CDC OPEN PROJECT 25-26

Satellite Imagery–Based Property Valuation

A Multimodal Machine Learning Study

Project Completed By: Sumit Sharma

Branch: Engineering Physics

Enrollment No.: 23123042

1. Overview: Approach and Modeling Strategy

1.1 Motivation

Property valuation is influenced not only by intrinsic attributes such as size, quality, and layout, but also by **extrinsic neighborhood characteristics** including greenery, infrastructure, road density, and proximity to water bodies. While traditional machine learning models rely heavily on structured tabular data, such neighborhood-level attributes are difficult to encode numerically.

This project investigates whether **satellite imagery**, when combined with structured housing data, can enhance property price prediction by providing visual environmental context.

1.2 High-Level Strategy

The overall strategy followed a **progressive, evidence-driven pipeline**:

1. Build a **strong tabular-only baseline** using classical ML models.
2. Perform **Exploratory Data Analysis (EDA)** to understand both numerical and spatial patterns.
3. Programmatically acquire **satellite imagery** using geographic coordinates.
4. Extract visual representations using a **pretrained CNN (ResNet-18)**.
5. Experiment with multiple **multimodal fusion architectures**.
6. Compare unimodal vs multimodal performance using RMSE and R^2.
7. Apply **Grad-CAM** to interpret visual features.
8. Select the **final model based on empirical performance**, not complexity.

This approach prioritizes **robustness, interpretability, and honest evaluation**.

2. Exploratory Data Analysis (EDA)

EDA was conducted to understand the distributional, spatial, and visual characteristics of the data.

2.1 Price Distribution Analysis

- Property prices exhibit a **right-skewed distribution**, with a small number of luxury properties.
- Median prices lie substantially below the maximum, indicating high variance.
- This skewness motivates the use of **tree-based models** and optional log-transformation during experimentation.

Key takeaway:
Price prediction is inherently noisy; absolute accuracy must be interpreted relative to scale.

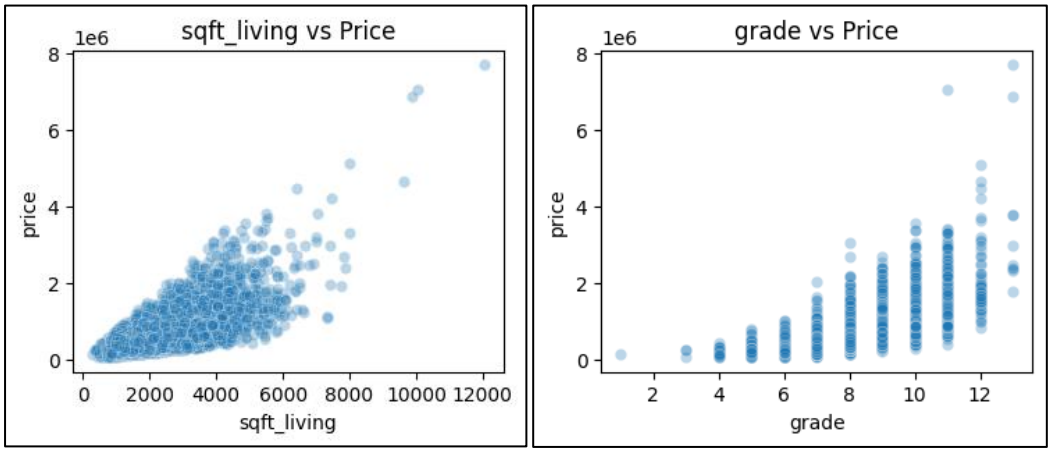
2.2 Feature Relationships (Tabular EDA)

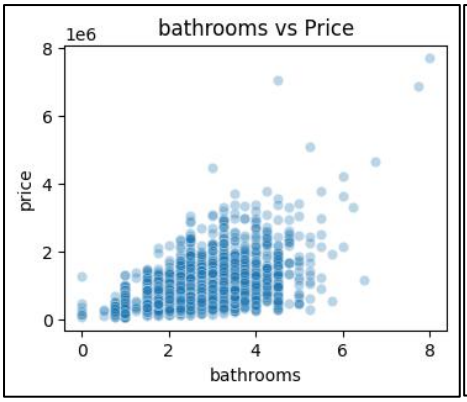
Key observations from correlation and scatter analyses:

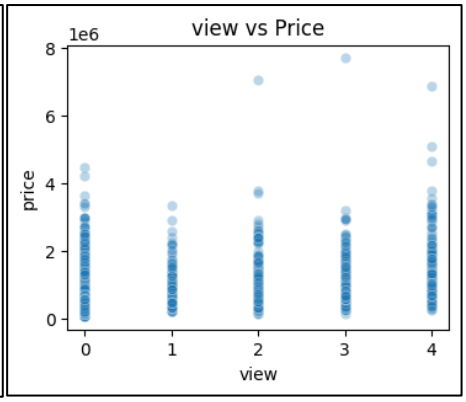
- sqft_living and grade show strong positive correlation with price.
- bathrooms and view contribute significantly to valuation.
- waterfront properties consistently command a premium.
- Latitude and longitude reveal **geographical clustering of high-value regions**.

