The Price of Power: A Data-Driven Investigation into Environmental Risk in Jharkhand's Uranium Belt

Naitik Agrawal — Track: Tech-AI/ML

June 25, 2025

Contents

1 Context and Motivation **2**

2 Methodology **3**
2.1 Task 1: Environmental Change via Satellite Imagery 3
2.2 Task 2: Environmental Risk Prediction Model 3

3 Findings and Results **4**
3.1 Landscape Transitions (Task 1) 4
3.2 Risk Classification Outcomes (Task 2) 7

4 Discussion **8**

5 Recommendations and Action Blueprint **9**

6 Resources and Code **10**

7 Appendix **11**

1 Context and Motivation

Mining activities across resource-rich regions often leave behind long-term ecological footprints and health-related challenges that remain underexplored and underreported. These impacts, while spatially visible, are rarely monitored systematically or communicated in a transparent manner.

This project aims to bridge that gap by leveraging satellite imagery, machine learning, and open-access geospatial data to make environmental risk visible, quantifiable, and actionable.

The core objective is to demonstrate how modern data tools can be used to:

- Map and quantify environmental degradation over time using remote sensing
- Develop a risk prediction model using ecological and health indicators
- Provide an accessible, interpretable dashboard for stakeholders — from local decision-makers to researchers and civil society

This work represents a step toward responsible resource governance — where extraction is continuously monitored, environmental change is transparent, and technology supports equitable and informed decision-making.

Objectives:

- Detect long-term environmental change using satellite data (NDVI, LULC, VIIRS Nightlights)
- Train a decision tree model to classify regions as "High Risk" or "Low Risk"
- Build an interpretable dashboard for policy support and risk simulation

Approach:

- Processed Landsat and VIIRS data to map deforestation, land use change, and urban expansion
- Generated a land-use transition map and visualized hotspot regions
- Developed a risk prediction model using proxy indicators: forest loss, illnesses, distance to mines

Key Findings:

- Significant **Forest** → **Urban** transitions reflect mining-led encroachment
- Nightlight analysis confirms rapid urban growth in settlement belts
- Decision Tree model shows forest loss and health issues are top risk predictors

Deliverables:

- Remote sensing-based maps of NDVI, LULC, Transitions, and Nightlights
- Risk prediction Streamlit dashboard with interactive sliders
- Actionable recommendations aligning mining oversight with ESG goals

2 Methodology

This project was divided into two major components:

2.1 Task 1: Environmental Change via Satellite Imagery

We used Google Earth Engine to process 10 years of Earth observation data (2013–2023), focusing on a uranium mining region in Jharkhand, India. The aim was to quantify ecological degradation and urban expansion. The following geospatial products were generated:

- **NDVI Change Map (2013–2023):** Normalized Difference Vegetation Index (NDVI) was computed using Landsat-8 images to detect changes in vegetation. Red regions in the map indicate areas of deforestation, while green regions indicate vegetation gain.
- **LULC Classification (2013 and 2023):** Random Forest classification was applied to identify major land cover types: *Forest (0)*, *Urban (1)*, *Barren (2)*, and *Water (3)*. ESA WorldCover data was used to label training points. This allowed detection of conversion from forest to urban zones and other transition types.
- **LULC Transition Map :** Pixel-wise class transitions (e.g., Forest → Urban) were visualized. This helped us identify top drivers of landscape transformation.
- **Nightlight Growth Analysis:** VIIRS data from 2014 and 2023 was used to detect nighttime brightness growth — a proxy for human settlement expansion. Yellow to blue regions indicate increasing urban activity.

2.2 Task 2: Environmental Risk Prediction Model

To simulate how environmental degradation affects public health, we developed a machine learning-based risk prediction system using a proxy dataset.

The tabular data includes the following features:

- **Forest Cover Loss (ha)** — Proxy for ecological disruption
- **Reported Illnesses** — Proxy for community health impact
- **Distance to Mines (km)** — Assumes inverse risk correlation
- **Zone and Year** — Region metadata

A Decision Tree classifier (scikit-learn) was trained on this data to classify regions into:

- **High Risk (1):** Zone likely affected by health and ecological decline
- **Low Risk (0):** Zone appears environmentally and socially stable

Interpretability was ensured through use of:

- Feature importance ranking
- Confusion matrix to analyze prediction reliability
- Full decision tree visualization

The proxy-based model enables scenario simulation and can be refined with real field data.

