Satellite Imagery-Based Property Valuation

A multimodal machine learning approach to predict residential property prices by combining structured housing data with satellite imagery.

Author: Yashwant Saini
Enrollment No: 23119058
Domain: Machine Learning & Remote Sensing

GitHub: github.com/Y-shwnt/

January 2026

Contents

1 Introduction

1.1 Traditional Property Valuation

Property valuation is a core problem in real estate analytics and is traditionally addressed using structured, tabular data. Most automated valuation models rely on features such as property size, number of bedrooms and bathrooms, construction quality, year of construction, and basic location indicators like zip codes or geographic coordinates. Such tabular datasets are easy to collect, store, and model using standard machine learning techniques. As a result, they have been widely used in both academic research and industry-grade real estate valuation tools.

1.2 Limitations of Tabular-Only Models

While structured features capture important property-level details, they often fail to represent the surrounding environment in which a property exists. Two houses with nearly identical internal characteristics can have significantly different market values due to differences in neighborhood conditions.
Factors such as proximity to water bodies, availability of green spaces, road connectivity, urban density, and nearby infrastructure strongly influence buyer perception and pricing. However, these neighborhood-level characteristics are either weakly represented or completely absent in most tabular datasets. As a result, tabular-only models may miss important contextual information, leading to inaccurate or overly generalized price estimates.

1.3 Why Satellite Imagery Matters

Satellite imagery provides a direct and scalable way to capture environmental and neighborhood context. High-resolution satellite images encode visual patterns related to land use, vegetation coverage, water proximity, road networks, and overall urban layout. These visual cues closely align with how humans assess the desirability of a location. For example, properties surrounded by greenery, well-connected roads, or organized urban layouts are often perceived as more valuable. By leveraging satellite imagery, machine learning models can access spatial information that is difficult to quantify using traditional numerical features alone.

1.4 Motivation for a Multimodal Approach

The central motivation of this project is to explore whether combining structured housing data with satellite imagery can lead to more context-aware property valuation models. A multimodal learning approach allows tabular features to represent intrinsic property attributes, while satellite images contribute external neighborhood and environmental information.
Rather than replacing traditional tabular features, this approach aims to enhance existing valuation frameworks by integrating visual context that is otherwise missing. By learning jointly from both data sources, the model is expected to produce more accurate, robust, and realistic property price predictions.

