Satellite Imagery Based Property Valuation

Submitted by:
Karra Nehru (Enrollment No: 23118039)
IIT Roorkee

Under:
CDC X Yhills Open Projects – 2025

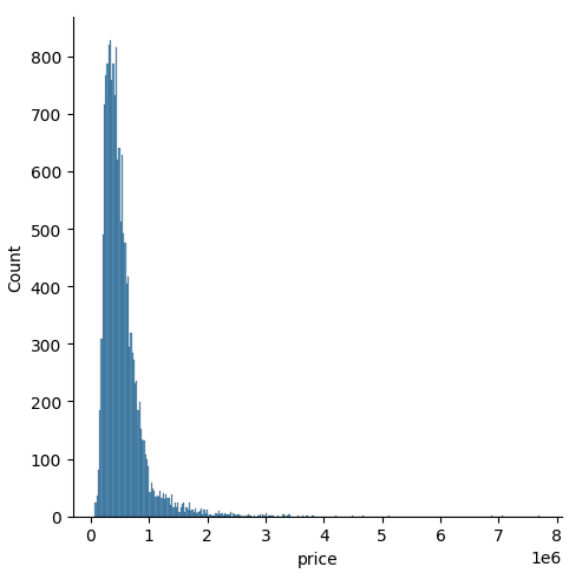
1. Overview

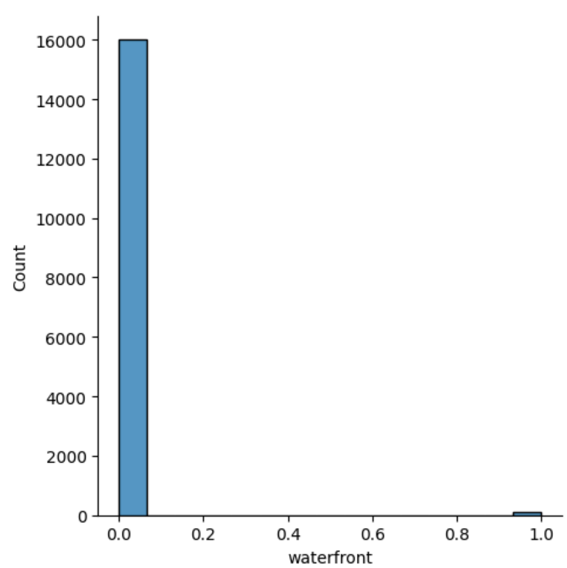
This project develops a multimodal regression pipeline to predict residential property prices by integrating structured tabular data with satellite imagery. A strong tabular XGBoost baseline is first established, followed by multimodal fusion approaches that incorporate visual environmental context extracted using convolutional neural networks (CNNs). The modeling strategy emphasizes data leakage prevention, robust feature engineering, and model explainability.

2. Exploratory Data Analysis (EDA)

2.1 Univariate Analysis

Sample data distribution plots :





Univariate distributions revealed that several numerical variables are **heavily right-skewed**, particularly *price , sqft_living, sqft_above, sqft_basement , sqft_lot, sqft_lot15 ,sqft_living15.*

The presence of extreme outliers and high skewness, especially in *price(skew=4.04)*, motivated the use of **logarithmic transformation** for the target variable and **yeo-johnson transformation** on skewed features to stabilize variance and improve model learning.

Analysis of discrete features showed:

