Multi-Task Learning for Fire Ignition Maps

Parastoo Hashemi : s339438

Yousef Fayyaz: s341999

FONDAZIONE Links PASSION FOR INNOVATION

1859

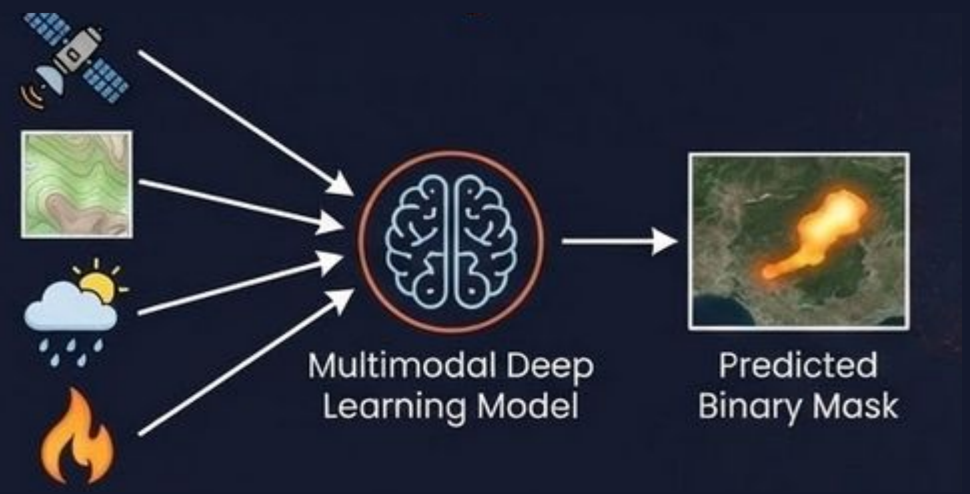
Project Objective

Core Objective

To develop a multimodal deep learning model capable of **Predicting the final burned area extent right at the moment of ignition,** using only **pre-fire multimodal data that** fused into a multimodal model

Input: Pre-fire environmental data (Satellite, Topography, Weather) + Ignition Point.

Output: Predicted binary mask of the final burned area.



The Paradigm Shift

Current State (Reactive): Remote sensing is primarily used for post-event damage assessment.

Target State (Proactive): To estimate the spread of fire immediately to allocate resources immediately during the "Golden Hour".

Value Proposition

Environmental Impact

Proactive predictions reduce environmental losses by enabling faster response before uncontrolled spread.

Fewer hectares burned → lower CO_2 emissions, preserved ecosystems, and reduced long-term regeneration costs.

Response Efficiency

Early burned-area forecasts help Civil Protection allocate firefighting teams, helicopters, and ground units efficiently.

Better situational awareness during the first critical hours improves containment and reduces operational risk.

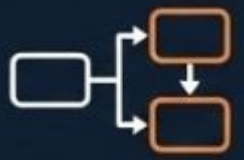
Innovation & Scalability

The multimodal AI system generalizes across regions where Sentinel-2, Landsat, DEM, ERA5, and infrastructure layers are available.

The framework can be replicated globally, providing a reusable template for data-driven wildfire risk management.

Research Questions

Can pre-fire multimodal data predict final burned area with sufficient accuracy to support early emergency decisions?



Does auxiliary-task learning improve spatial segmentation performance compared to single-task models?

What is the relative contribution of satellite, meteorological, and topographical modalities?

Pre-Fire Data: Satellite Imagery & Environmental Drivers

Satellite Imagery (Sentinel-2 & Landsat)

Sentinel-2 Imagery: 12-band multispectral, 10-20m res.
- why needed? captures immediate pre-fire condition (vegetation & fuel availability)

Landsat Imagery: Extra spectral info (SWIR/NIR)
- why needed? Sensitive to vegetation moisture and fuel dryness.
- Complementary spectral response for fuel characterization.

Environmental Drivers (DEM & ERA5)

Topograph (DEM): Terrain , slopes and valleys (30m res).
- why needed? Fire spreads uphill, elevation is crucial for behavior.

ERA5 weather: Temperature & wind at ignition time.
- why needed? Meteorological conditions are key drivers of fire expansion.
- wind vectors provide essential dynamic information for propagation direction.

Ground-Truth Fire Labels

BURNED AREA MASK & IGNITION DATA

Burned Area Mask: Shape of the area that actually burned,
Binary Segmentation
(1 = burned, 0 = unburned)

why needed? Serve as a target label for supervised learning segmentation

Ignition Data: The exact location where the fire started. (Binary marker), derived from Civil Protection point data.

why needed? The model learns how the fire is likely to spread from that specific starting position.

Land cover and Roads

Land cover map: Categorical mask of vegetation types (Forest, Shrub, Grassland, etc.).

why needed? Critical for estimating fuel type, Assists the model in differentiating propagation potential based on available fuel.

Roads: Binary road mask (10m resolution).

why needed? Captures both human-caused ignition likelihood and fire break potential.

Data Preprocessing: Cloud Exclusion

Why it matters: Clouds obscure ground features (occlusion) and alter pixel reflectance values (radiometric noise), creating "blind spots" in the dataset.

