GeoAuPredict (GAP): Deep Learning for Geospatial Prediction of Gold

Final Project — Deep Learning Course

Edward Calderón & Team

October 13, 2025

Motivation and Problem Statement

- Why geospatial prediction of gold is relevant (exploration, cost, risk).
- Traditional approaches: surveys, interpolation, geostatistics.
- Challenges: heterogeneous data, spatial autocorrelation, missingness.
- Our goal: learn from multi-modal geospatial data to predict Au occurrence.

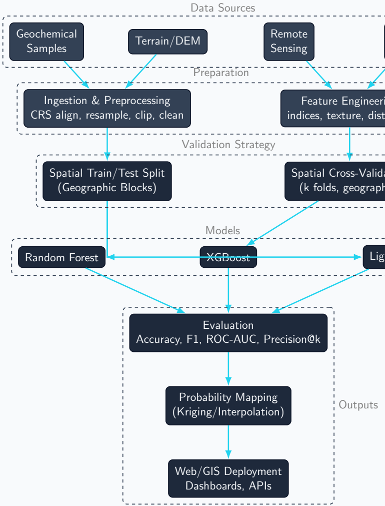
Data Ingestion and Preprocessing

- Sources: DEM, remote sensing indices, geology layers, geochemical assays, distances to structures.
- CRS harmonization, resampling, study area clipping, normalization.
- Spatially aware splits; class balance checks.

Exploratory Data Analysis

- Correlations and distributions; outliers and missingness.
- Spatial plots of assays and features; clustering and hotspots.
- Early insights to guide feature engineering.

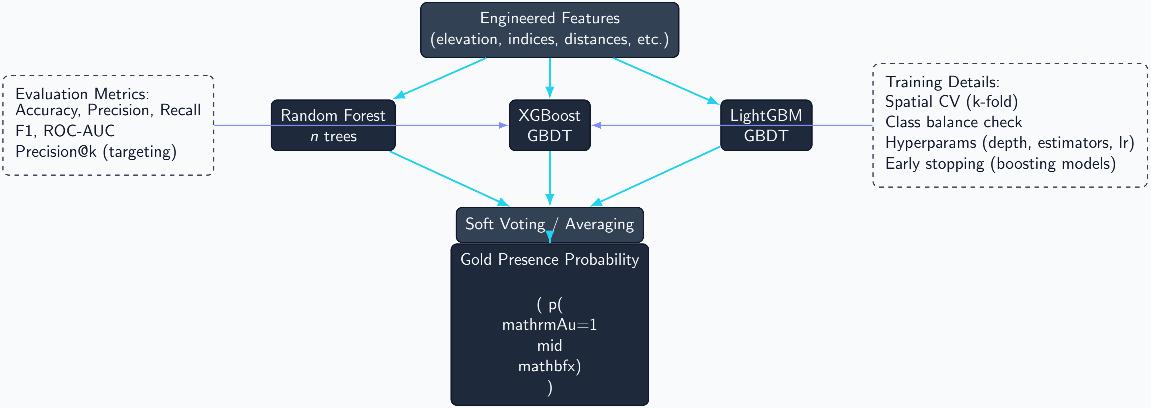
Pipeline Flow



Modeling Approach

- Tree ensembles: Random Forest, XGBoost, LightGBM (tabular geospatial features).
- Spatial cross-validation (geographic blocks) to assess generalization.
- Probability mapping via interpolation; uncertainty estimation.

Model Architecture (Ensemble)



Training & Validation

- Metrics: Accuracy, Precision, Recall, F1, ROC-AUC, Precision@k.
- Early stopping for boosting models; hyperparameter sanity checks.
- Spatial CV (geographic blocks) for honest generalization.
- Key results from notebooks:
  - **LightGBM (best base)**: AUC = **0.9243**
  - **Voting Ensemble (production)**: AUC = **0.9208** (spatially separated test)
  - **Spatial Blocks AUC**: **0.86** (geographic CV)

Results Summary

- **Best base model**: LightGBM — AUC **0.9243**
- **Production model**: Voting Ensemble — AUC **0.9208** (spatially separated test)
- **Spatial Blocks AUC (CV)**: **0.86 —** honest, geography-aware estimate
- Precision@k supports efficient targeting (top areas). Demo runs show high P@10; full report omitted here

Comparison with Classical Methods

- Baselines: kriging, logistic regression, WoE, SVM (literature typical AUC ≈ 0.82).
- **Improvement**: $+\sim$**22%** AUC over best classical ($0.82 \rightarrow 0.9208/0.9243$).
- Strengths: DL/ensembles capture non-linear, multi-modal interactions.
- Limitations: classical methods often assume stationarity/linearity.

Ablation & Sensitivity

- Feature importance / permutation importance; remove/keep studies.
- Hyperparameters: depth, estimators, learning rate.
- Robustness: spatial CV folds, noise perturbations.

Deployment & Use Cases

- Deployment: web dashboards, GIS export, APIs.
- Use cases: guide field sampling; prioritize zones; decision support.
- Limitations: domain shift, data availability, generalization.

Q&A Strategy (Judge Questions)

Why DL vs geostatistics?	Non-linear interactions; heterogeneous modalities; scale. Classical often assume stationarity / linearity.
Avoiding overfitting?	Spatial CV; regularization; early stopping; non-contiguous holdouts; augmentation.
Interpretability?	Feature importance, SHAP, LRP, surrogate models; spatial attribution maps.
Transferability?	Domain shift awareness; fine-tuning; multi-region training; domain adaptation.
Uncertainty?	If absent: propose MC-dropout, ensembles, quantile reg.; report as future work.
Compare to SOTA?	Cite TorchGeo/remote sensing DL; compare metrics/approach; highlight novelty.
Limitations?	Overfitting risk; sparse labels; sensor noise; domain shift; interpretability.
Operational integration?	Data ingestion pipeline; real-time inference; GIS integration; feedback loop.

Conclusions

- Novelty: multi-modal integration, spatial CV, probability mapping.
- Impact: reduce exploration costs; improve targeting; scalable pipeline.
- Future: uncertainty quantification; domain adaptation; explainability.

Questions

Thank you!

Questions welcome.

▸ Demo

github.com/edwardcalderon/GeoAuPredict