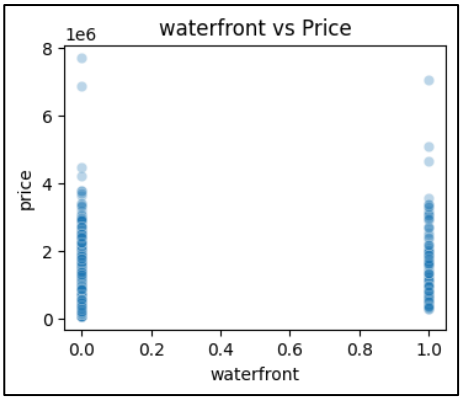
These findings justify:

- The use of ensemble methods (Random Forest)
- The inclusion of **location-derived visual context**









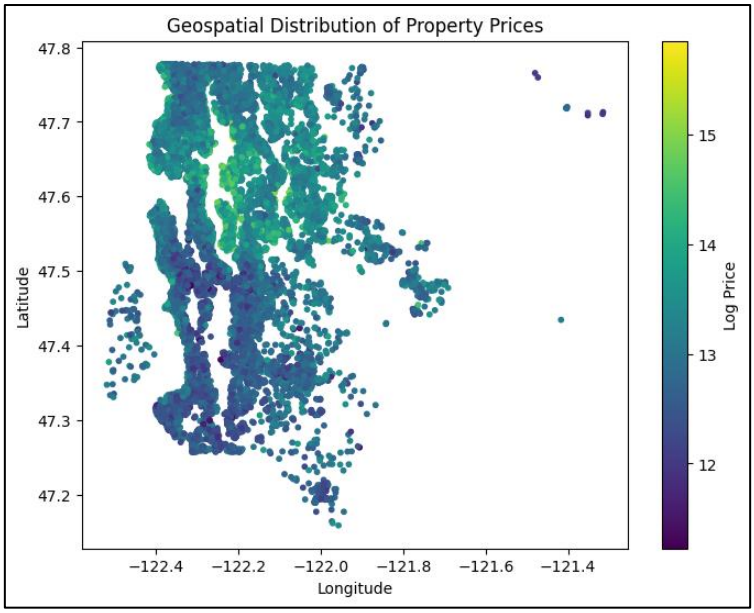
2.3 Geospatial & Visual EDA (Satellite Images)

Sample satellite images were visualized directly during EDA to qualitatively assess neighborhood context.

Sample Satellite Images downloaded using Mapbox API

From visual inspection:

- High-value properties are often located near:
  - Water bodies
  - Dense greenery
  - Well-connected road networks
- Lower-value regions tend to show:
  - High concrete density
  - Sparse vegetation
  - Industrial or highly congested layouts



EDA Insight:
Satellite imagery captures *macro-level neighborhood signals* that are not explicitly present in tabular data.

3. Financial and Visual Insights

This section connects **visual cues** to **economic interpretation**.

3.1 Green Cover vs Concrete Density

- Areas with visible tree cover and parks generally correspond to higher property values.
- Vegetation acts as a proxy for:
    - Environmental quality
    - Affluence
    - Lower population density

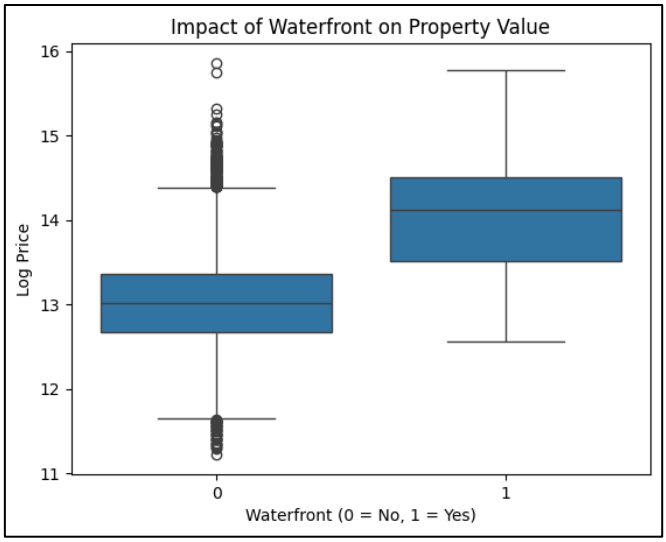
Conversely:

- Highly concrete-dense areas often reflect urban congestion or industrial zones, associated with lower average prices.

