Who Decides the Risk? Permafrost Thaw Risk Assessment in Wainwright Community, Alaska

Emine Senkardesler [1]*, Anna Liljedahl [2], Elhan Ersoz [3], and Simon Zwieback [4]

[1] Department of Informatics, University of Illinois Urbana-Champaign, Urbana, IL, USA

[2] Woodwell Climate Research Center, Falmouth, MA, USA

[3] Department of Crop Sciences, University of Illinois Urbana-Champaign, Urbana, IL, USA

[4] University of Alaska Fairbanks, Fairbanks, AK, USA

* emine2@illinois.edu

Keywords: arctic infrastructure, machine learning, permafrost, remote sensing, risk assessment

Introduction and Research Question

Risk can be defined as a function of hazard, exposure, and vulnerability. Across Arctic Alaska, permafrost thaw is the hazard that threatens Arctic settlements, putting much of the region's infrastructure at risk by mid-century (Hjort et al., 2018). Community infrastructure is the exposure: buildings, roads, water systems, and fuel tanks that cannot be cheaply replaced when the ground beneath them subsides. Wainwright, an Inupiat village on the Chukchi Sea coast, is already facing this problem. Buildings tilt as ice wedges melt, roads buckle, and coastal erosion advances toward critical facilities. In Kipnuk, hundreds of residents displaced by Typhoon Halong remain in temporary housing. Both communities face the same urgent question: where can we safely live, build, and expand? The answer is a risk map showing which patches of ground are stable and which are not. But producing such a map means combining heterogeneous hazard datasets, each at a different resolution, measurement scale, and level of certainty, into a single defensible product. Risk is computed per cell by weighted overlay: each hazard layer is assigned a weight, the layers are summed, and the result is classified into risk zones. This is transparent and reproducible, but the weights are rarely validated. This motivates our main research question: how should ice-wedge polygons, terrain features, and drained lake basins be weighted to map permafrost risk?

To answer this question, we compare three approaches: (1) a rule-based weighting scheme, (2) a convolutional neural network (CNN) that learns weights from spatial patterns, and (3) Alpha Earth Embeddings (AEE) that provide independent satellite-derived validation, and analyze where they agree, where they disagree, and what those disagreements reveal about the underlying risk structure. Two secondary questions follow: (a) how well can satellite-derived embeddings substitute for expert hazard layers in communities where ground-truth data is unavailable, and (b) how effectively can a large language model (LLM) translate per-cell risk data into plain-language narratives for community planners?

Study Area and Data Integration

A 50-meter resolution grid was created over a 10-mile buffer surrounding Wainwright, Alaska. This grid is the spatial unit for all analyses. Each cell was attributed with multiple thematic layers harmonized through overlay analysis and zonal statistics, including: (1) drained lake basin fractions from a classification derived from Landsat-8 imagery and ArcticDEM for the North Slope of Alaska (Bergstedt et al., 2021); (2) terrain features (slope, curvature, and topographic wetness index) derived from the 2 m ArcticDEM (Porter et al., 2018); (3) ice-wedge polygon presence from CNNs applied to sub-meter WorldView-2 imagery (Witharana et al., 2023); (4) a Sentinel-1 InSAR-derived mean excess-ice content product for Wainwright, where values above 0.2 indicate ice-rich, thaw-sensitive ground (unpublished data, co-author S. Zwieback); (5) contaminated site status from the Alaska Department of Environmental Conservation; (6) infrastructure footprints from the HABITAT deep-learning pipeline on 0.5 m Maxar imagery (Manos, 2024); (7) coastal erosion forecast zones (Buzard et al., 2021); (8) NDWI-derived trough ponds, which form in the degrading troughs of ice-wedge polygon networks, summarized as pond presence and area fraction per cell; and (9) geolocated community observations from the Local Environmental Observer Network (LEO Network, n.d.), used as attention flags and ground truth for the data layers. These inputs span both physically observable features and human-knowledge features (contamination records, erosion assessments, local observations). The 50 m cell size harmonizes inputs ranging from 2 m (ArcticDEM) to 30 m (Landsat) into a consistent unit while matching the building-and-parcel scale of siting decisions; the 10-mile buffer covers the plausible search area for expansion and relocation.

Methods

Stage 1 — Rule-Based Risk Index. A composite risk index was constructed by assigning weights to hazards with multiplicative modifiers for infrastructure and contamination or erosion (risk set to maximum). This produces a transparent baseline. To validate this baseline, a Monte Carlo analysis recomputed the hazard core under 5,000 Dirichlet weight vectors centered on the expert weights; about 99% of cells keep their risk class, so the map is robust to weight uncertainty.

