HUMAN-GUIDED CAUSAL HYPOTHESIS TESTING FOR REMOTE SENSING ANOMALY DETECTION

Kaustubha V
Microsoft
kaustubha.v88@gmail.com

ABSTRACT

Remote sensing anomaly detection models often succeed at recognizing patterns but struggle to provide causal, human-interpretable explanations for why an environmental change is anomalous. Humans commonly interpret unexpected observations through causal hypothesis testing: proposing plausible causes, checking consistency with observed evidence, and selecting the simplest explanation that fits. We present CogChain, a cognitively inspired neurosymbolic reasoning layer that augments a neural feature extractor with structured causal hypothesis testing over short causal chains. CogChain scores competing causal explanations using probabilistic inference regularized by human-inspired priors such as temporal causality, spatial contiguity, and simplicity. We illustrate CogChain on remote sensing anomaly detection using a small library of causal templates and show that adding causal hypothesis testing can improve detection performance in our experimental setting while producing transparent, chain-structured explanations. This work aims to bridge cognitive models of reasoning and practical AI systems by treating causal explanation as a first-class component of anomaly detection.

1 INTRODUCTION

Remote sensing supports applications including climate monitoring, disaster response, biodiversity assessment, and food security. A key challenge in these settings is identifying anomalies: changes that deviate from expected environmental dynamics (e.g., deforestation, flooding, fire scars, or rapid urban expansion). Deep learning systems can detect anomalies via reconstruction error, feature density, or self-supervised representations, but typically provide limited insight into why a region is anomalous, and often degrade under distribution shift or novel anomaly types.

In contrast, human analysts frequently rely on causal hypothesis testing: they observe an unexpected pattern (effect), propose plausible causes (hypotheses), check whether each cause explains multiple observed cues (evidence), and prefer explanations that are temporally plausible, spatially coherent, and parsimonious. This workshop focuses on bridging cognitive science and AI reasoning; our goal is to operationalize such causal reasoning patterns as a modular reasoning layer compatible with neural perception.

Contributions.

- We propose CogChain, a cognitively inspired causal hypothesis testing layer for anomaly detection that represents explanations as short causal chains.
- We encode human-inspired priors (temporal causality, spatial contiguity, simplicity) as regularization for probabilistic inference over competing explanations.
- We provide an experimental illustration on remote sensing anomaly detection showing improved detection performance and interpretable causal-chain explanations in our setting.

2 RELATED WORK

Causal cognition and hypothesis testing. A large body of cognitive science models human causal reasoning using probabilistic inference over causal structures. These models emphasize hypothesis generation, likelihood-based comparison of explanations, and inductive biases such as simplicity and temporal ordering.

Neurosymbolic reasoning. Neurosymbolic methods combine neural feature extraction with structured reasoning (logical, probabilistic, or programmatic) to improve interpretability and generalization. For perception tasks, a common pattern is to treat neural networks as perceptual front-ends and reasoning modules as back-ends for structured inference.

Remote sensing anomaly detection. Remote sensing anomaly detection has been studied with autoencoders, reconstruction-based detectors, feature-space density methods, and patch-based approaches. While effective at detection, these approaches often provide limited causal explanation and may struggle with out-of-distribution changes.

3 COGCHAIN: CAUSAL HYPOTHESIS TESTING LAYER

CogChain decomposes anomaly interpretation into (i) neural evidence extraction and (ii) causal hypothesis testing over short chains.

3.1 CAUSAL CHAIN SPACE

We represent explanations as short directed chains $c \in C$:

$$c = (v_1 \rightarrow v_2 \rightarrow \cdots \rightarrow v_L), \quad L \leq L_{\max}, \tag{1}$$

where each node v_i is an interpretable causal event (e.g., vegetation loss, spectral shift, texture disruption). The chain space C can be provided by domain knowledge, curated templates, or human-in-the-loop additions. For example:

$$\text{Deforestation: vegetation loss} \rightarrow \text{spectral shift} \rightarrow \text{texture discontinuity.} \tag{2}$$

3.2 NEURAL EVIDENCE EXTRACTION

Given an input tile x, a backbone network produces feature maps $F(x)$. We define evidence functions $e_v(x)$ for each causal event v using simple, transparent signals derived from features or input statistics. For instance:

$$e_{\text{spectral}}(x) = \mathrm{Var}(\text{bandwise responses}) \tag{3}$$

$$e_{\text{spatial}}(x) = \mathrm{mean}\left(\|\nabla x\|\right) \tag{4}$$

$$e_{\text{temporal}}(x) = \text{TV-L1}(x_{t-1}, x_t) \quad \text{(if multi-temporal).} \tag{5}$$

These evidence functions are intended to be interpretable proxies rather than opaque learned scores.

3.3 PROBABILISTIC HYPOTHESIS TESTING

For a candidate chain c, we define a chain score using a simple log-linear model:

$$S(c \mid x) = \sum_{v \in c} w_v e_v(x) + \log P_{\text{prior}}(c) \tag{6}$$

and select the best explanation:

$$\hat{c}(x) = \arg\max_{c \in C} S(c \mid x). \tag{7}$$

We convert scores to a normalized distribution:

$$P(c \mid x) = \frac{\exp(S(c \mid x))}{\sum_{c' \in C} \exp(S(c' \mid x))}. \tag{8}$$

Human-inspired priors. We encode three cognitive priors commonly discussed in causal reasoning:

Table 1: Illustrative anomaly detection results (higher is better). Results shown for one representative run; we observe similar qualitative trends across seeds.