2 Architecture Diagram

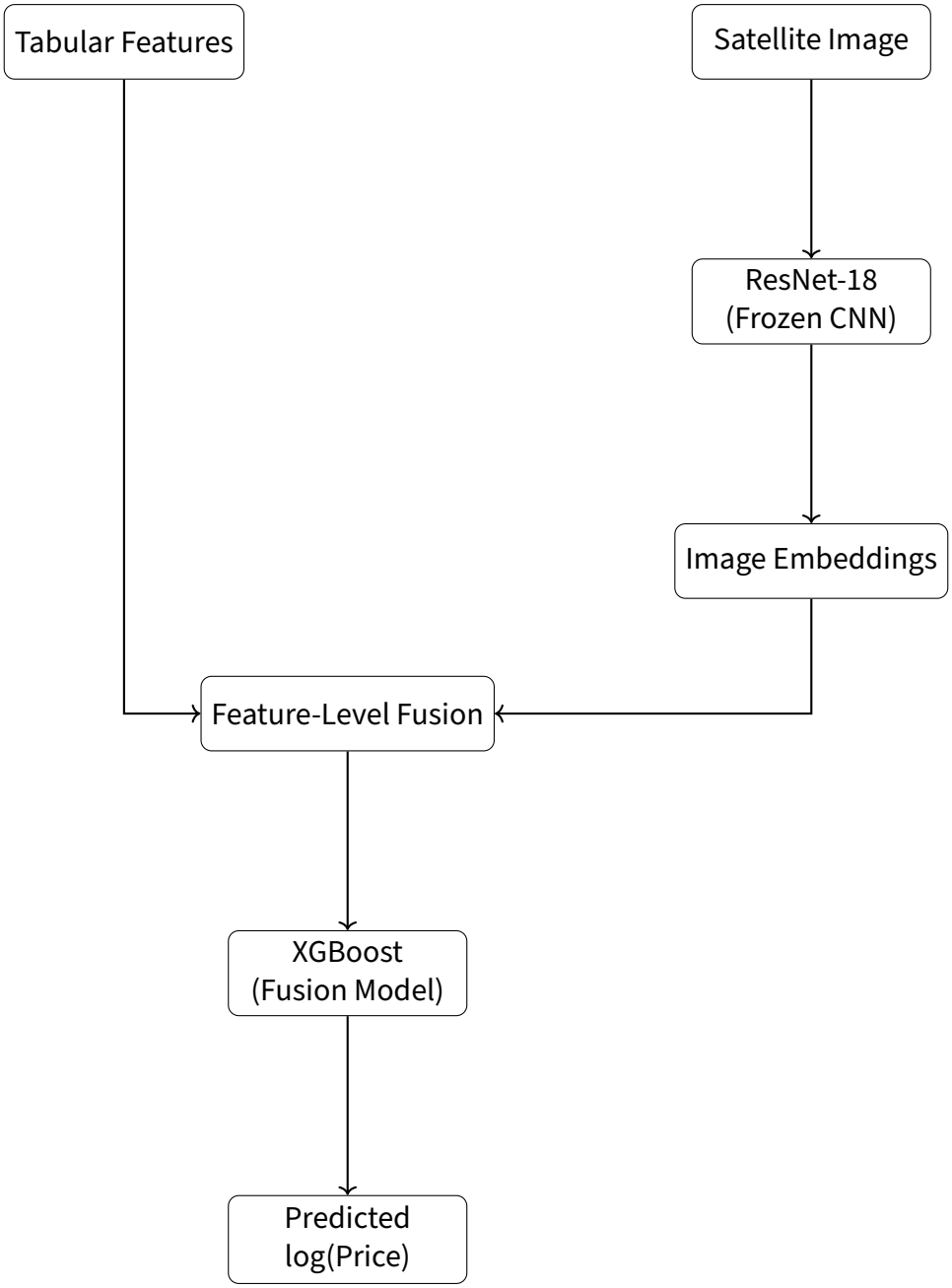

Figure 1: Architecture of the satellite–tabular fusion pipeline for property price prediction.

3 Problem Statement, Objectives, and Scope

3.1 Problem Statement

Traditional housing datasets primarily focus on structured property attributes such as size, number of rooms, and basic location information. While these features are important, they often lack critical environmental and neighborhood context that strongly influences property value. Factors such as neighborhood quality, availability of green spaces, surrounding infrastructure, and urban layout are either weakly captured or completely absent in tabular data.
As a result, models trained only on structured housing data may fail to distinguish between properties that are internally similar but located in very different environments. This project addresses this limitation by proposing a multimodal pipeline that integrates tabular property data with satellite imagery. The core challenge lies in effectively extracting meaningful visual features from satellite images and integrating these heterogeneous data types into a single, high-performance predictive model.

3.2 Project Objectives

The objectives of this project are as follows:

- To build a robust baseline property valuation model using only structured tabular housing data.
- To programmatically acquire satellite images using geographic coordinates (latitude and longitude) associated with each property.
- To extract high-level visual features from satellite images using a pretrained Convolutional Neural Network (CNN).
- To develop and evaluate image-only and multimodal regression models for property price prediction.
- To compare the performance of tabular-only, image-only, and multimodal models using standard regression metrics.
- To apply explainability techniques such as Grad-CAM to understand which visual regions in satellite images influence model predictions.

3.3 Scope and Expected Outcomes

The scope of this project is limited to residential property valuation using historical housing transaction data and publicly available satellite imagery. The study does not attempt to model short-term market fluctuations or external economic shocks.
The expected outcomes include a deeper understanding of the strengths and limitations of multimodal learning in real estate analytics, insights into how environmental and neighborhood context affects property prices, and the development of a reproducible end-to-end pipeline. This pipeline demonstrates both the potential benefits and practical challenges of integrating visual data into traditional machine learning workflows for property valuation.

4 Dataset Description

4.1 Tabular Dataset

Each row in the base dataset corresponds to a single residential property, with the target variable being the property price. The dataset includes a diverse set of structured features describing the physical characteristics of the property, its location, and its surrounding environment.
The dataset is split into a training set containing the target variable and a test set used solely for evaluation. After preprocessing, the training dataset contains **16,209 records**, while the test dataset contains **5,404 records**. No missing values were observed in either split.

Table 1: Summary of the housing dataset used in this study.

Dataset Property	**Value**
Training set shape	(16,209, 21)
Test set shape	(5,404, 20)
Number of modeling features	15
Null values	0

4.2 Target Variable

The target variable is the transaction price of a residential property. Exploratory analysis shows that raw property prices are highly right-skewed, with a small number of extremely high-valued properties.
To stabilize model training and improve regression performance, the target variable is transformed using a logarithmic transformation:

$$y = \log(1 + \text{price})$$

All models in this project are trained to predict log-transformed price values.