3 Findings and Results

3.1 Landscape Transitions (Task 1)

Over the decade from 2013 to 2023, satellite imagery reveals significant environmental changes:

- **NDVI Change Map:** Large regions show a shift from high to low NDVI, indicating deforestation. Small green patches suggest areas of vegetation regrowth.

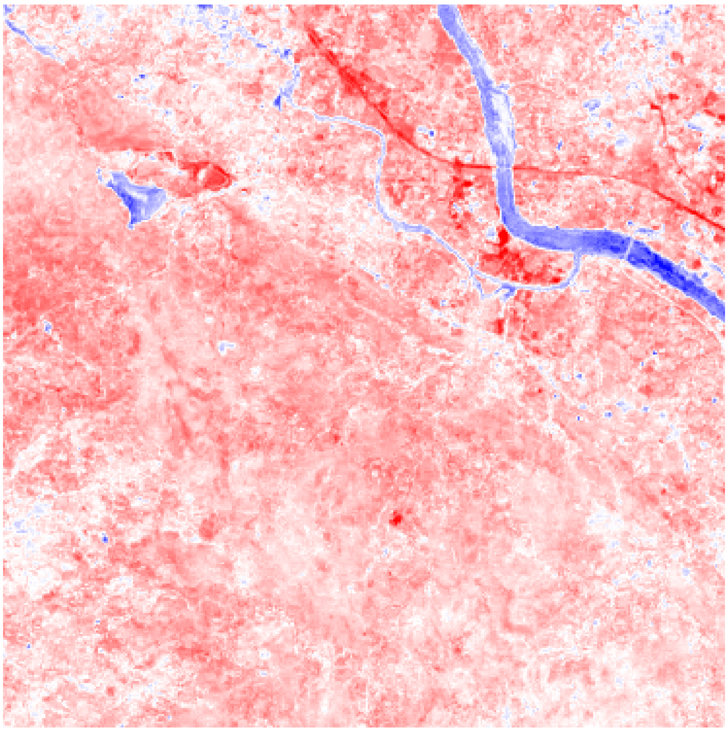
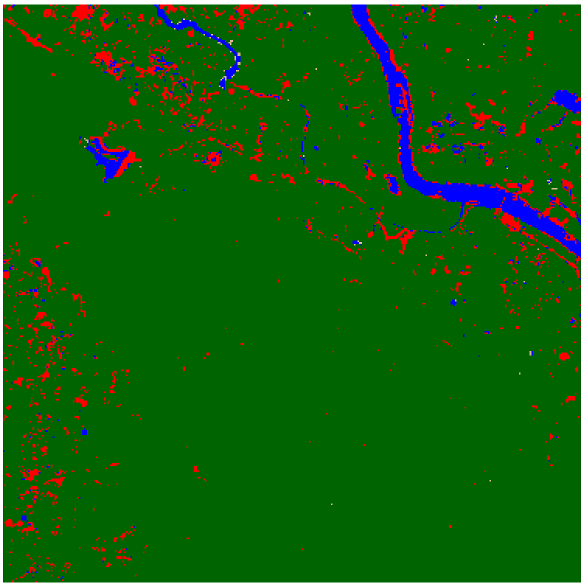
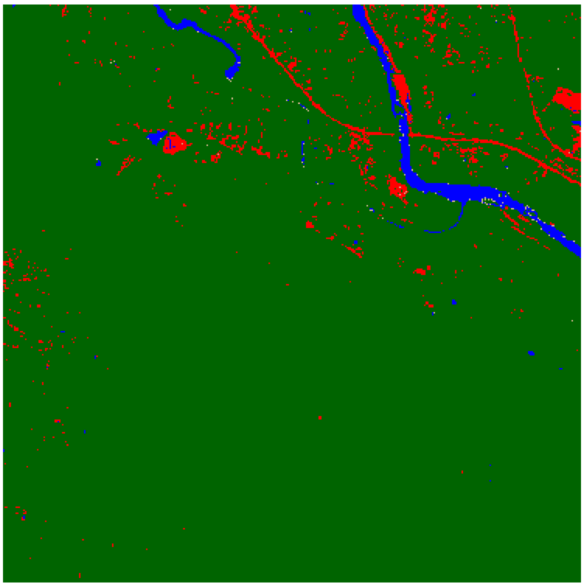