3.2 Waterfront Proximity

Satellite images clearly distinguish waterfront properties.

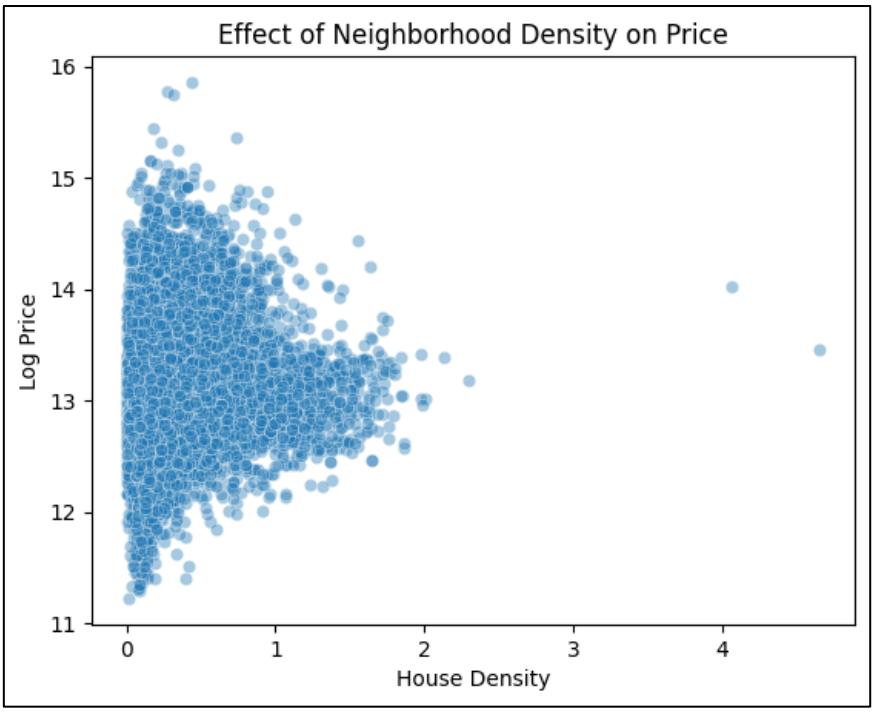
- Visual proximity to lakes or rivers strongly correlates with premium pricing.
- This aligns with tabular feature waterfront, validating that satellite imagery encodes meaningful economic signals.