Cloud Masking: Used CloudSen12 to select the clearest Sentinel-2 acquisition within a 7-day pre-fire window. including cloudy parts forces the model to learn atmospheric artifacts rather than actual terrain features

After: Clearest Acquisition (Selected)

Sentinel-2 - Data: 2017-10-22 (Crop)

Filtering Process

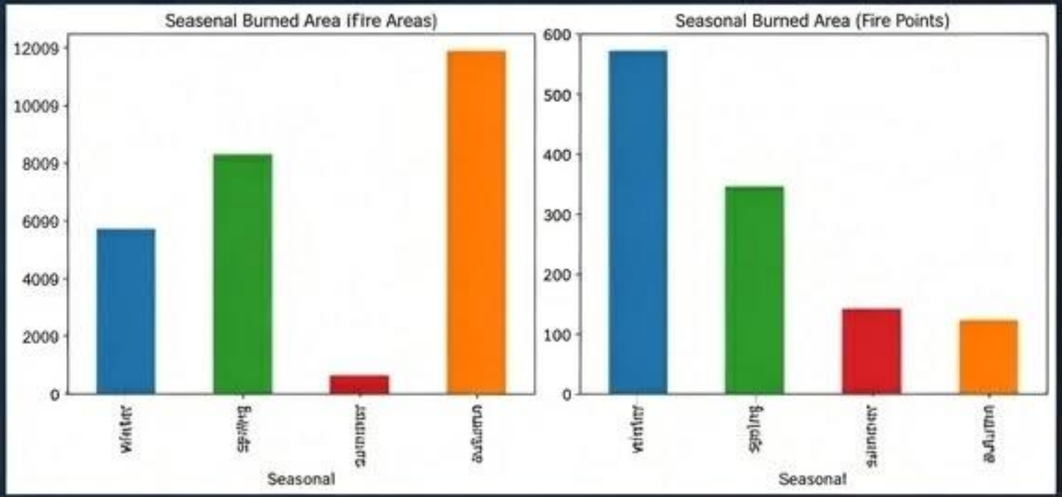
Before: Cloudy Acquisition

Sentinel-2 - Data: 2017-10-27 (Crop)

Dataset Characteristics & Challenges

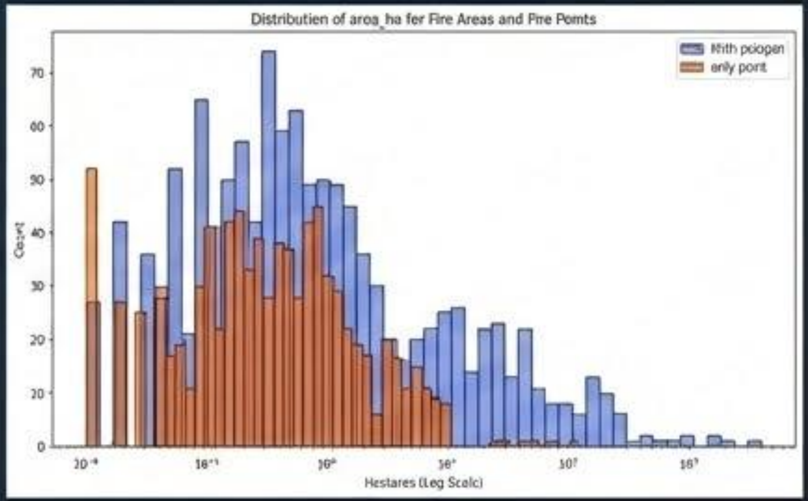
Scientific Context: Seasonal Patterns & Imbalance

Seasonal Disparities: Frequency vs. Severity



- The number of fires doesn't match how much land gets burned.
- Winter has the most fire starts, but they usually stay small.
- Autumn has fewer fires, but they grow much bigger and burn more land.
- This means we need to predict how *bad* a fire will be, not just if it will start.

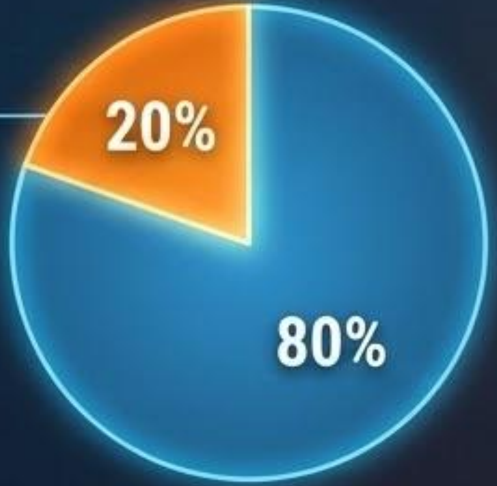
Size Distribution & Dataset Imbalance



- Our data is very unbalanced, with a 'long tail' of rare events.
- Most fires are tiny (less than 1 hectare, like a small field).
- The "Polygon" data (blue) shows the rare but huge fires (>100 ha) that cause the most damage.
- This imbalance makes it hard to teach a computer model to predict the big, important fires.

DATASET PREPARATION & CHALLENGES

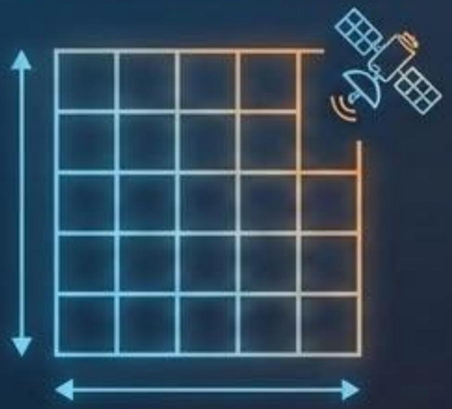
DATASET SPLIT (80/20)



TRAINING SET (872 Fires)
Diverse geographical locations (Piedmont)

VALIDATION SET (218 Fires)
Independent evaluation, maintaining diversity

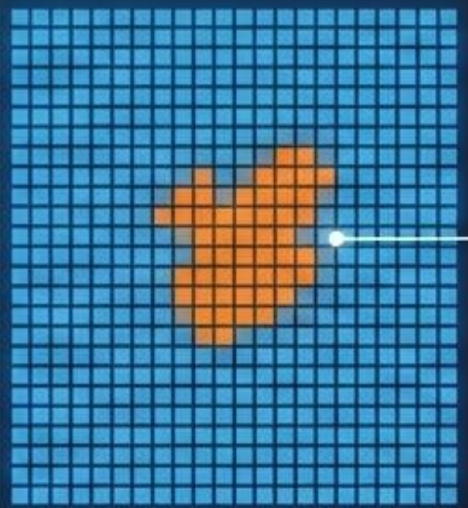
INPUT CONFIGURATION



256x256
PATCHES

All inputs processed as 256×256 patches for sufficient spatial context and computational feasibility.