Figure 1: NDVI Change (2013–2023): regrowth, deforestation

- **LULC Classifications:** Comparison of 2013 and 2023 land cover shows increasing urban spread and a reduction in forest cover. Barren and water areas remained largely stable.



(a) LULC 2013



(b) LULC 2023

Figure 2: Land Use Land Cover Classification over 10 years Forest Urban Barren Water

- **LULC Transition Map:** Pixel-wise transitions highlight critical shifts:

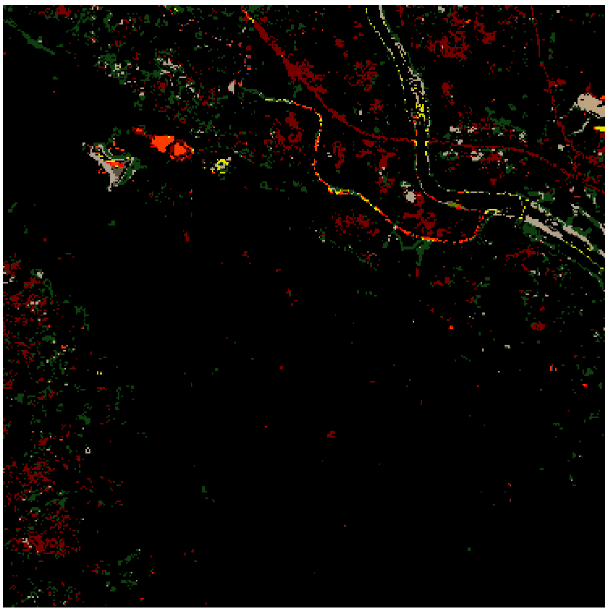


Figure 3: LULC Transition Map (2013–2023)

Forest → Urban
Forest → Barren
Forest → Water

Barren → Urban
Water → Urban
Urban → Barren
Urban → Forest
Water → Forest
No Change (00, 11, 22, 33)

  - Forest → Urban : Most prominent transition, implying urban encroachment into green zones.
  - Forest → Barren : Indicates ecosystem degradation.
  - Urban → Forest : Sparse but possible evidence of restoration.

- **Nightlight Growth (VIIRS):** Urban areas show increased brightness from 2014 to 2023, correlating strongly with urban expansion observed in LULC maps.

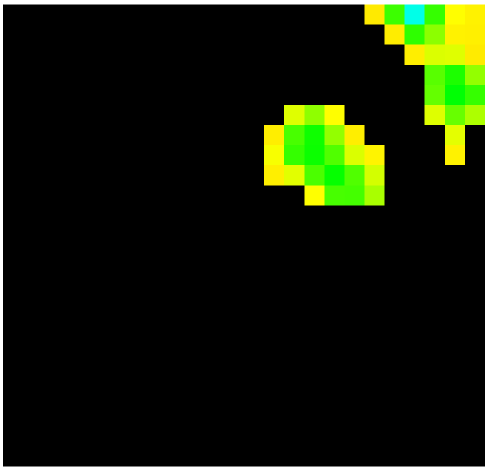


Figure 4: VIIRS Nighttime Light Growth (2014–2023): A proxy for urban expansion based on nightlight intensity.

- Very Low Brightness
- Low Brightness
- Moderate Brightness
- Growth Region
- High Urban Growth
- Very Bright Urban
- City Centers / Peak Urban
- No Change / Masked Area

Interpretation: These visual insights reflect unregulated land use, loss of green cover, and expanding human footprint — aligning with documented challenges in UCIL-adjacent zones.

