Masked Mineral Modeling: Continent-Scale Mineral Prospecting via Geospatial Infilling

Sujay Nair[1,2], Evan Austen Coleman[1], Sherrie Wang[1], Elsa Olivetti[1]

[1] Massachusetts Institute of Technology
[2] Georgia Institute of Technology

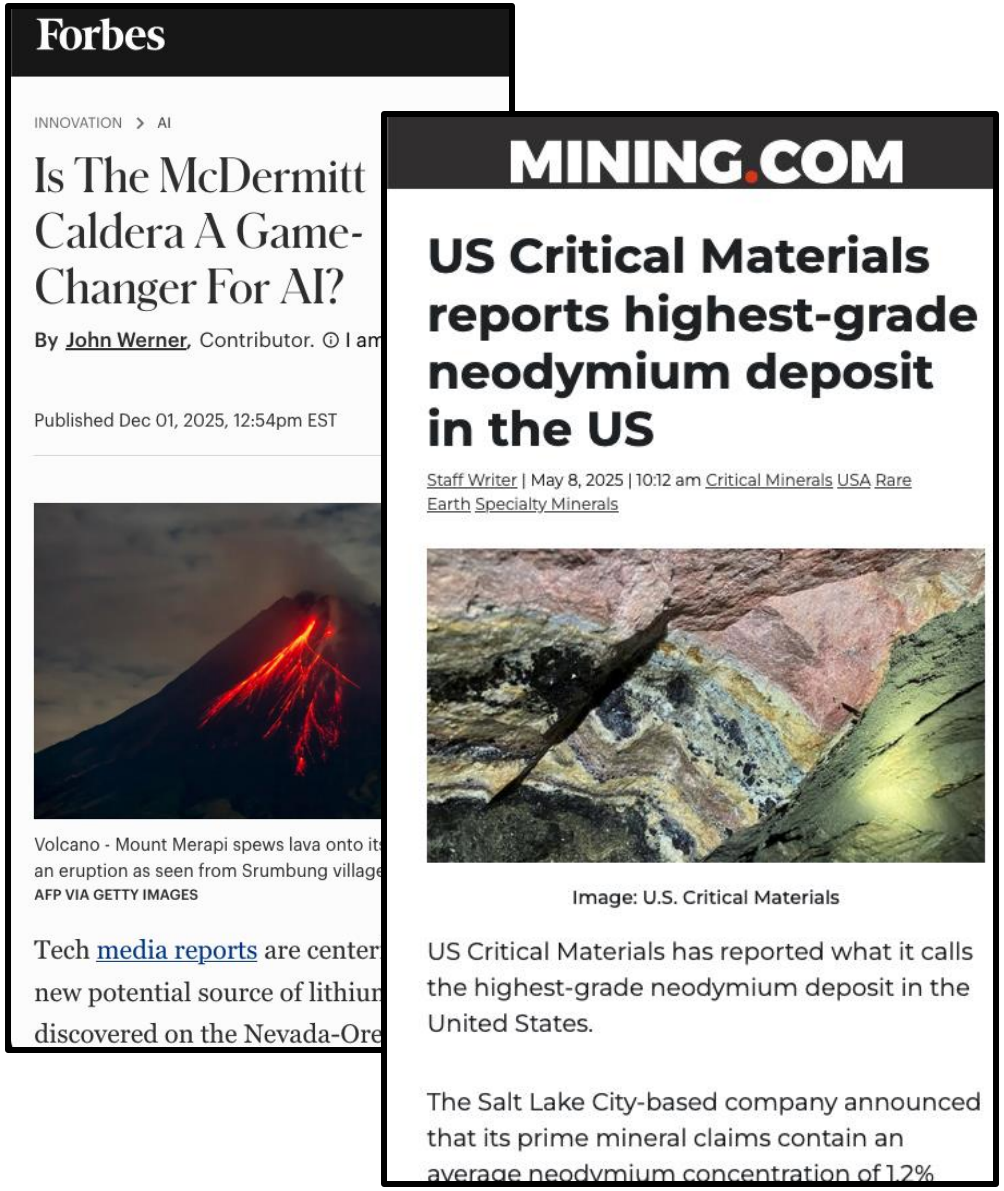
MOTIVATION

Achieving the energy transition and addressing climate change will require unprecedented quantities of mineral resources to be extracted.

At the same time, many of these minerals have only recently gained economic value.

Recent discoveries suggest that even regions which have been thoroughly surveyed may contain large deposits of climate-critical resources.

Could it be that resource records are incomplete because climate change is changing what we consider a resource?