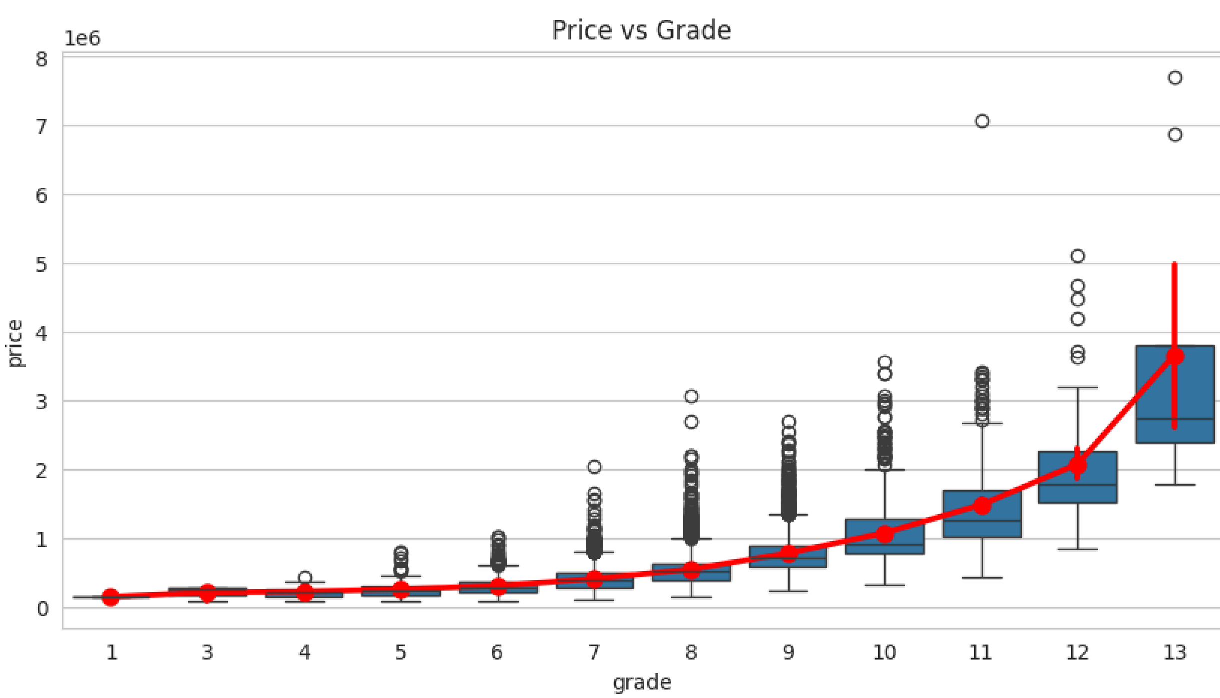
- There are Very few properties with *bedrooms* > 6
- Limited representation of *floors* > 3
- Strong imbalance in variables such as *waterfront* and high *view* ratings

These findings helped later in feature engineering decisions to reduce noise from sparse categories.

2.2 Bivariate & Correlation Analysis

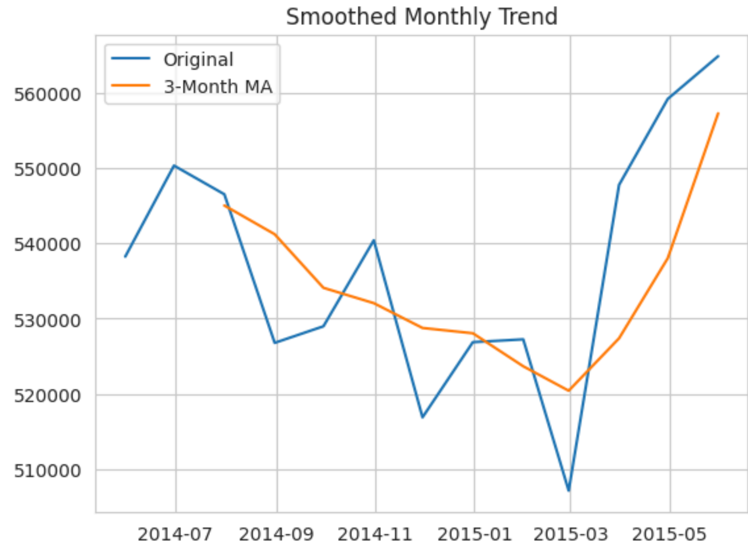
Correlation analysis between features and price revealed:

- **Strong positive correlation** with: *sqft_living (0.70),grade (0.66),sqft_above (0.60), sqft_living15 (0.58)*



2.3 Temporal Analysis

The *date* feature was decomposed into *year, month*, and *quarter*.



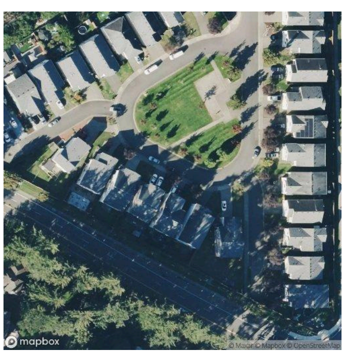
(blue: monthly average , orange: 3 months moving average)

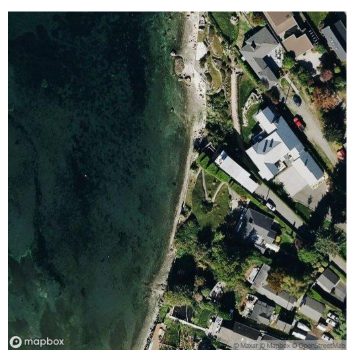
Temporal trends showed:

- A gradual price decline from mid-2014 to early-2015
- A sharp price rebound within the following 2–3 months, surpassing 2014 levels.