3.2 Risk Classification Outcomes (Task 2)

The risk prediction model — based on proxy environmental data — revealed patterns tied to ecological stress and health vulnerability.

- **Feature Importance:** Forest Cover Loss and Reported Illnesses are the dominant predictors, confirming their value as indicators of risk.

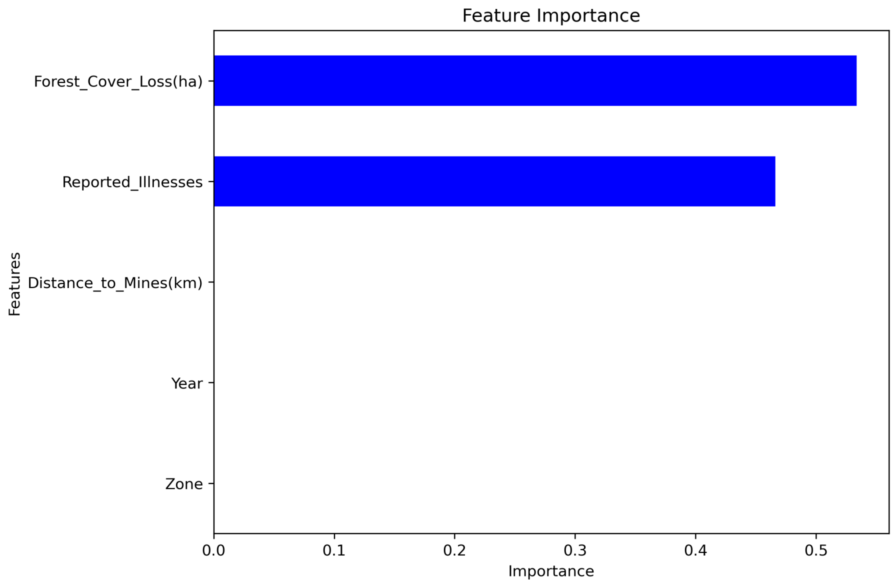


Figure 5: Feature Importance based on Decision Tree Model. Forest Cover Loss and Reported Illnesses dominate the predictive logic.

- **Confusion Matrix:** Model performs well in distinguishing high-risk and low-risk zones.

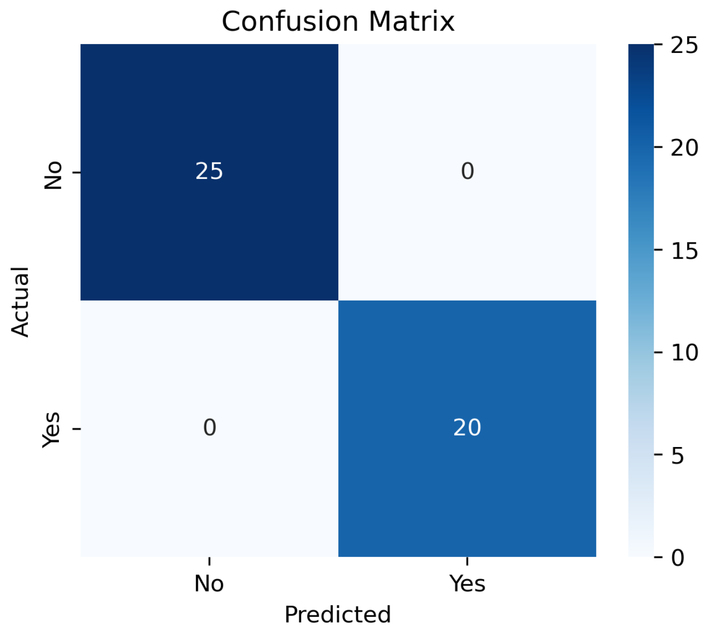


Figure 6: Confusion Matrix showing the model's ability to distinguish between high and low risk zones.

- **Decision Tree:** Shows clear thresholds for risk (e.g., > 500 ha forest loss and > 1000 illnesses ⇒ High Risk).