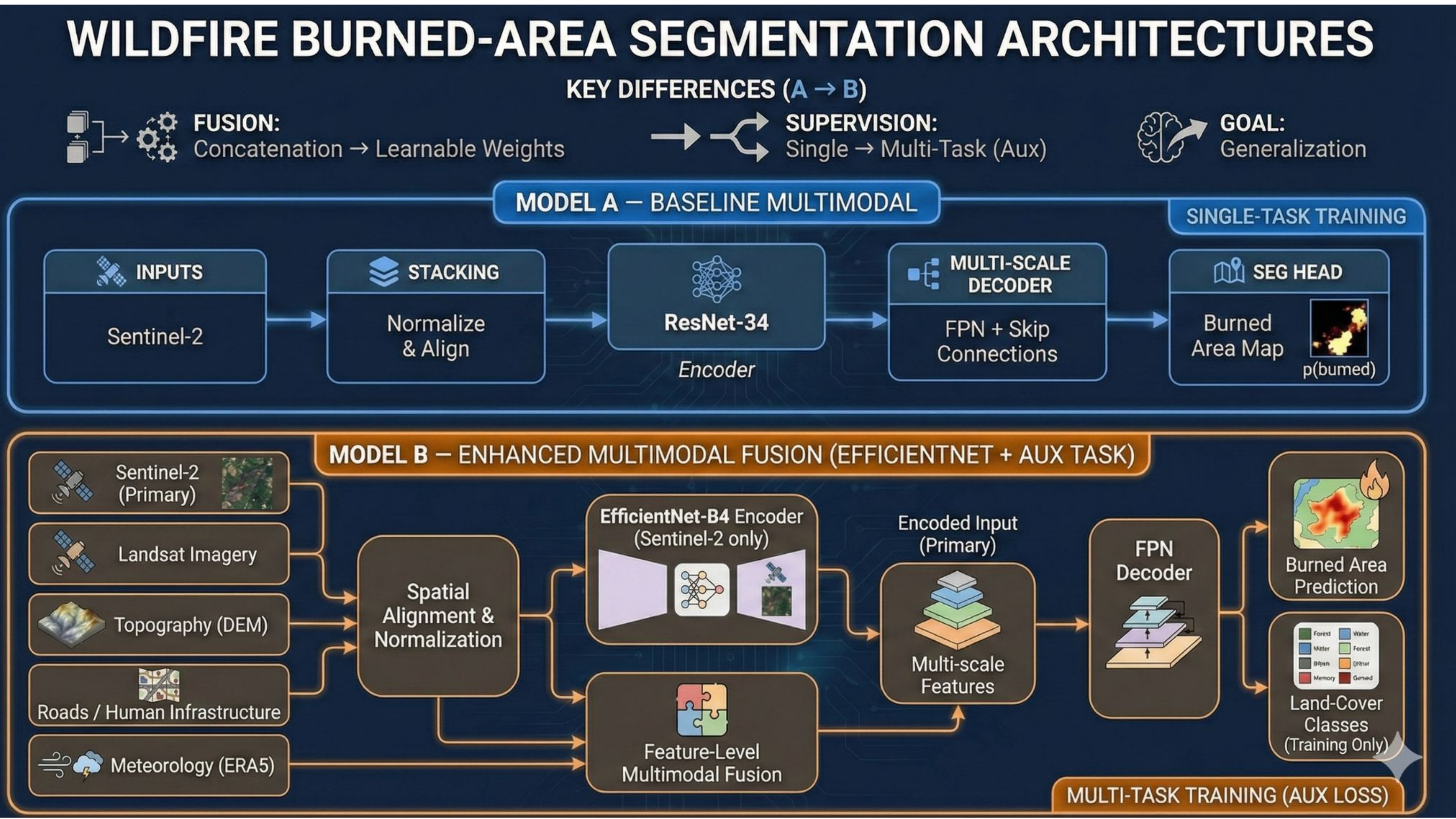
DATASET CHALLENGE: EXTREME CLASS IMBALANCE