3.3 Road and Infrastructure Patterns

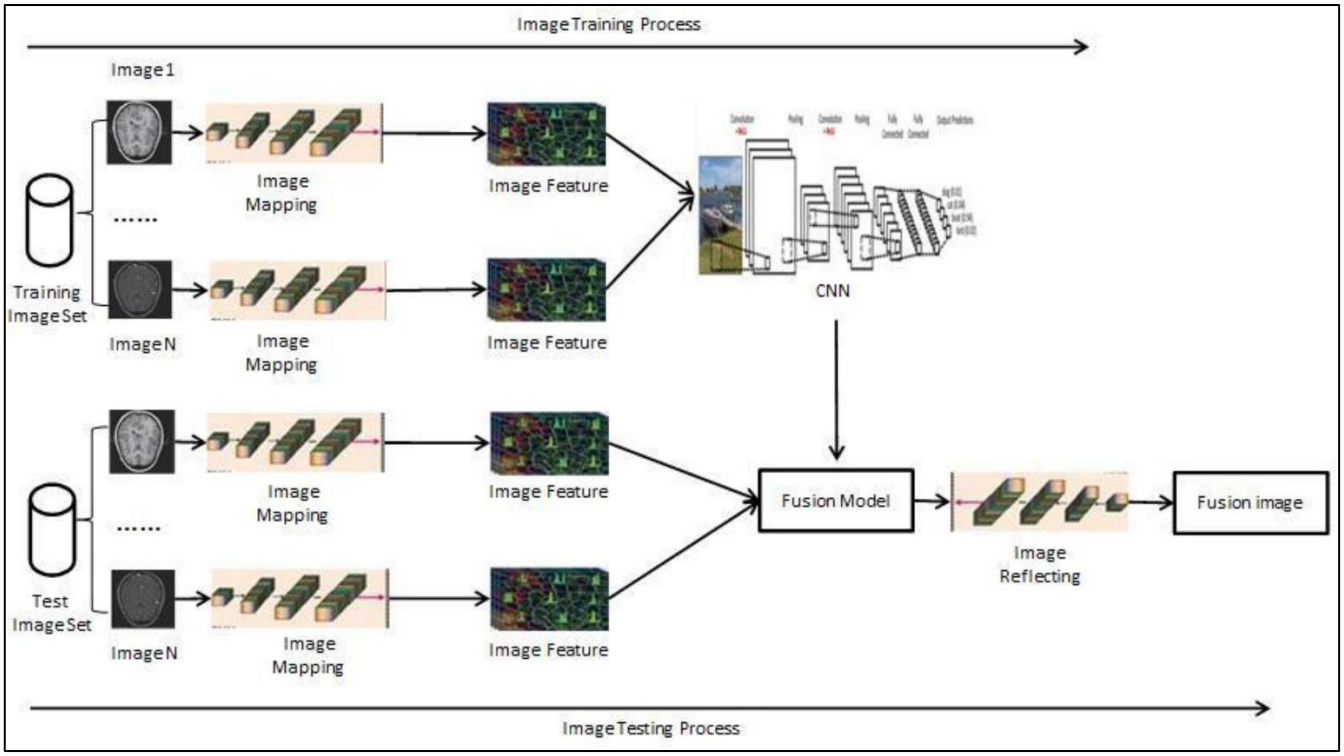
- Well-structured road networks indicate planned neighborhoods and accessibility.
- Gradual road layouts often outperform chaotic dense grids in valuation.

Conclusion:
Visual features extracted from satellite images have clear **economic intuition**, even if they do not always improve numerical model performance.



4. Architecture Diagram

The following diagram summarizes the multimodal pipeline explored in this project.



Architecture Description (Textual)

1. **Tabular Branch**
    - Structured housing attributes
    - Direct input to regression model
2. **Image Branch**
    - Satellite image
    - Pretrained ResNet-18
    - 512-dimensional visual embedding
3. **Fusion Layer**
    - Concatenation of tabular and image features
    - Optional dimensionality reduction (PCA)
4. **Prediction Head**
    - Regression model (Random Forest / Ridge / MLP)

5. Modeling and Results

5.1 Tabular-Only Models

Multiple models were tested on structured data alone.

Best-performing model: Random Forest Regressor

Metric	Value
RMSE	~129,000
R^2	~0.87

This establishes a **strong baseline**.

5.2 Multimodal Models (Tabular + Satellite Images)

Several fusion strategies were evaluated.

Model	RMSE	R^2
RF + CNN embeddings	~160,000	~0.78
MLP + CNN embeddings	~195,000	~0.66
PCA + Ridge (CNN + Tabular)	~170,000	~0.74

5.3 Interpretation of Results

Key observations:

- Multimodal models **did not outperform** the tabular baseline.
- High-dimensional CNN embeddings introduced noise.
- ImageNet-pretrained CNNs are not task-aligned for real estate valuation.
- Limited dataset size (~4k images) constrained multimodal learning.

Critical Insight:

Multimodal learning is not inherently superior; its success depends on alignment between representation, task, and data scale.

6. Explainability with Grad-CAM

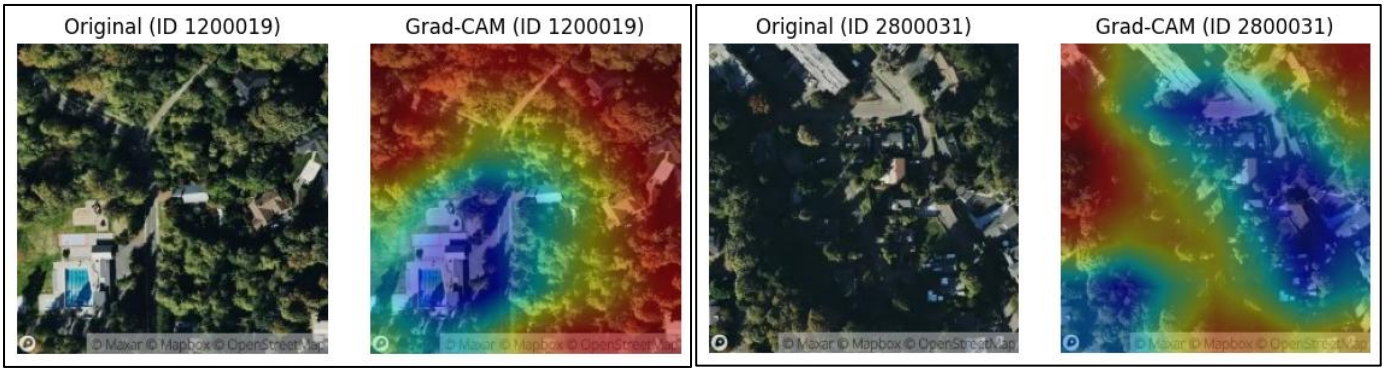
Grad-CAM was applied to the CNN to understand **what visual regions influenced feature extraction**.