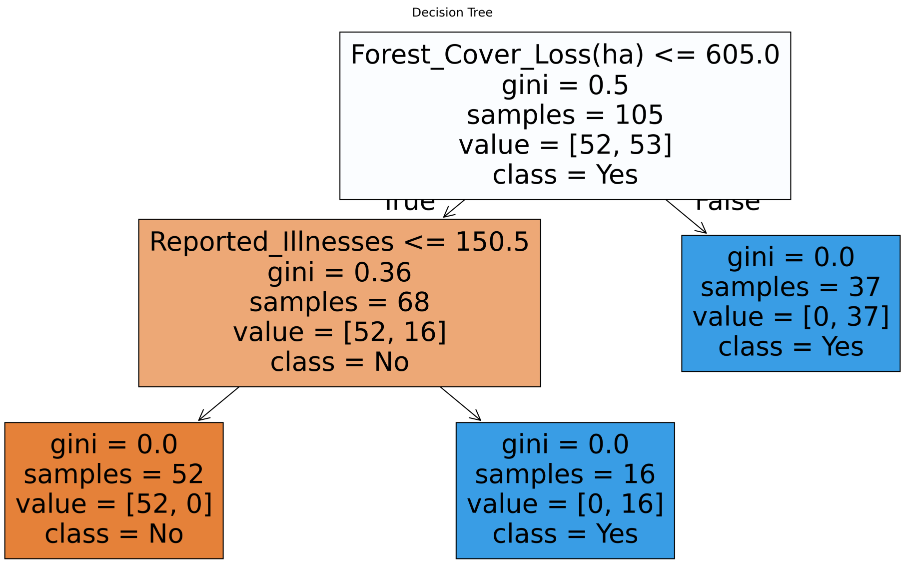


Figure 7: Full Decision Tree structure used to classify zones. Splits are based on forest loss and reported illnesses.

Scenario Simulation: The model allows for user-driven policy simulation (e.g., increasing forest restoration decreases predicted risk). This can inform proactive governance.

4 Discussion

The combined insights from satellite imagery and proxy-based modeling reveal a multi-faceted picture of environmental risk in the UCIL mining belt.

Environmental Degradation is Measurable and Spatially Explicit. Using NDVI and LULC change detection, we can trace the systematic disappearance of vegetation over a 10-year period. The most dominant land transition — **Forest → Urban** — reflects encroachment tied to mining activity, infrastructure development, and unregulated land use. This spatial evidence counters narratives of "minimal ecological impact" and provides empirical grounding for community concerns.

Urban Expansion is Confirmed Across Datasets. VIIRS Nightlight analysis serves as an independent proxy validating LULC findings. Increased nighttime brightness aligns with new settlements, road networks, and industrial activity — particularly in the **southern and southeastern segments of the ROI** defined by coordinates `(86.31°E, 22.58°N) to (86.41°E, 22.68°N)`. This rectangular region in eastern Jharkhand was selected due to its proximity to known UCIL uranium mining operations. The cross-verification of LULC change and VIIRS-based brightness growth **strengthens the case for accelerated urban encroachment** over a 10-year period.

Risk is Not Random — It Follows Ecological Fault Lines. Our decision tree model confirms that forest loss and illness burden are the two strongest predictors of whether a zone is "high risk." Distance to mines played a lesser role — suggesting that indirect ecological pathways (like air/water degradation or psychosocial stress) may be more impactful than proximity alone.

This Model is a First Draft, Not a Final Diagnosis. While useful for simulation and scenario testing, the risk classifier is trained on proxy data. Future versions should incorporate:

- Real hospital records and mining health data
- Ground-truthed environmental samples (e.g., air, water quality)
- Temporal progression and seasonal patterns

Still, the current version offers a low-cost, interpretable decision aid that can flag zones for further investigation.

Technology Can Amplify, Not Replace, Justice. This project is not a replacement for community-led movements or field activism. Rather, it complements those struggles with visibility, traceability, and pressure. It equips regulators, citizens, and watchdogs with tools to demand accountability from extractive systems.