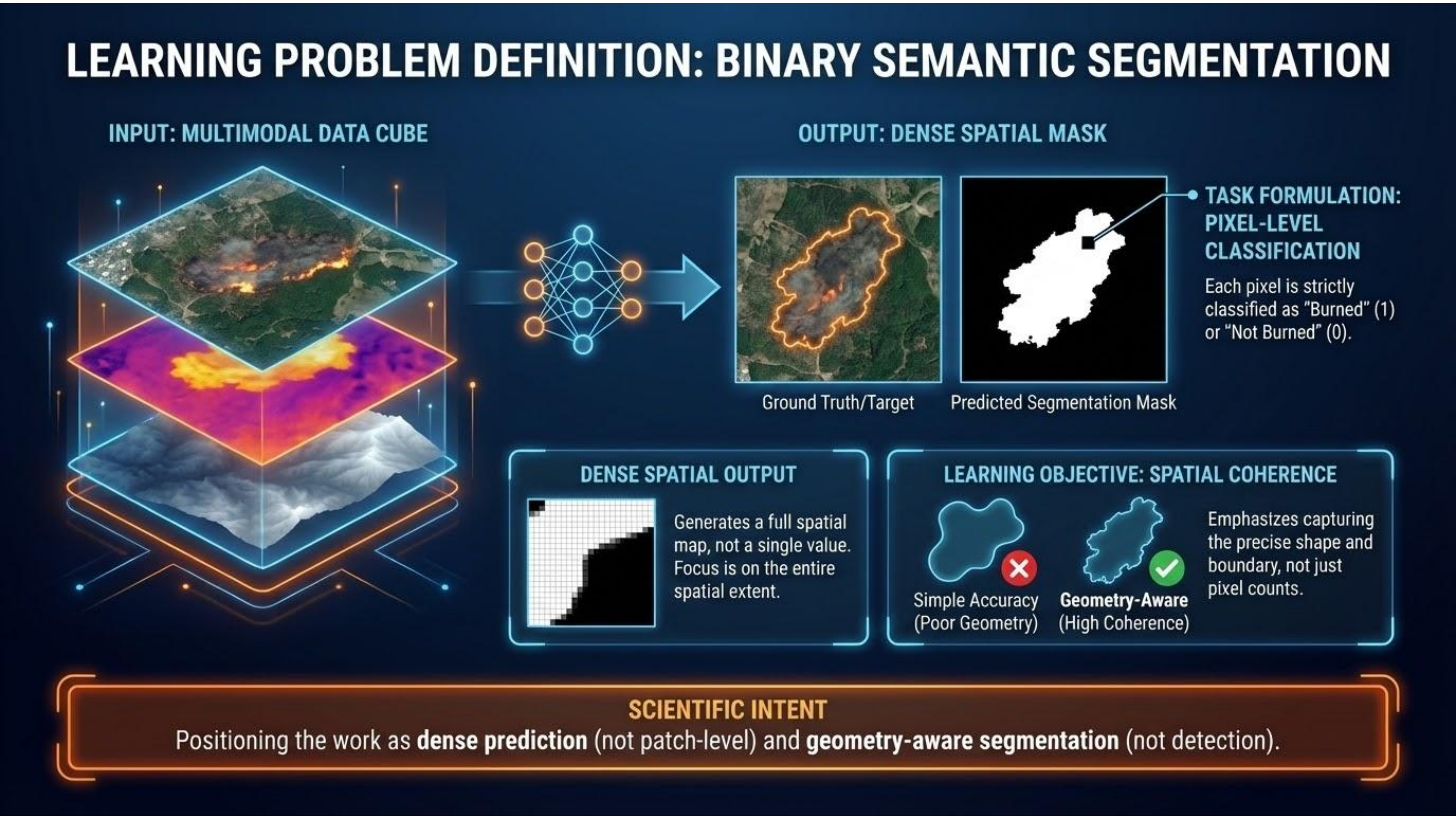
Burned pixels are rare (e.g., <1%). Standard models bias towards the majority background class.

SOLUTION: MULTITASK LEARNING STRATEGY

Jointly optimizing burned area segmentation to address imbalance.



WILDFIRE BURNED-AREA SEGMENTATION ARCHITECTURES
KEY DIFFERENCES (A → B)
FUSION:
Concatenation → Learnable Weights
SUPERVISION:
Single → Multi-Task (Aux)
GOAL:
Generalization
MODEL A — BASELINE MULTIMODAL
SINGLE-TASK TRAINING
INPUTS
Sentinel-2
STACKING
Normalize & Align
ResNet-34
Encoder
MULTI-SCALE DECODER
FPN + Skip Connections
SEG HEAD
Burned Area Map
p(bumed)
MODEL B — ENHANCED MULTIMODAL FUSION (EFFICIENTNET + AUX TASK)
Sentinel-2 (Primary)
Landsat Imagery
Topography (DEM)
Roads / Human Infrastructure
Meteorology (ERA5)
Spatial Alignment & Normalization
EfficientNet-B4 Encoder (Sentinel-2 only)
Feature-Level Multimodal Fusion
Encoded Input (Primary)
Multi-scale Features
FPN Decoder
Burned Area Prediction
Land-Cover Classes (Training Only)
MULTI-TASK TRAINING (AUX LOSS)


LEARNING PROBLEM DEFINITION: BINARY SEMANTIC SEGMENTATION
INPUT: MULTIMODAL DATA CUBE
OUTPUT: DENSE SPATIAL MASK
Ground Truth/Target
Predicted Segmentation Mask
TASK FORMULATION: PIXEL-LEVEL CLASSIFICATION
Each pixel is strictly classified as "Burned" (1) or "Not Burned" (0).
DENSE SPATIAL OUTPUT
Generates a full spatial map, not a single value. Focus is on the entire spatial extent.
LEARNING OBJECTIVE: SPATIAL COHERENCE
Simple Accuracy (Poor Geometry)
Geometry-Aware (High Coherence)
Emphastzes capturing the precise shape and boundary, not just pixel counts.
SCIENTIFIC INTENT
Positioning the work as dense prediction (not patch-level) and geometry-aware segmentation (not detection).

Baseline vs Multinodal: Quantitative Gain

EXPERIMENTAL SETUP

Parameter	Configuration
Model	Baseline vs Fusion
Encoder	ResNet34 vs EffNet-B4
Input	Sentinel vs Multimodal
Epochs	40 vs 120

Table 1. *Controlled setup differences including encoder architecture, input modalities, and training schedule duration.*

PERFORMANCE METRICS

IOU IMPROVEMENT
+0.251 `0.152 → 0.403`

F1 IMPROVEMENT
+0.311 `0.263 → 0.574`

Insight: Main gain comes from recall (0.212 → 0.572), indicating significantly fewer missed burned pixels in early detection scenarios.