Forbes

INNOVATION > AI

Is The McDermitt Caldera A Game-Changer For AI?

By John Werner, Contributor.

Published Dec 01, 2025, 12:54pm EST

Volcano - Mount Merapi spews lava onto its
an eruption as seen from Srumbung village
AFP VIA GETTY IMAGES

Tech media reports are center
new potential source of lithium
discovered on the Nevada-Ore

MINING.COM

US Critical Materials reports highest-grade neodymium deposit in the US

Staff Writer | May 8, 2025 | 10:12 am Critical Minerals USA Rare Earth Specialty Minerals

Image: U.S. Critical Materials

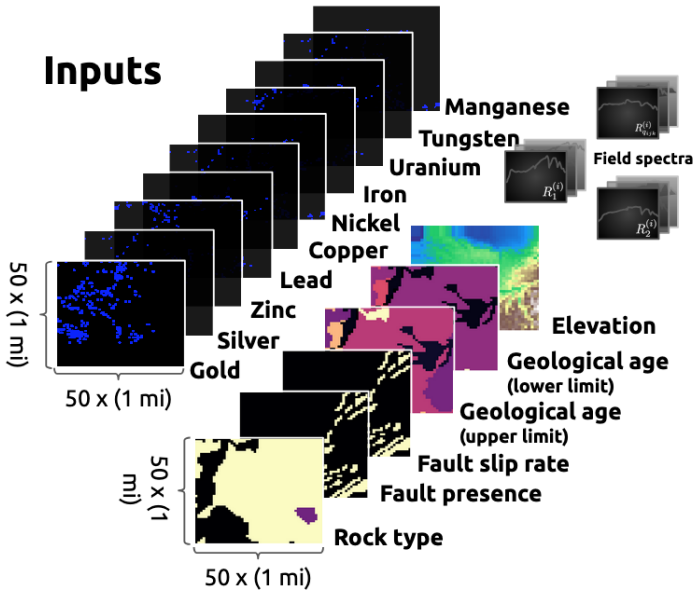
US Critical Materials has reported what it calls the highest-grade neodymium deposit in the United States.

The Salt Lake City-based company announced that its prime mineral claims contain an average neodymium concentration of 1.2%

BACKGROUND

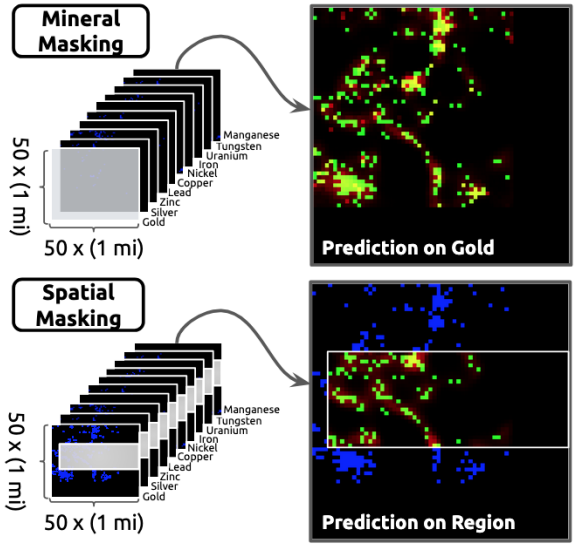
Mineral records are sparse and clustered, but incomplete global datasets are available.

Can we mine copresence correlations to uncover resources hidden by limited or biased recordkeeping?



The missing or ablated nature of these records presents a generative problem stencil which merits exploration.

We consider the task of artificially discarding entire mineral layers or contiguous regions and recovering them by exploiting diverse datasets.