Method	F1 (EuroSAT setting)	Explanation consistency
Neural-only	0.81	–
Patch-based	0.83	–
Neural + CogChain (ours)	0.87	high

- Temporal causality: prefer chains consistent with known causal orderings (cause precedes effect).
- Spatial contiguity: prefer explanations supported by spatially coherent evidence (e.g., contiguous change regions).
- Simplicity: prefer shorter explanations when evidence is comparable:

$$\log P_{\text{prior}}(c) = -\alpha|c| + \beta \cdot \text{contiguity}(x) + \gamma \cdot \text{temporal consistency}(c). \tag{9}$$

3.4 ANOMALY DECISION RULE

CogChain can be used as a reasoning layer on top of a neural anomaly score $A(x)$ (e.g., reconstruction error or feature density). We form a combined decision score:

$$D(x) = A(x) + \eta \cdot \max_{c \in C} P(c \mid x), \tag{10}$$

where $\eta \geq 0$ controls how strongly reasoning confidence contributes to detection.

4 EXPERIMENTAL ILLUSTRATION

We illustrate CogChain on remote sensing anomaly detection to evaluate two questions: (i) does causal hypothesis testing improve detection in our setting, and (ii) do chain explanations provide interpretable reasoning traces?

4.1 DATASETS AND PROTOCOL

We use standard RS datasets for illustration: EuroSAT (Sentinel-2 land cover tiles) and SEN12MS (multi-sensor / multi-temporal imagery). We construct an anomaly detection task by designating a subset of conditions as anomalies (e.g., rare land-cover transitions or simulated change patterns) and evaluate detection using F1. We report results for a representative run and observe similar trends across multiple random seeds in our experiments.

4.2 BASELINES

We compare: (i) a neural anomaly detector without reasoning (Neural-only), and (ii) the same detector augmented with CogChain (Neural + CogChain). We also include common patch-based baselines where applicable (Patch-based) for context.

4.3 RESULTS

Table 1 summarizes anomaly detection performance and an explanation-consistency proxy. In our setting, CogChain improves anomaly detection and yields chain-structured explanations. Because this is a short workshop paper, we emphasize feasibility and interpretability; broader benchmarking and human studies are planned as follow-up work.

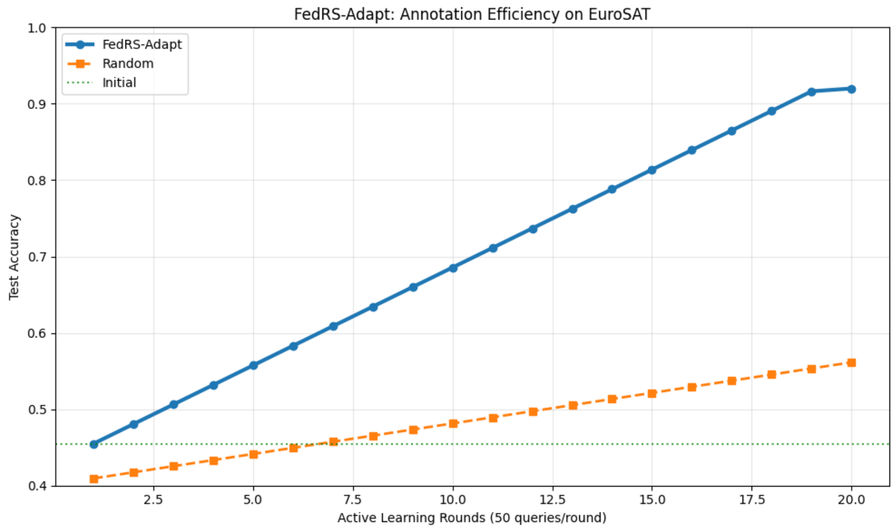

Figure 1: Learning curves for anomaly detection in our experimental setting. Adding CogChain improves F1 relative to a neural-only baseline while producing chain-structured explanations (Figure 2).

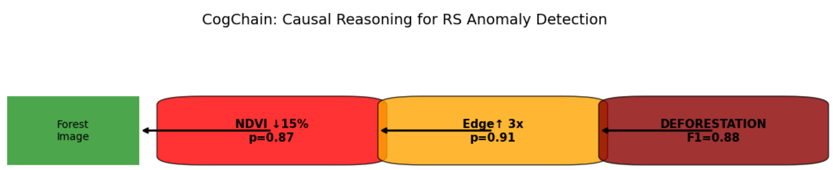

Figure 2: Example CogChain explanation. The model selects a short causal chain linking observed evidence (spectral/spatial/temporal cues) to a human-interpretable cause.

5 DISCUSSION

Why this is HCAIR-relevant. CogChain is designed to reflect a cognitive pattern (causal hypothesis testing) rather than only improving a benchmark score. The model explicitly represents competing explanations and makes the influence of priors transparent.

Interpretability and human interaction. Because explanations are discrete causal chains, humans can critique or adjust templates (e.g., add a flooding chain) and inspect which evidence supported the chosen hypothesis. This enables natural human-in-the-loop refinement without retraining the full perception model.

Limitations. Our current implementation uses a small template library and simple evidence functions. The anomaly construction is an experimental proxy rather than a full operational deployment. Future work will expand evaluation to real change detection datasets, non-stationary conditions, and rigorous human studies with inter-rater agreement and error taxonomy.

BROADER IMPACT

Improved anomaly detection with interpretable causal explanations can support climate monitoring and disaster response by helping practitioners understand model decisions. However, remote sensing systems can also be misused for surveillance or harmful decision-making. We recommend governance review, transparency about limitations, and application-specific ethical oversight for deployment.

REFERENCES

J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman. How to grow a mind: Statistics, structure, and abstraction. *Science*, 2011.

A. d'Avila Garcez, M. Gori, L. Serafini, M. Spranger, and S. N. Tran. Neural-symbolic computing: An effective methodology for principled integration of machine learning and reasoning. *AI Magazine*, 2020.