Stage 2 — CNN Encoder-Decoder. To test whether the data support a different weighting, we designed a convolutional encoder-decoder (TensorFlow/Keras). The attribute layers were rasterized into a multi-channel image at 50 m resolution. The encoder compresses through two convolutional layers (32 and 16 filters) with max-pooling; the decoder reconstructs a single-channel risk prediction via transposed convolutions. Trained for 30 epochs on an 80/20 pixel-level split, the model achieves RMSE = 0.146, MAE = 0.115, and Pearson $r = 0.632$. To diagnose what the CNN learned differently, we performed a spatial disagreement analysis and a channel ablation study (zeroing each input layer to quantify its contribution).

Stage 3 — Alpha Earth Embeddings (AEE). To provide validation independent of the expert-defined layers, we extracted AEE (Google DeepMind, 2025) for each grid cell. AEE are per-pixel embedding vectors from the AlphaEarth Foundations model, which distills multi-sensor satellite image time series into a compact representation of spectral, textural, and temporal landscape context, built with no knowledge of our hazard layers. Because they capture the surface expression of thaw processes such as polygonized terrain and drained basins, they offer an independent check on expert-defined inputs. We then tested whether individual AEE bands can predict individual hazard layers.

Stage 4 — LLM Risk Narratives. To make risk assessments accessible to non-technical decision-makers, we developed a narrative generation layer using a large language model. For each grid cell, the risk score, input feature values, expert-AI agreement status, and attribution data are formatted into a structured prompt. The LLM produces a 3–4 sentence plain-language explanation of the risk level, its drivers, and its planning implications. Narratives are currently evaluated through domain-expert review before sharing with community stakeholders; a systematic evaluation with community planners is ongoing work.

Results

Rule-based weights and the CNN agree on 72% of cells, validating that domain knowledge captures the dominant risk patterns. The 28% disagreement is systematic, not random. Table 1 summarizes the three mapping approaches and the narrative layer.

Approach	**Input data**	**Output**
Rule-based index (expert weighting)	All hazard and knowledge layers	Cell-level risk index (0–1)
CNN encoder-decoder (supervised learning)	Rasterized multi-channel layer stack	Cell-level risk prediction (0–1)
AEE (foundation model embeddings)	Satellite imagery only	Per-cell embedding vectors
LLM narratives (generative language model)	Risk score, features, agreement status	Plain-language cell explanation

Table 1: Comparison of the three mapping approaches and the narrative layer.

In 1,455 cells (10.8%), the CNN predicts significantly higher risk than the expert. These cells are characterized by high drained lake fraction (mean 0.62) but low ice-wedge presence (0.52); the equal-weight scheme assigned them a mean risk of only 0.17, undervaluing drained lake terrain. In 2,304 cells (17.0%), the CNN predicts lower risk. These cells have near-universal ice-wedge presence (0.99), but the CNN learned that ice wedges alone, without converging hazard signals from other layers, do not warrant the high penalty (mean 0.60) that the expert's additive rules produced. Channel ablation confirms this pattern: the CNN's learned feature importance ranking diverges from the expert's equal weights. Some hazard layers contribute roughly twice the predictive influence of others. An independent structure analysis supports this: unsupervised clustering of the standardized hazard layers, using no weights or labels, recovers landscape types that align with the expert hazard classes (χ^2 test, $p < 0.001$).

Embeddings corroborate these findings independently. AEE bands achieve moderate-to-strong correlations with physically observable hazard layers — ice-wedge polygons (r = 0.34), drained lakes (r = 0.31), and infrastructure (r = 0.33) — but cannot predict contamination status (r =

0.04) or erosion forecasts (r = 0.22). This asymmetry establishes an empirical boundary: satellite-derived foundation models can serve as proxies for physical hazards in data-poor communities, but regulatory and planning layers must still come from local knowledge and ground studies.

LLM-generated narratives make these results actionable. For a disagreement-zone cell, the model produces: *"This area sits within a drained lake basin. Although expert rules rated it as low risk because no ice-wedge polygons were detected, the AI model flags elevated risk — drained lake terrain makes this ground vulnerable to thaw settlement. A geotechnical assessment is recommended before construction."* These narratives make the most uncertain zones — where expert and AI diverge — interpretable for planners. Figure 1 shows the resulting risk map for the Wainwright community.