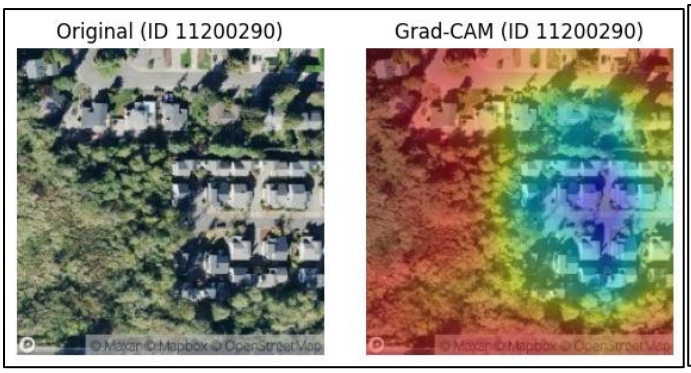
Observations:

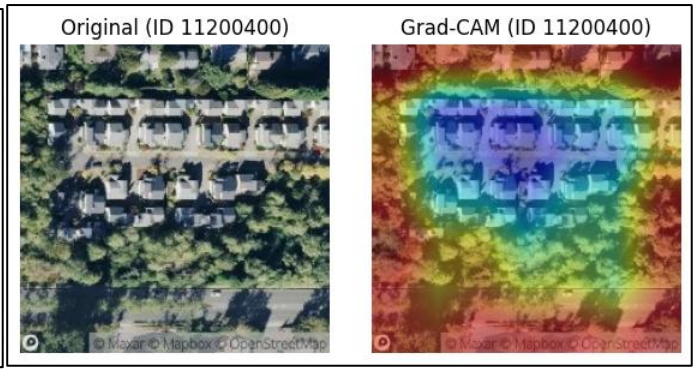
- Strong activation on:
  - Water bodies
  - Green spaces
  - Major road structures
- Minimal attention on homogeneous regions

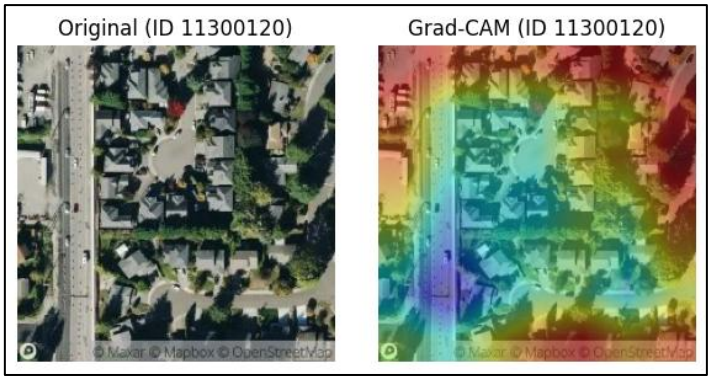
This confirms that:

- The CNN captures **meaningful environmental context**
- Lack of performance gain is due to **fusion limitations**, not irrelevant visual features









7. Final Model Selection

Despite extensive multimodal experimentation, the **tabular-only Random Forest** was selected for final predictions because:

- It achieved the lowest RMSE
- It was the most stable across validation
- It was easier to interpret and deploy

The final prediction file was generated using this model.

8. Conclusion

- This project demonstrates a **complete, end-to-end multimodal ML pipeline**, from data acquisition to explainability. While satellite imagery provided valuable qualitative insights into neighborhood context, naive fusion with pretrained CNN embeddings did not yield quantitative improvements over a strong tabular baseline.
- The key takeaway reflects a real-world ML principle:
- **Model effectiveness is determined by empirical evidence, not architectural complexity.**

9. Future Work

- Fine-tuning CNNs on real estate–specific imagery
- Learning region-level embeddings instead of full-image features
- Graph-based modeling of spatial neighborhoods
- Temporal modeling of housing price trends