Methodology: Baseline evaluated on quick CPU subset (~120 val samples); thresholds {0.50, 0.75, 0.90, 0.95}. Multimodal data sourced from

RQ3 – Which modalities contribute most?

METHOD

We estimate modality contribution using inference-time ablation: run the same trained multimodal model on the same validation subset, but **zero-out one modality at a time.** The performance drop (ΔIoU, ΔF1) indicates contribution.

Conclusion: The modality causing the largest drop in IoU/F1 is the most influential. **Landsat** and **ERA5** show the highest contribution.

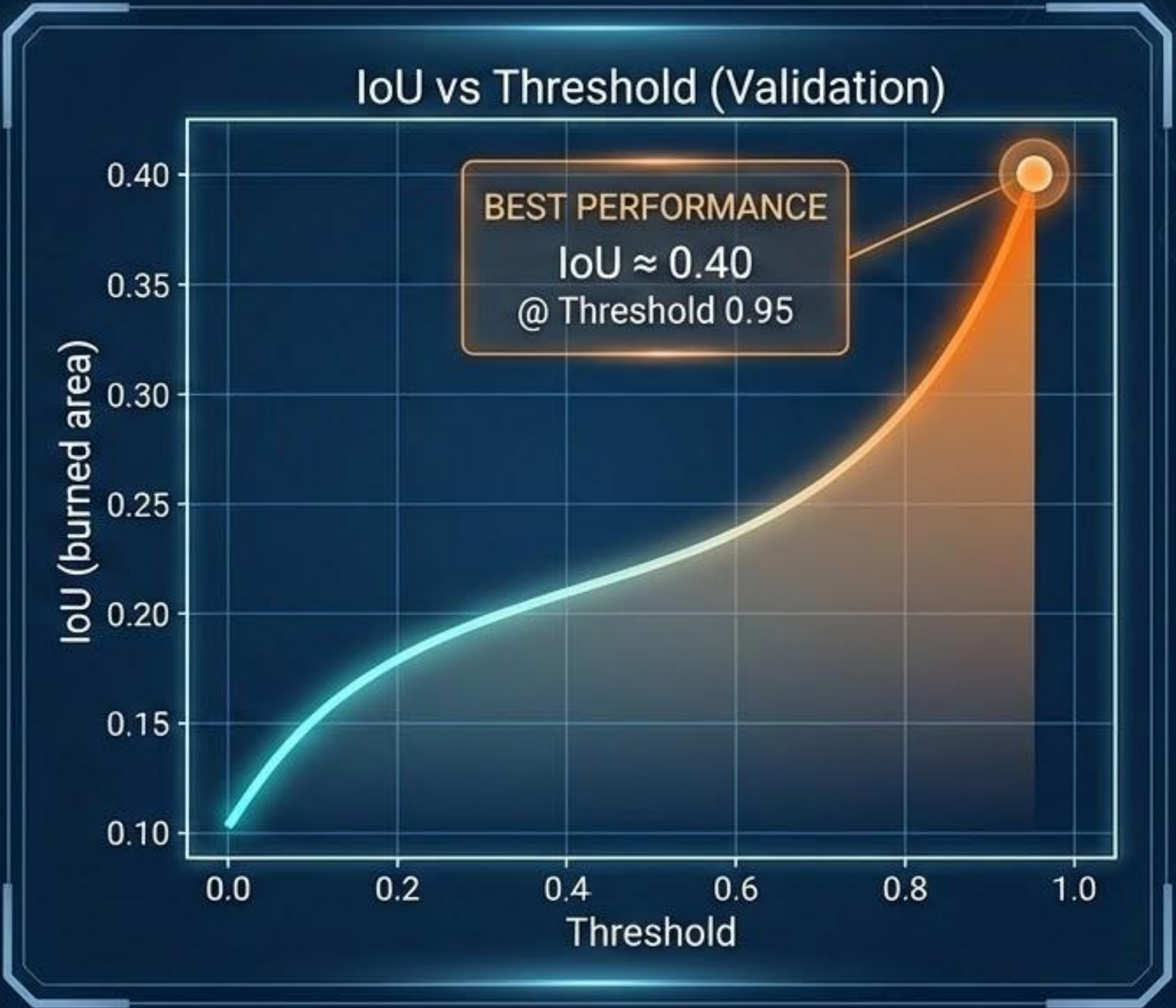
Landsat & ERA5 (High) → Other (Medium) → Ignition (Low)

Modality Importance: Performance Drop (Higher is Better)

EFFECT OF DECISION THRESHOLD ON IOU

- Default threshold (0.5) significantly underestimates performance.
- IoU increases monotonically with the threshold.
- **Best Performance:** Validation IoU ≈ 0.40 at threshold 0.95.