This confirmed the importance of incorporating temporal context into the feature set.

Satellite image samples (fetched by *data_fetcher.py* source: MapBox) **:-**



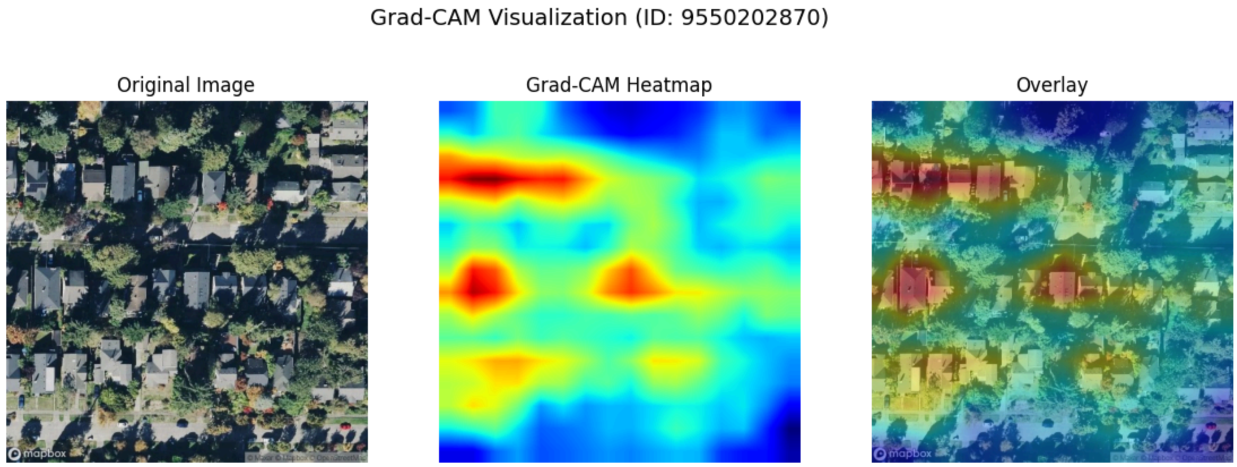


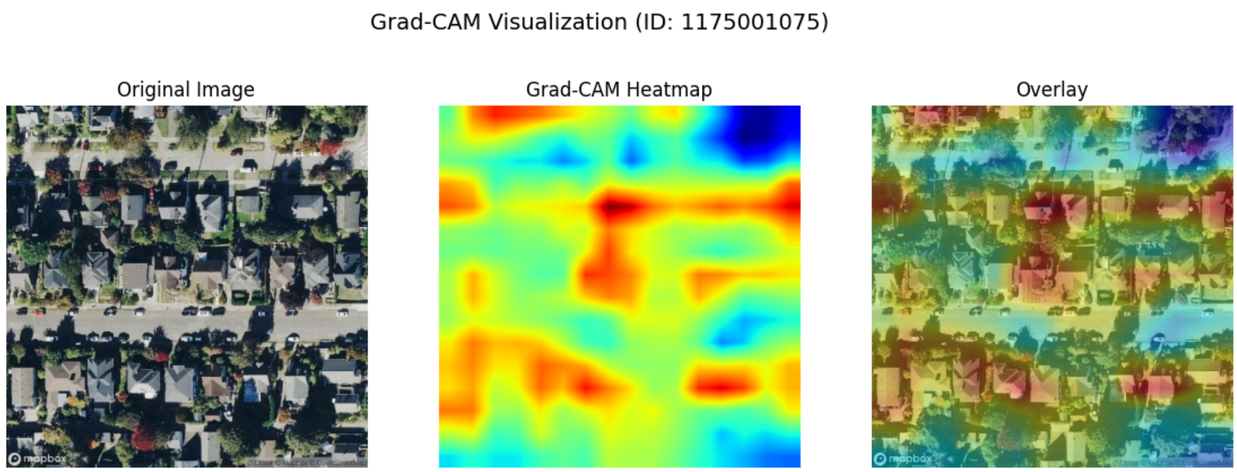
(id.jpg) {(53500760.jpg) (120059044.jpg) (121039083.jpg) }