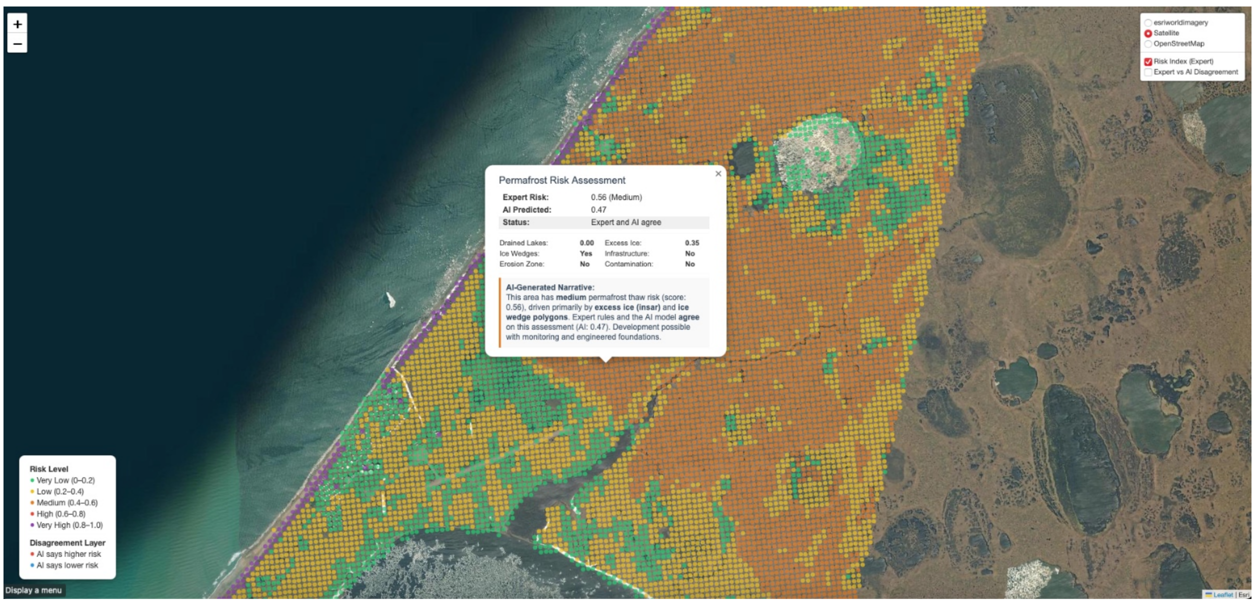

Figure 1: Risk map demonstration for the Wainwright community showing cell-level risk scores with expert-AI agreement assessment.

Discussion and Conclusion

Arctic communities need to know where it is safe to live, build, and expand on thawing permafrost. We make three contributions. First, the **grid-based approach produces readable, cell-level risk maps** that community planners can use directly for relocation siting and neighborhood extension as well as which parts of the community are already experiencing the thaw related problems. Each 50 m cell carries a risk score, a confidence assessment (expert-AI agreement), and a plain-language explanation of what drives the risk. Second, **foundation model embeddings can predict physically observable hazard layers** (ice wedges, drained lakes) from satellite imagery alone, but not human-knowledge layers (contamination, erosion forecasts), establishing what AI can substitute in data-sparse communities and what still requires local expertise. Third, **LLM-generated cell-level narratives** close the last mile between quantitative risk grids and community action, giving tribal councils and emergency managers text they can act on.

The 50 m grid framework is reproducible and is currently being extended to Kipnuk, Alaska, to test whether the approach transfers across permafrost regimes and data availability conditions.

Acknowledgements

This work was supported by the team working with the Permafrost Discovery Gateway. We thank Dr. Nicolas Federico Martin for their valuable guidance and support. We are grateful to the Local Environmental Observer (LEO) Network and the Alaska Native Tribal Health Consortium for providing community-based observation data, and to the Arctic Data Center for data archiving and access support. We also thank the many individuals, teams, and organizations who created and shared the data used in this study.

References

Bergstedt, H., Jones, B., Hinkel, K., Farquharson, L., Gaglioti, B., et al. (2021). Drained Lake Basin classification based on Landsat-8 imagery, North Slope, Alaska 2014–2019 [Data set]. Arctic Data Center. doi: 10.18739/A2K35MF71.

Buzard, R. M., Turner, M. M., Miller, K. Y., Antrobus, D. C., and Overbeck, J. R. (2021). Erosion exposure assessment of infrastructure in Alaska coastal communities. Report of Investigation 2021-3, Alaska Division of Geological & Geophysical Surveys. doi: 10.14509/30672.

Google DeepMind (2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv:2507.22291.

Hjort, J., Karjalainen, O., Aalto, J., Westermann, S., Romanovsky, V. E., Nelson, F. E., Etzelmüller, B., and Luoto, M. (2018). Degrading permafrost puts Arctic infrastructure at risk by mid-century. *Nature Communications*, 9:5147. doi: 10.1038/s41467-018-07557-4.

LEO Network (n.d.). Local Environmental Observer (LEO) Network. Alaska Native Tribal Health Consortium. Retrieved from `https://www.leonetwork.org/`.

Manos, E. (2024). Buildings and roads detected from 0.5 meter resolution Maxar satellite imagery, Alaskan communities, 2018–2023 [Data set]. Arctic Data Center.

Porter, C., Morin, P., Howat, I., et al. (2018). ArcticDEM [Data set]. Harvard Dataverse. doi: 10.7910/DVN/OHHUKH.

Witharana, C., Udawalpola, M. R., Liljedahl, A. K., et al. (2023). Ice-wedge polygon detection in satellite imagery from pan-Arctic regions, Permafrost Discovery Gateway, 2001–2021 [Data set]. Arctic Data Center.