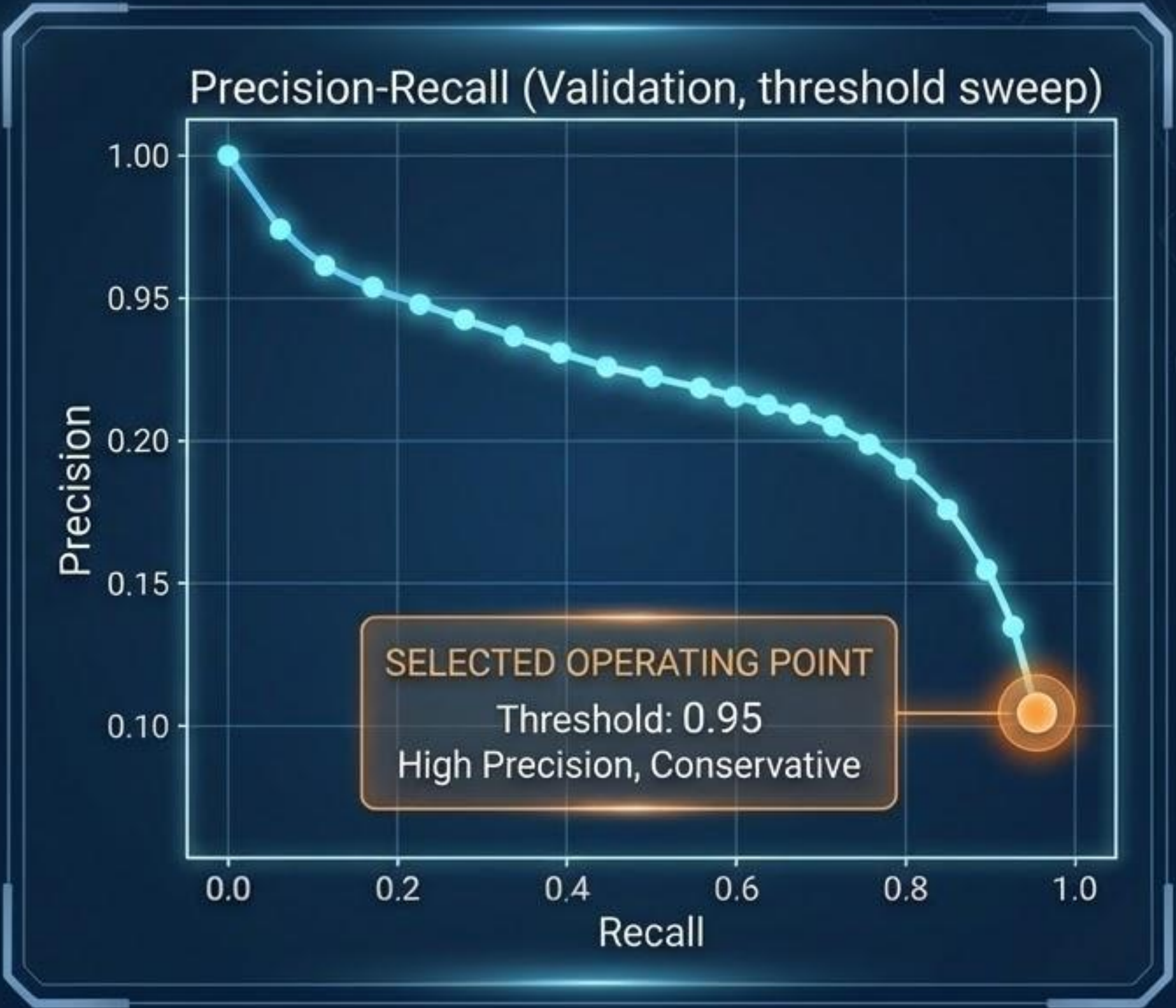
Threshold calibration improves IoU by **~70%** without retraining.



PRECISION–RECALL BEHAVIOUR

- **Low thresholds:** High recall but many false positives.
- **High thresholds:** Conservative predictions (high precision).

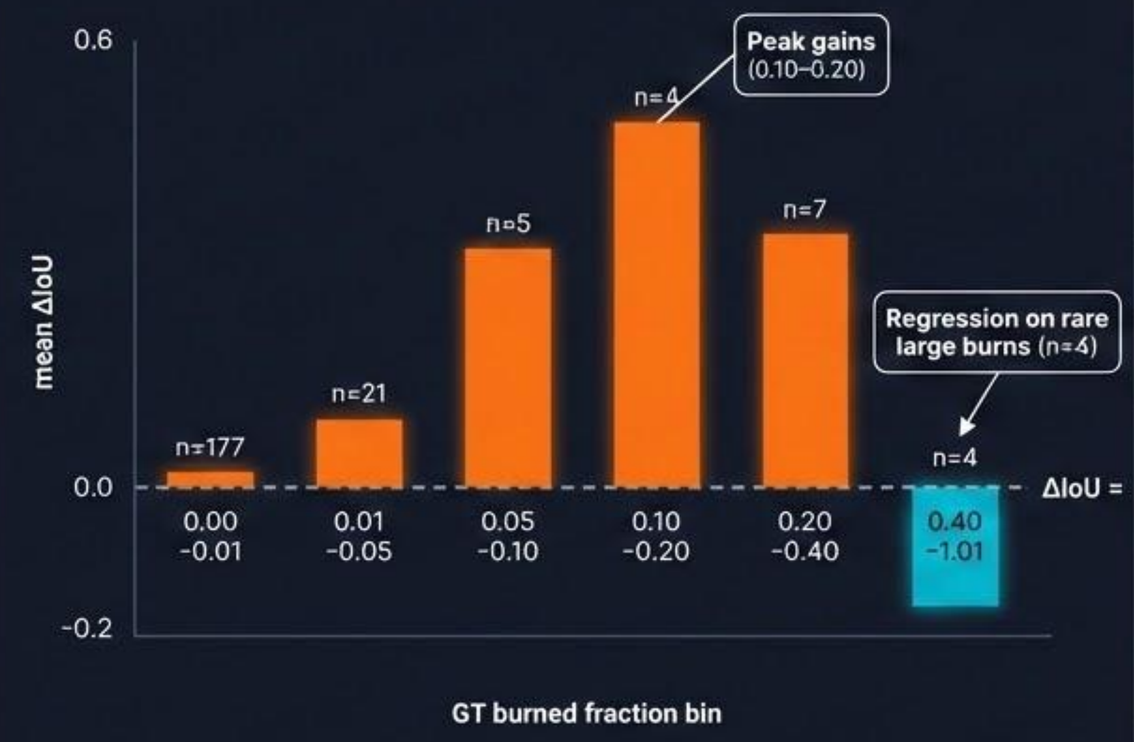
Decision: Selected operating point (0.95) prioritizes spatial accuracy to minimize false alarms.



