

# AlphaEarth: Geo-Risk using Earth Observation and ML

Shubhankar Mahanta

mailto:hello@spass.uk

## Challenge Tackled

After a disaster, slow paperwork delays insurance claims and heightens distress for victims and their families. Physical damage assessment is often impossible immediately due to inaccessible sites, significantly prolonging response times. Inspired by Google DeepMind’s AlphaEarth initiative, this project leverages satellite imagery, terrain data, and real-time weather to deliver instant geo-risk estimation. Currently trained on synthetic data, the model generates risk scores based on location’s weather and environmental conditions.

## Approach and Model

The system is centered around a compact deep learning architecture called **CompactGeoEmbed**, a lightweight CNN derived from MobileNetV2. It fuses optical satellite imagery, elevation maps, and environmental factors such as precipitation, humidity, and wind speed to estimate a composite risk score for any coordinate. The inference pipeline retrieves Sentinel imagery and elevation data through Google Earth Engine, processes it locally using a quantized PyTorch model under 10MB, and presents the results via a Gradio web interface hosted on Hugging Face Spaces. The design supports CPU-only inference, ensuring low deployment cost and accessibility.

## Data and Implementation

The model was trained on synthetically curated geospatial data combining open datasets such as MODIS (fire burn area), SMAP (soil moisture), and SRTM (terrain elevation). Although limited by API quotas and inconsistent spatial resolutions, careful sampling, caching, and cross-sensor alignment enabled construction of a functional dataset suitable for experimentation. The model generalizes well across diverse landscapes and maintains high consistency under quantized inference.

## Challenges

Development revealed several difficulties, including user memory limit while querying on Earth Engine for sampling and getting a balanced distribution of

data, harmonizing multi-sensor resolutions (250m vs. 1km), and a general scarcity of open hazard datasets from agencies like UNEP or SEDAC. Extreme denial rate while fetching forecast data from <https://open-meteo.com/>. Despite these limitations, the model learned to predict stable inference points, end-to-end inference using only geographic coordinates and weather data.

## Time Allocation

Data preparation and testing: 8 hours. Model training and visualization: 3.5 hours. Deployment and hosting optimization: 2.5 hours. Code structuring and documentation: 2 hours.

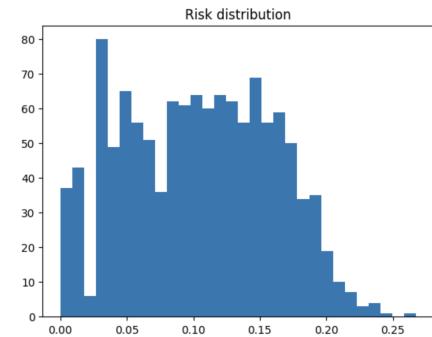


Figure 1: Geo-Risk Distribution on 1200 samples

## Project Summary

The final prototype demonstrates a real-time geo-risk estimation framework that integrates Earth observation data with machine learning for actionable insurance intelligence. Its architecture emphasizes reproducibility, interpretability, and low computational overhead, serving as a foundation for scalable systems that can automate damage assessment, enhance transparency, and accelerate post-disaster recovery.

## Live Demo

Hosting using Huggingface CPU Spaces  
[/spaces/Pingsz/3rd-hack-nation](https://spaces.huggingface.co/Pingsz/3rd-hack-nation)

Project Reproducible version hosted on GitHub  
[/sp4s-s/3rd-hack-nation](https://github.com/sp4s-s/3rd-hack-nation)