5 Recommendations and Action Blueprint

Based on both remote sensing insights and risk prediction outcomes, the following recommendations are proposed for stakeholders including UCIL, state policymakers, environmental regulators, and civil society actors:

1. Integrate Remote Sensing into Regulatory Oversight

- Use NDVI and LULC transitions to monitor environmental degradation annually.
- Make it mandatory for mining leases to be accompanied by geospatial degradation audits.
- Flag rapid **Forest → Urban** transitions as zones of interest for immediate inspection.

2. Health–Environment Data Fusion

- Link remote-sensed forest loss and urban growth with public health surveillance (e.g., local illnesses, groundwater contamination).
- Develop an early warning system where increased forest loss + reported illness spike trigger an alert.

3. Public Dashboard and Open Monitoring

- Maintain public-facing dashboards like this project's Streamlit app.
- Let NGOs, researchers, and community members access live updates and satellite maps.
- Increase transparency by integrating UCIL's environmental clearances with geospatial monitoring.

4. Mandate Environmental Restoration

- Zones with high deforestation or **Forest** → **Barren** transitions must be brought under afforestation or restoration schemes.

5. Data-Driven Policy Interventions

- Use risk prediction tools to simulate intervention outcomes (e.g., increasing forest cover reduces risk).
- Allocate healthcare and environment budgets to zones flagged as high-risk by the model.

This blueprint aims to move from retrospective accountability to **real-time monitoring and proactive mitigation**. The integration of satellite data, AI, and community-facing dashboards can shift the paradigm from extractive opacity to environmental justice.

6 Resources and Code

This project uses a mix of remote sensing, machine learning, and interactive visualization. All code and outputs are publicly available.

Web Dashboard

- **Streamlit App:**
  `https://atw8vyxjpqxwh5fq4yhrqx.streamlit.app/`
  Live policy simulation + map visualizations

Codebase

- **GitHub Repository:**
  `https://github.com/NaiTi-K/CSI.git`
  Contains model training code and full dashboard code

Geospatial Datasets

- **NDVI Change Map (2013–2023)**
- **LULC Classification Maps (2013, 2023)**
- **LULC Transition Map**
- **VIIRS Nightlight Growth (2014–2023)**

ML-Based Risk Model

- Trained using a decision tree classifier on proxy tabular data
- Predicts High vs Low Risk zones based on:
  - Forest Cover Loss
  - Reported Illnesses
  - Distance to Mines
- Interpretable via decision tree plots, feature importance, and confusion matrix

7 Appendix

A. Supporting Images

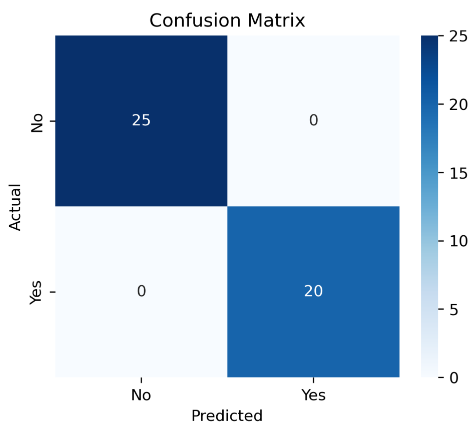


Figure 8: Confusion Matrix of the Risk Prediction Model

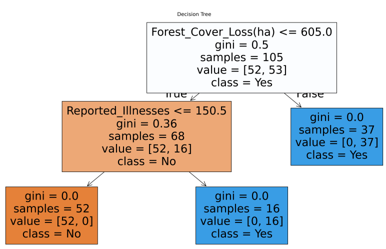


Figure 9: Decision Tree Classifier for Environmental Risk

C. LULC Label Codes

For interpretation of raster legends:

- 0 — Forest (Green)
- 1 — Urban (Red)
- 2 — Barren (Brown/Tan)
- 3 — Water (Blue)

D. Transition Class Codes

These codes were used in the pixel-wise transition maps.

- 01 — Forest → Urban (Red)
- 02 — Forest → Barren (Tan)
- 03 — Forest → Water (Blue)
- 10 — Urban → Forest (Bright Green)
- 12 — Urban → Barren (Gray)
- 21 — Barren → Urban (Dark Red)
- 30 — Water → Forest (Dark Green)
- 31 — Water → Urban (Brick Red)