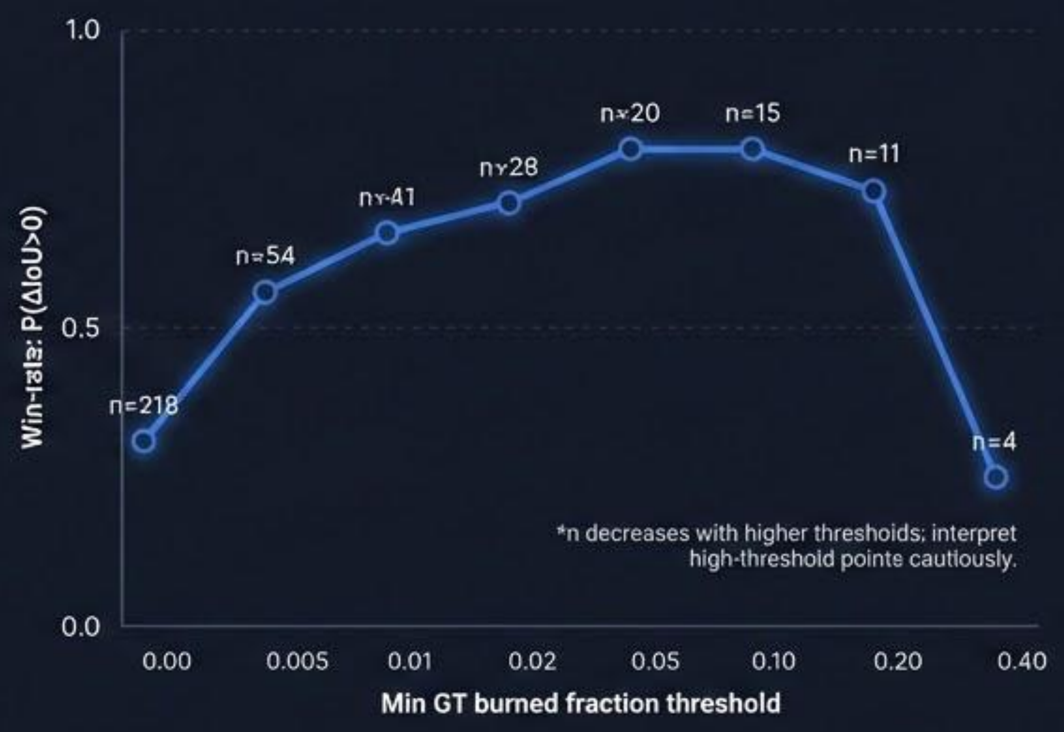
Effect of Adding Auxiliary Land-Segmentation Task in the Multimodal Model

Analysis of **ΔIoU by Burned-Fraction Bins** and **Cumulative Win-Rate Conditioned on Minimum Burn Size** (thr=0.95)