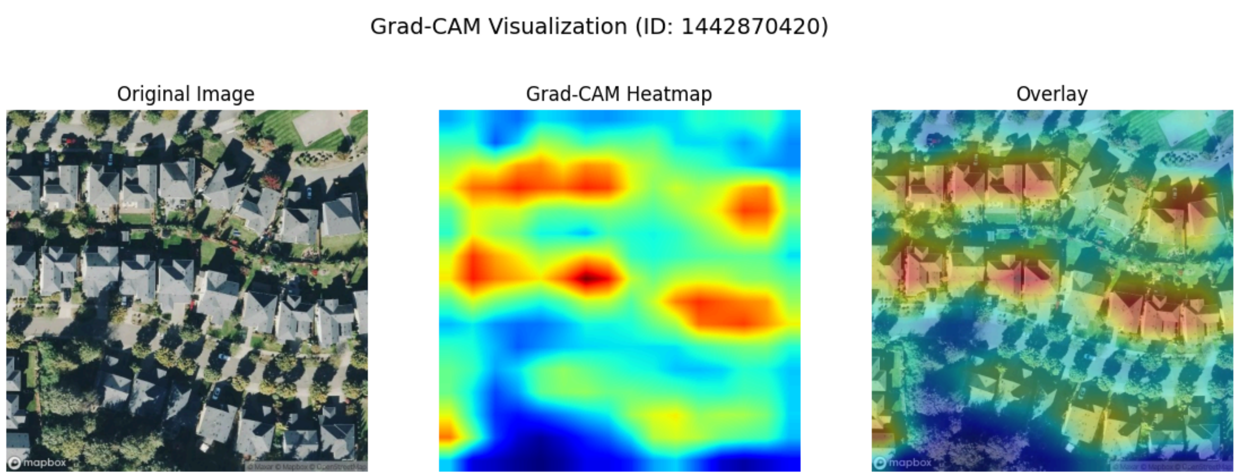
Sample satellite images show significant variation in neighborhood characteristics such as greenery, building density, road structure, and proximity to water. These observations justify the inclusion of satellite imagery to capture environmental context not available in tabular data.

