towards geospatial foundation
118 models via continual pretraining. *arXiv Preprint arXiv:2302.04476*. [https://arxiv.org/abs/](https://arxiv.org/abs/2302.04476)
119 [2302.04476](https://arxiv.org/abs/2302.04476)
- 120 Stewart, A. J., Robinson, C., Corley, I. A., Ortiz, A., Lavista Ferres, J. M., & Banerjee, A. (2025).
121 *TorchGeo: Deep learning with geospatial data*. 11(4). <https://doi.org/10.1145/3707459>
- 122 Szwarcman, D., Roy, S., Fraccaro, P., Gislason, P. E., Blumenstiel, B., Ghosal, R., Oliveira,
123 P. H. de, Almeida, J. L. S., Sedona, R., Kang, Y., Chakraborty, S., Wang, S., Gomes, C.,
124 Kumar, A., Truong, M., Godwin, D., Lee, H., Hsu, C.-Y., Asanjan, A. A., ... Moreno, J. B.
125 (2025). Prithvi-EO-2.0: A versatile multi-temporal foundation model for earth observation
126 applications. *arXiv Preprint arXiv:2412.02732*. <https://arxiv.org/abs/2412.02732>
- 127 Xiong, Z., Wang, Y., Zhang, F., Stewart, A. J., Hanna, J., Borth, D., Papoutsis, I., Le Saux,
128 B., Camps-Valls, G., & Zhu, X. X. (2024). Neural plasticity-inspired multimodal foundation
129 model for earth observation. *arXiv Preprint arXiv:2403.15356*. [https://arxiv.org/abs/2403.](https://arxiv.org/abs/2403.15356)
130 [15356](https://arxiv.org/abs/2403.15356)