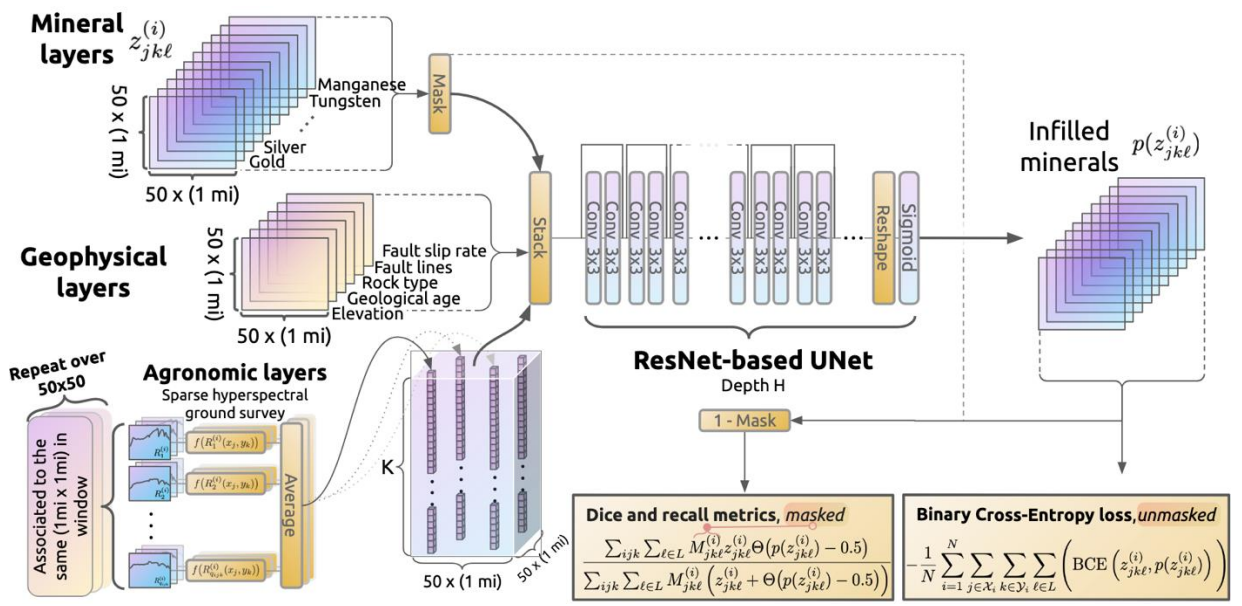


METHOD OVERVIEW

M3 Architecture: ResNet-backed UNet admitting auxiliary inputs

Trained to infill masked binary class labels indicating the confirmed presence of 10 mineral species.

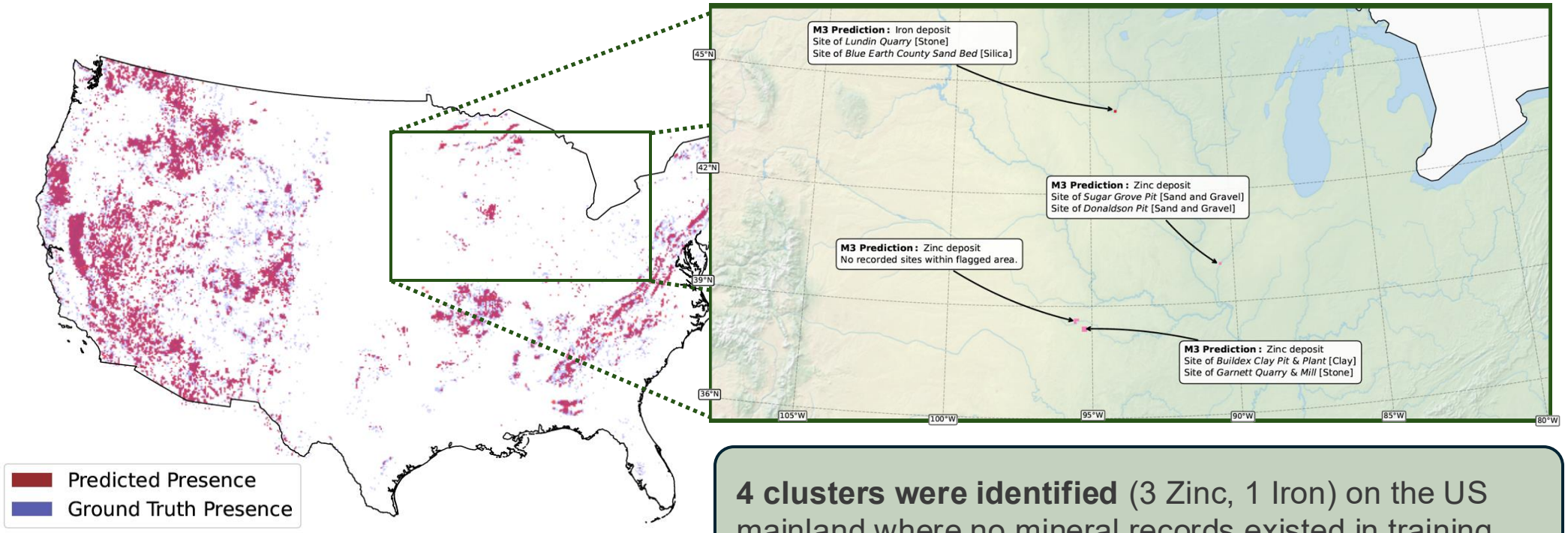
Dice and recall combined with BCE loss to handle missing true negatives in the dataset.