3. Financial and Visual Insights

3.1 Grad-CAM visualization of predicted high valuation properties

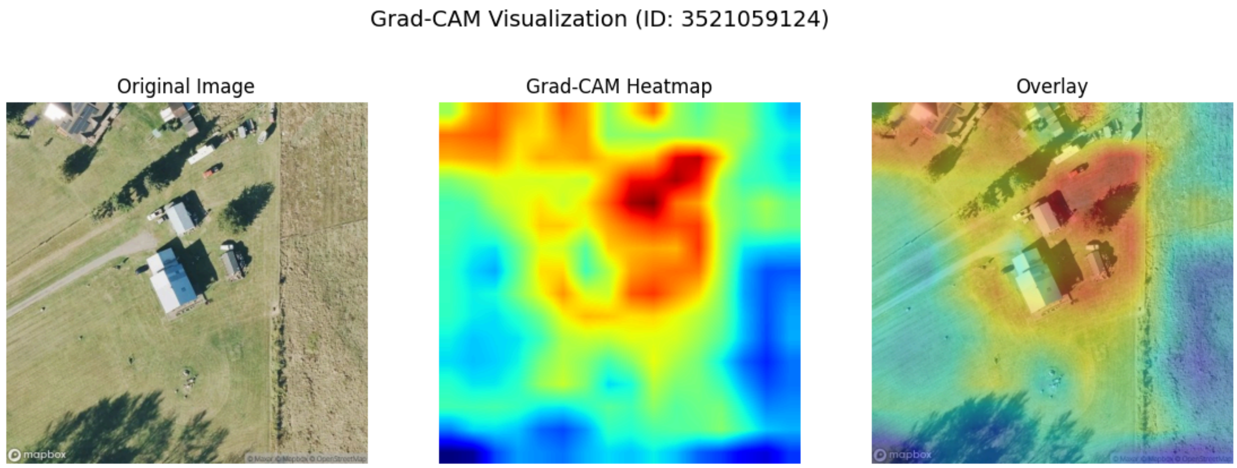


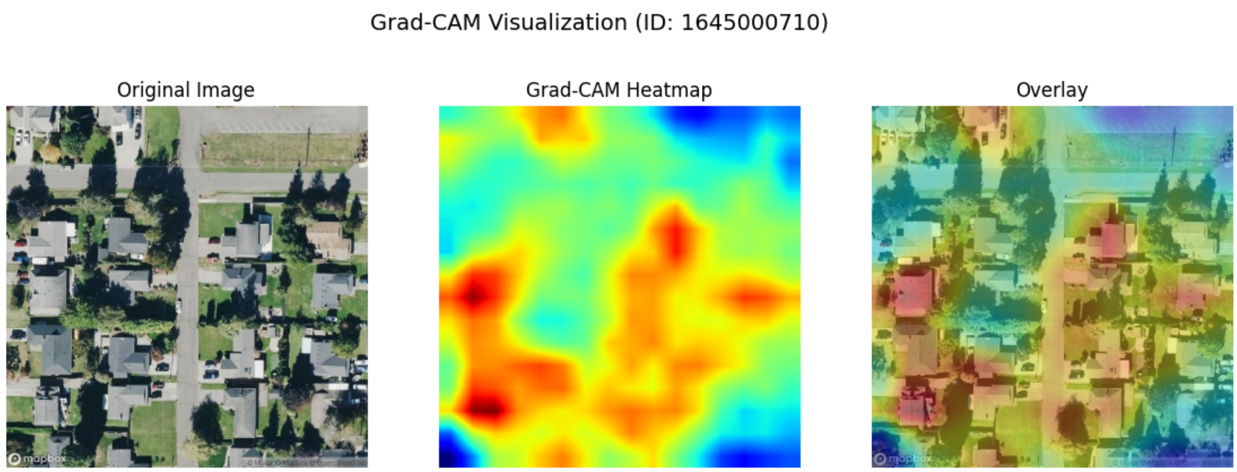


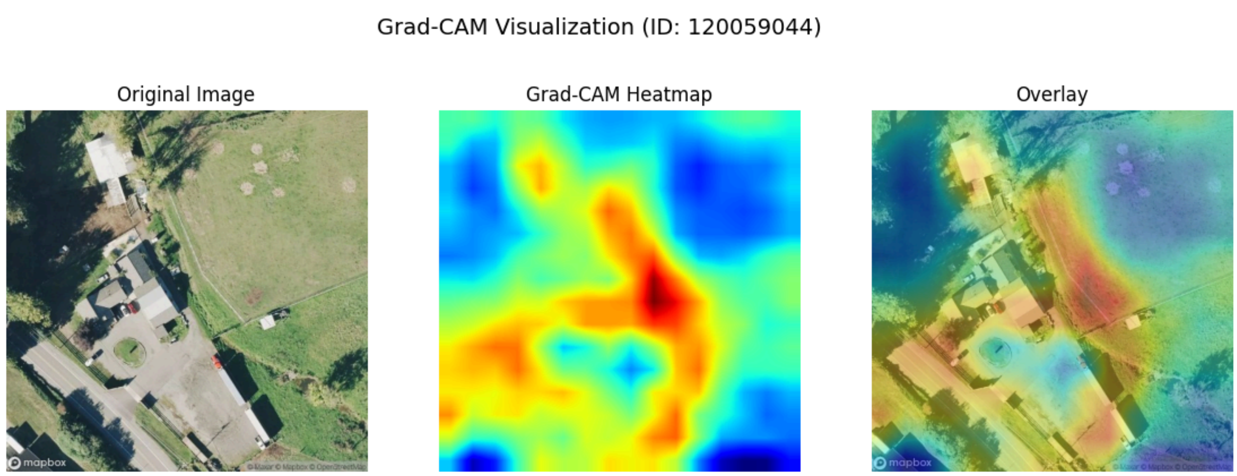


NOTE : The Red color indicates Max importance followed by the Yellow then the Green and Blue indicates Min importance in the displayed Heatmap

3.2 GRAD-CAM visualization of predicted low valuation properties







NOTE : The Red color indicates Max importance followed by the Yellow then the Green and Blue indicates Min importance in the displayed Heatmap