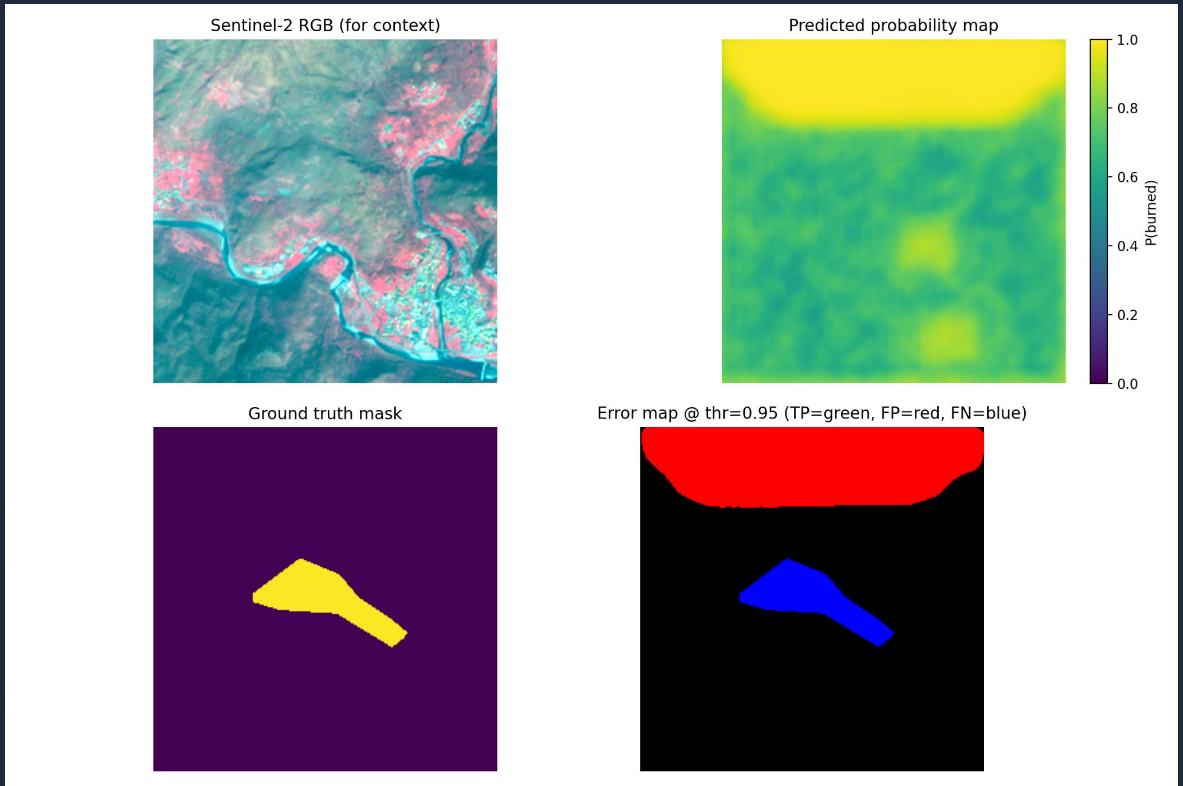
Mean ΔIoU by GT Burned-Fraction Bins

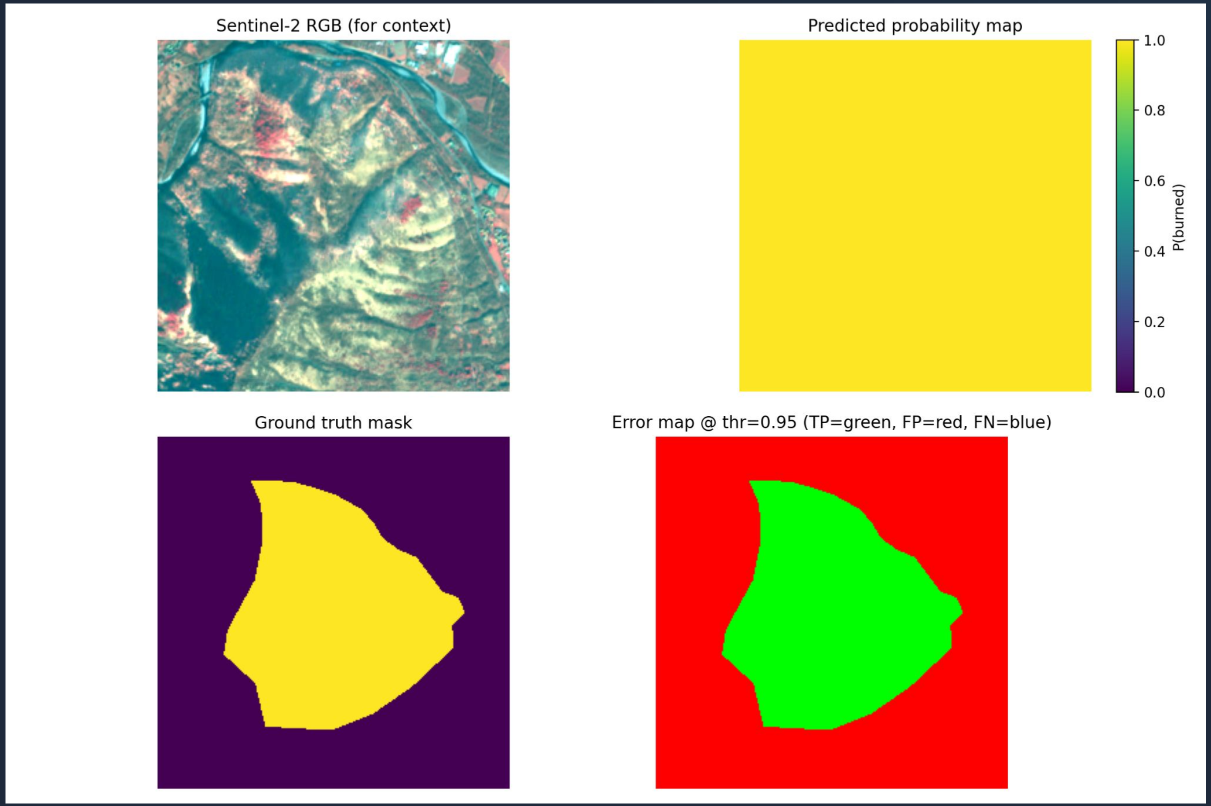


Cumulative Win-Rate: NEW (Auxiliary Task) vs. OLD (Baseline)



QUALITATIVE EVALUATION





SMALL / DIFFICULT FIRE

Model shows conservative behavior, avoiding noise.

LARGE / WELL-DETECTED FIRE

Large fires are detected with high spatial coherence.

Qualitative Results: Auxiliary task Model Captures Burn Scars More Accurately

Top improvement examples at thr = 0.95 (ΔIoU shown per case)

Case A: idx=125 | ΔIoU=+0.650 | **Key change:** Auxiliary task recovers missing interior burned area while keeping boundaries tight.

Case B: idx=10 | ΔIoU=+0.644 | **Key change:** Auxlllary task expands to match GT extent and reduces fragmented false gaps.

What improved (overall)?

- **Higher overlap** with GT (IoU↑) by reducing missed burned pixels (FN↓).
- **Better boundary alignment:** fewer under-segmented regions than baseline.
- At thr=0.95, predictions are cleaner and more spatially coherent.

CONCLUSION

Signal Utility
Pre-fire multimodal data provides useful signals for predicting final burned areas.

Data Integration
Integrating multiple data sources improves segmentation performance compared to a baseline approach.

Encoder Architecture
Encoder design plays a key role: a unified high-capacity encoder achieved better IoU than alternative encoder strategies.

Design Balance
Careful architectural choices can improve accuracy while keeping the model pipeline simple and reproducible.

"Overall, this project shows that **effective multimodal integration** combined with a **well-chosen encoder strategy** can significantly enhance burned-area prediction performance."

THANKS FOR YOUR ATTENTION!

ANY QUESTIONS?

Disclaimer: Our models are good, but they can't predict *your* questions.