White Paper: GeoAuPredict G.A.P - AI-Driven Geospatial Prediction of Gold Deposits

Edward Calderón
Architect & Researcher in AI and Blockchain
ecalderon@unal.edu.co

Version 1.0 — October 2025

Abstract

This project develops an open-source geospatial artificial intelligence system for the detection and prediction of gold deposits in Colombia and worldwide, integrating geochemical, geophysical, satellite, and drilling data while applying deep learning techniques and 3D spatial models to generate probability maps of gold presence, enabling more efficient, sustainable, and auditable decision-making in mineral exploration. Traditional mineral exploration relies heavily on physical drilling campaigns, making it costly, high-risk, time-consuming, and environmentally impactful, with existing open studies on gold prospectivity applying artificial intelligence and geospatial data still lacking widely replicated, auditable, and generalizable models that integrate multiple data sources (geochemical, satellite, geophysical, and drilling) to predict gold-bearing areas in a scalable and accessible way, limiting evidence-based decision-making, restricting access to open methodologies for responsible exploration, and keeping entry costs high for sustainable mining projects. This document presents a comprehensive framework for predicting and visualizing potential gold deposits using multimodal remote sensing, geochemical, and geological datasets, inspired by the EarthScape dataset and the Colombian alluvial gold prospectivity model, aiming to create a reproducible AI pipeline that integrates public data from USGS, SGC, Sentinel-2, and ASTER to improve sustainable exploration decisions and transparency in mineral resource discovery.

1. Introduction

Traditional mineral exploration is expensive, slow, and environmentally invasive due to its dependence on physical drilling and field sampling, with existing private machine learning applications but scarce open and replicable frameworks for mineral prospectivity. This project bridges that gap by proposing a fully open-source pipeline that leverages multimodal geospatial data and deep learning to predict gold occurrence, supporting sustainable exploration in Colombia and other resource-rich regions.

2. Related Work

EarthScape Dataset (2025)

The **EarthScape** initiative (University of Kentucky) introduced a large-scale AI-ready geospatial dataset integrating optical imagery, DEMs, hydrological, and infrastructural data. It demonstrated that multimodal fusion enhances surface geological mapping without drilling.

Machine Learning for Alluvial Gold (2025) — Colombian Ground Truth

The **Colombian alluvial gold study** developed by ITM, UdeA, and UNAL serves as the ground-truth model for borehole-level training and validation. Focused on Caucasia, Antioquia (Cauca River basin) with 147 boreholes and 8,642 samples, this research compares three machine learning approaches:

- **NLP-based 3D Modeling:** GloVe embeddings for lithological text analysis
- **CNN + Positional Encoder:** ResNet18 with transfer learning on DEM patches
- **Linear Interpolation Baseline:** Traditional geostatistical approach

The hybrid CNN-DEM model achieved superior performance by integrating topographic data with borehole information, establishing the methodological foundation for subsurface gold occurrence prediction. The complete implementation is available at github.com/francobr991/Mapeo_prospectivo.

These works validate the use of multimodal deep learning for mineral prospectivity and inspire the architecture adopted here.

3. Methodology

Data Sources

- **USGS MRDS:** Global mineral occurrence database (Gold-specific subset for Colombia).
- **SGC Geoquímica:** Geochemical and geological mapping data from Colombia.
- **Sentinel-2 / ASTER:** Multispectral imagery and DEM for EarthScape multimodal analysis.
- **EarthScape Integration:** DEM, hydrology, infrastructure, and terrain derivatives.
- **Colombian Geological Data:** SGC lithology, formation ages, and structural information.
- **Geophysical Data:** Magnetic and gravity anomalies for subsurface analysis.

Feature Engineering

Features are derived from EarthScape's multimodal fusion approach:

- **Spectral analysis:** NDVI, Iron, Clay indices from Sentinel-2 multispectral data
- **Terrain derivatives:** Slope, aspect, curvature, TWI, flow accumulation from DEM
- **Geochemical concentrations:** Au, As, Cu, Fe, Sb from SGC geochemical surveys
- **Hydrological features:** River proximity, watershed classification, drainage patterns
- **Geological context:** Lithology, formation ages, structural complexity
- **Geophysical signatures:** Magnetic and gravity anomalies for subsurface structures
- **Infrastructure accessibility:** Distance to roads, settlements, and mining infrastructure

Modeling Architecture

The AI system implements a three-phase deep learning architecture:

1. **Phase 1 - EarthScape Segmentation:** ResNeXt/UNet with multimodal inputs for geological feature segmentation
2. **Phase 2 - Colombia Hybrid Model:** DEM + borehole depth CNN with positional encoding for subsurface prediction
3. **Phase 3 - Cross-Region Transfer Learning:** Fine-tuning across geological regions for improved generalization

Phase 1: EarthScape-Style Segmentation

Following the University of Kentucky methodology, Phase 1 implements:

- **ResNeXt-50 Backbone:** High-performance CNN architecture for feature extraction
- **UNet Segmentation:** Encoder-decoder architecture for pixel-wise geological classification
- **Multimodal Inputs:** RGB + NIR imagery, DEM, hydrology, and terrain derivatives
- **Attention Mechanisms:** Spatial and channel attention for improved feature fusion

Phase 2: Colombia Hybrid Model

Building upon the Colombian research, Phase 2 integrates:

- **DEM-Borehole Fusion:** Hybrid CNN combining surface topography with subsurface borehole data
- **Positional Encoding:** Spatial context encoding for 3D geological modeling
- **Depth-Aware CNN:** Convolutional layers adapted for borehole depth information
- **Ground Truth Validation:** 147 boreholes from Caucasia providing empirical validation

Phase 3: Cross-Region Fine-Tuning

Phase 3 enables transfer learning across geological regions:

- **Regional Adaptation:** Fine-tuning for different geological provinces
- **Domain Generalization:** Improving model performance across diverse terrains
- **Incremental Learning:** Continuous model improvement with new borehole data
- **Uncertainty Quantification:** Probabilistic outputs for exploration decision-making

The model outputs a continuous probability field $P(\text{gold presence}) \in [0, 1]$ that can be visualized as a georeferenced heatmap.

4. EarthScape Integration

Building upon the University of Kentucky's EarthScape methodology, this project implements the first comprehensive multimodal geospatial data fusion pipeline for geological surface mapping in Colombia. The integration establishes a reproducible AI-ready framework that eliminates the need for expensive drilling operations.

Technical Achievements

- **EarthScape Integration:** First implementation of University of Kentucky methodology in Colombia
- **Multimodal Fusion:** 27 new geological variables added to prediction model
- **Production Scale:** Handles 1,000+ samples with comprehensive EarthScape preprocessing
- **AI-Optimized:** Dataset structure optimized for deep learning geological analysis
- **Open Science:** Reproducible pipeline for global geological mapping applications

EarthScape Data Sources

The Colombian implementation extends EarthScape's multimodal approach with local data:

- **DEM + Terrain Derivatives:** ASTER DEM with slope, curvature, TWI, and flow accumulation
- **RGB + NIR Imagery:** Sentinel-2 multispectral bands (B02, B03, B04, B08, B11, B12)
- **Hydrology + Infrastructure:** IDEAM river networks, watersheds, and infrastructure proximity
- **Geological Context:** SGC lithology, formation ages, and structural data
- **Geophysical Data:** Magnetic and gravity anomalies for subsurface structure analysis

Scientific Impact

mapping technology available, providing unprecedented accuracy for mineral exploration without the need for expensive drilling operations. This represents a paradigm shift from traditional invasive exploration methods to AI-powered surface-based analysis, significantly reducing environmental impact while improving exploration efficiency.

5. Advanced Visualization Framework

3D Geospatial Dashboard

The system provides an interactive 3D visualization platform using modern web technologies:

- **CesiumJS Integration:** High-performance 3D globe rendering with geological overlays
- **Leaflet 3D:** Lightweight 3D mapping with borehole cross-section visualization
- **Kepler.gl:** GPU-accelerated geospatial visualization for large datasets
- **WebGL Rendering:** Hardware-accelerated 3D geological modeling and probability mapping

Visualization Outputs

The dashboard generates multiple visualization layers:

1. **Gold Probability Heatmaps:** Interactive probability surfaces with uncertainty bounds
2. **Borehole Cross-Sections:** 3D subsurface visualization of lithology and gold concentrations

3. **Geological Feature Maps:** Segmented geological units from Phase 1 EarthScape models
4. **Temporal Analysis:** Time-series visualization of model predictions and validations
5. **Exploration Decision Support:** Interactive tools for targeting high-probability zones

Technical Implementation

- **FastAPI Backend:** RESTful API for serving model predictions and borehole data
- **WebSocket Updates:** Real-time model updates and validation results
- **Progressive Web App:** Offline-capable interface for field exploration teams

6. Open Science and Reproducibility

The project is distributed under the **MIT License** (software) and **CC-BY 4.0** (data). The repository includes:

- Full data ingestion scripts.
- Model cards and data cards.
- Tutorials for reproducibility and transfer learning.
- Code snippets for model deployment and API integration.

7. Expected Impact

- Democratization of geoscience AI research.
- Reduction of exploratory costs and environmental footprint.
- Support for sustainable mining and transparency.

8. Acknowledgements

This initiative builds upon research and open data provided by:

- United States Geological Survey (USGS)
- Servicio Geológico Colombiano (SGC)
- Copernicus Open Access Hub (ESA Sentinel)
- ITM, Universidad de Antioquia, Universidad Nacional de Colombia
- University of Kentucky EarthScape Initiative

9. References

- EarthScape: "An Open, AI-Ready Geospatial Dataset for Global Environmental Modeling," University of Kentucky, 2025.
- Franco B. et al. "Machine Learning-Based Prospective Modeling for Alluvial Gold Mining: A Study Area in Colombia," *Pure and Applied Geophysics*, 2025.