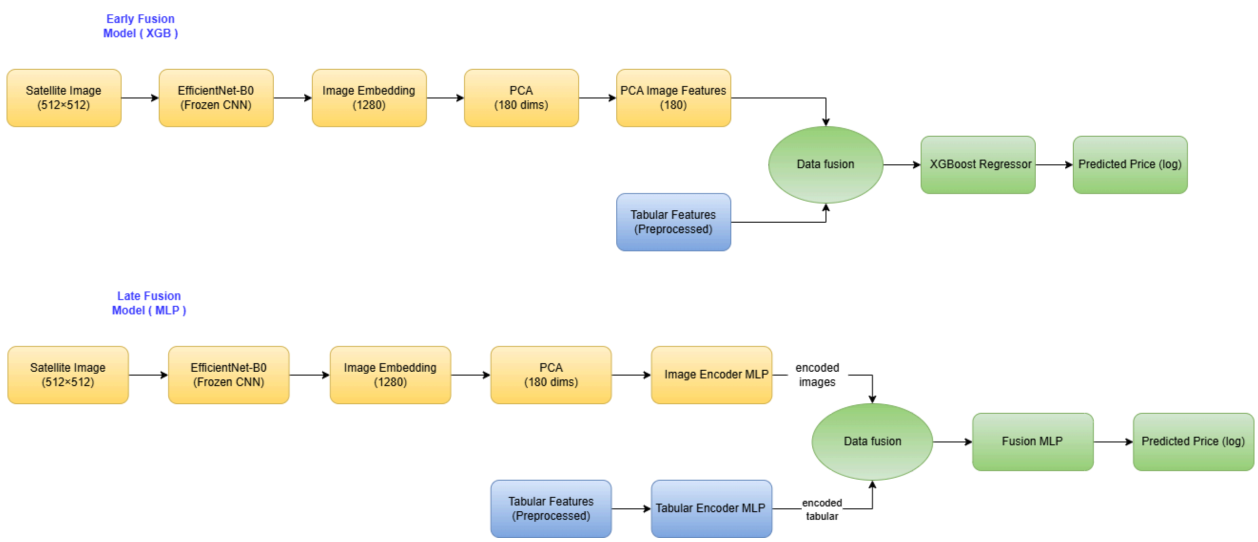
Observation :

An analysis of CNN-based image embeddings combined with Grad-CAM visualizations shows a clear link between visible environmental features and property values. Interestingly, the results challenge the common assumption that greener environments are always more desirable. In this dataset, areas with more vegetation and open land tend to be associated with lower property values.

In contrast, higher-valued properties are most often located in dense urban settings. These areas are visually defined by concrete surfaces, closely packed buildings, and repetitive structural patterns.

Overall , in the dataset analyzed, higher property values are mainly associated with denser urban areas and closer access to developed infrastructure. The results highlight that the relationship between visual environmental features and real estate values is more complex and sometimes less intuitive than commonly assumed.

4. Architecture Diagram



The multimodal architecture consists of two main components. Tabular features are processed through a feature engineering and preprocessing pipeline. Satellite images are encoded using a pretrained EfficientNet-B0 CNN, producing high-dimensional embeddings that are reduced via Principal Component Analysis (PCA).

In the **early fusion** approach, PCA-compressed image embeddings are concatenated with tabular features and passed to the XGBoost regressor, enabling the model to learn from both the structured and visual information at once.

In the **late fusion** approach, tabular and image features are processed separately through different neural network branches and are only combined at a higher, more abstract level. This setup allows us to test whether modeling complex interactions between the two feature sets leads to improved predictive performance.

5. Results and Model Comparison

Model performance was evaluated using RMSE and R^2 metrics on the original price scale.

Model	**RMSE**	**R^2**
Tabular XGBoost	104,680	0.910
Early Fusion (Tab + Img)	108,035	0.904
Late Fusion (Neural)	196,925	0.682

The tabular-only XGBoost model remained the strongest performer, underscoring the importance of structured property features in price prediction. However, when satellite imagery was incorporated, **the early fusion XGBoost model demonstrated noticeably better performance than the late fusion approach and achieved results comparable to the tabular baseline**. This suggests that early fusion is a more effective strategy for combining visual and tabular information, allowing satellite imagery to enhance predictions without overwhelming the structured signal.