ResNet-backed UNet out-performs both ViT and kriging

Spatial inductive bias is superior, likely due to sparse clustering of resources. Coarse-graining by MaxPooling inputs destroys any ViT-ResNet gap.

AUXILIARY FEATURES EXPAND MODEL COVERAGE



Addition of geophysical and agronomic data enables model evaluation in regions with no logged resource presence.

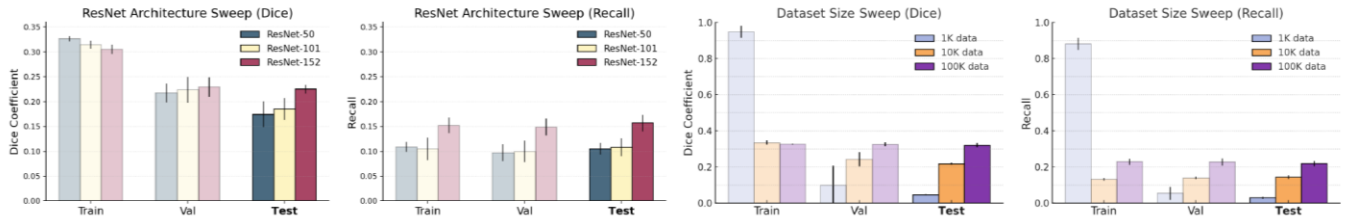
We aggregate model predictions into a map by evaluating the best-performing model configuration 150× over entire CONUS [3 seeds × 50 evals].

4 clusters were identified (3 Zinc, 1 Iron) on the US mainland where no mineral records existed in training data. 3 match known surface mines for sand, gravel, clay, and silica. Zinc has Elevated criticality, above that of e.g. Lithium (which is Moderate).

$$\text{Dice} = 0.28 \pm 0.02$$
$$\text{Recall} = 0.14 \pm 0.01$$

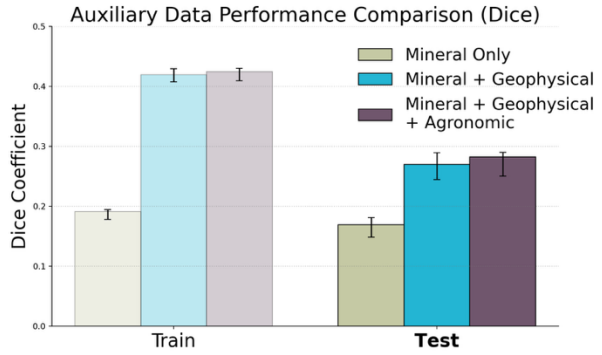
($N = 10K,\ A = 0.8,\ H = 152,\ K = 64,\ 30K$ gradient steps)

SCALABILITY LEVERS



General improvement with model and dataset size.

This motivates time-intensive tests over global datasets.



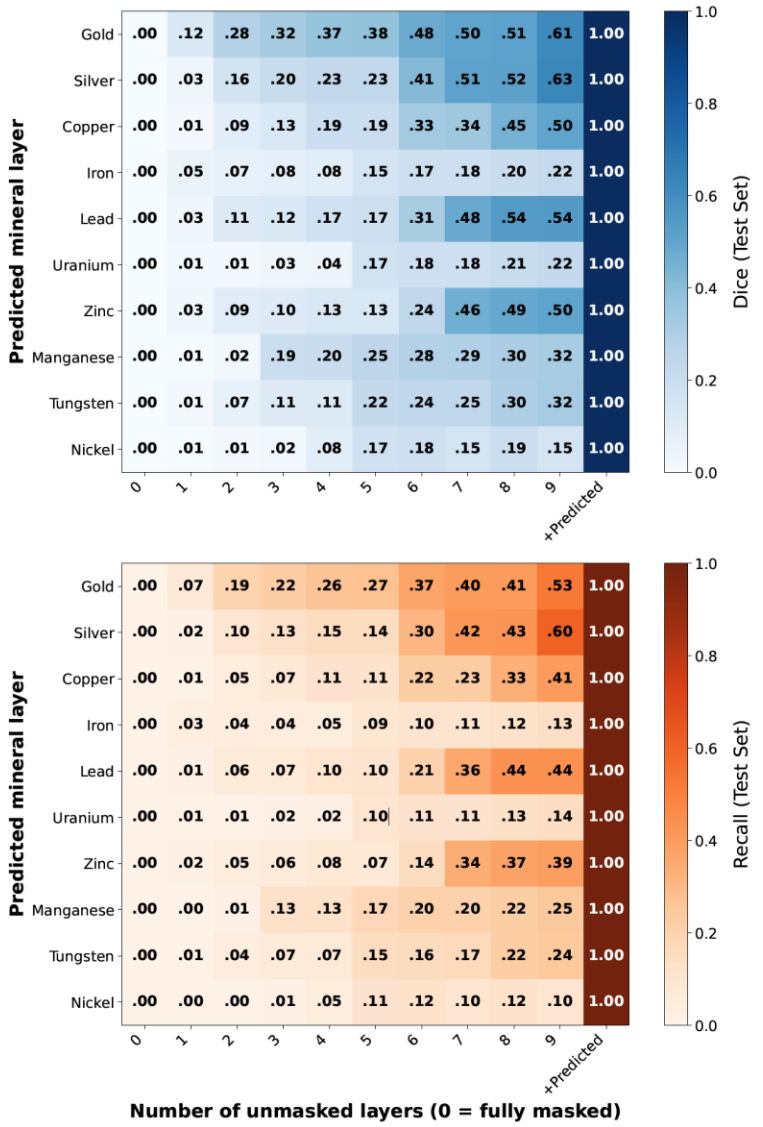
Significant gains due to geophysical inputs.

Sparse hyperspectral survey data generally improved training stability.

Learned features broadly interdependent across species.

Mineral prospecting literature has only co-inferred ≤ 4 species at once.

Poor performance for Fe, U is possibly due to leakage of processing sites into mine sites and intentional scrambling for national security.



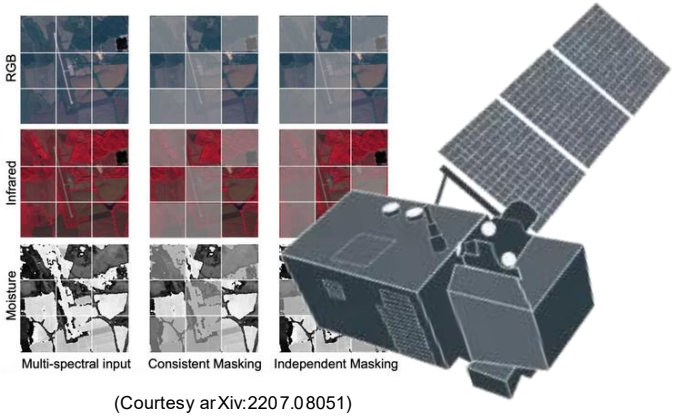
FUTURE WORK

Dedicated analysis of feature impacts from remote sensing data

Mining activity is easily detected by satellite. It is unclear to what extent such features might bias M3 with respect to resources which have already been extracted.

Harmonizing global geophysical datasets is nontrivial

Databases such as Mindat have global coverage of resource presence, but some work will be required to cross-reference it against fault line data, lithography, hydrography, and elevation data.

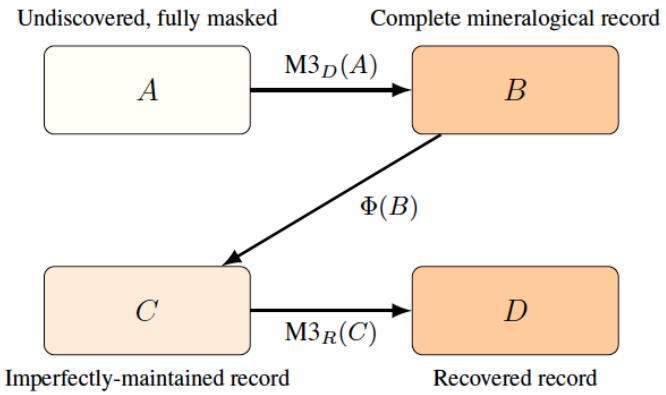


(Courtesy arXiv:2207.08051)

mindat.org®

Relaxing rigid masking via end-to-end inference

Using 2 M3 instances enables us to back out the structure of the ablation impacting mineralogical recordkeeping. This is an extension of positive-unlabeled (PU) learning with connections to diffusion.



$$\mathcal{L}_{\mathrm{M3}} = \mathrm{BCE}\left(\mathrm{M3}_R\left(\Phi\left(\mathrm{M3}_D(A)\right)\right), \{\mathrm{M3}_D(A) > T\}\right)$$

$$\mathcal{L}_{\Phi} = \mathrm{BCE}\left(\Phi\left(\mathrm{M3}_D(A)\right), \{z_{jk\ell}^{(i)}\}\right)$$

$$\mathcal{L} = \mathcal{L}_{\mathrm{M3}} + \beta\mathcal{L}_{\Phi}$$