4.3 Tabular Feature Groups

The structured tabular dataset contains **21 columns in the training set** (including the target variable) and **20 columns in the test set**. After excluding the target variable, a total of **15 tabular features** are used for modeling. These features capture structural attributes, quality indicators, location information, and neighbourhood-level characteristics of residential properties.
The tabular features are summarized below:

- **Structural attributes:** `bedrooms`, `bathrooms`, `sqft_living`, `sqft_above`, `sqft_basement`, and `floors`.
- **Lot and neighbourhood characteristics:** `sqft_lot`, `sqft_living15`, and `sqft_lot15`.

- **Quality and condition indicators:** `grade` and `condition`.
- **Environmental indicators:** `view` and `waterfront`.
- **Location and temporal features:** `lat`, `long`, `zipcode`, `yr_built`, `yr_renovated`, and engineered `house_age`.

All tabular features are numeric and are standardized prior to model training. No missing values are present in the dataset, and therefore no imputation is required.

4.4 Satellite Imagery Dataset

To capture neighbourhood-level and environmental context not available in tabular data, satellite images are programmatically fetched using the geographic coordinates associated with each property. Satellite imagery is retrieved using a static maps API at a fixed zoom level of **18** and an image resolution of **256** $\times$ **256 pixels**, ensuring consistency across samples.
Due to API usage limits and computational constraints, satellite images are not collected for the entire dataset. Instead, images are fetched for a **stratified subset of properties** selected from the training data. Stratification is performed using price bins to ensure balanced representation across low-, mid-, and high-price ranges.
A slight discrepancy between the number of tabular records and available satellite images arises because duplicate property entries are removed by retaining only the most recent transaction record for each property. After this filtering:

- Training images used: **5982**
- Test images: **5396**

Each satellite image is strictly aligned with its corresponding tabular record, ensuring a one-to-one mapping between visual inputs and structured features.

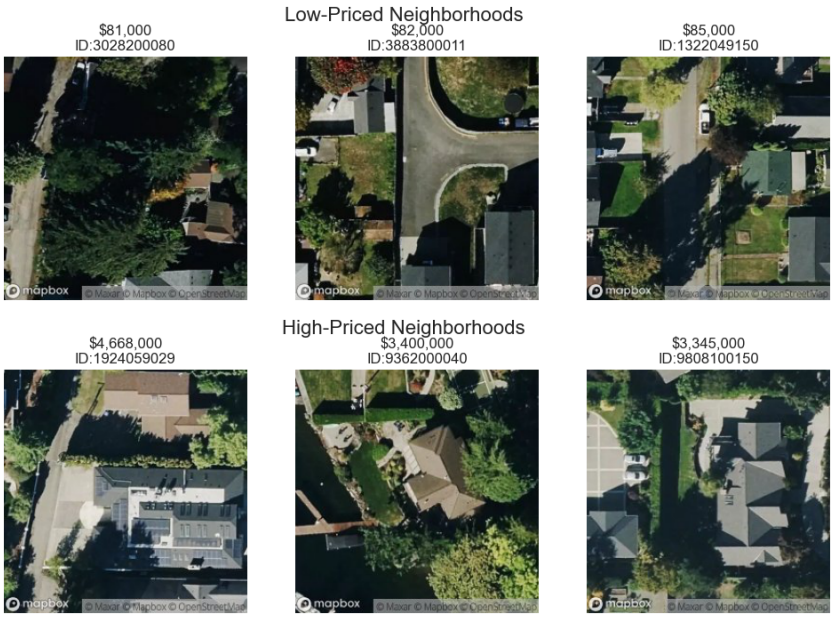

Figure 2: Representative satellite images sampled across low-, and high-price ranges.

4.5 Dataset Alignment and Integrity

All satellite image sampling and feature extraction are performed exclusively on the training dataset to avoid data leakage. The test dataset remains completely unseen during model training and feature extraction. Careful indexing and alignment ensure that each fused data point correctly represents the same property across tabular features, satellite images, and extracted image embeddings.

5 Exploratory Data Analysis (EDA) and Insights

5.1 Price Distribution

Property prices exhibit a highly right-skewed distribution, with a small number of extremely high-valued properties. This skewness increases variance and negatively affects regression performance, particularly for high-price outliers.

Figure 3: Distribution of raw property prices.

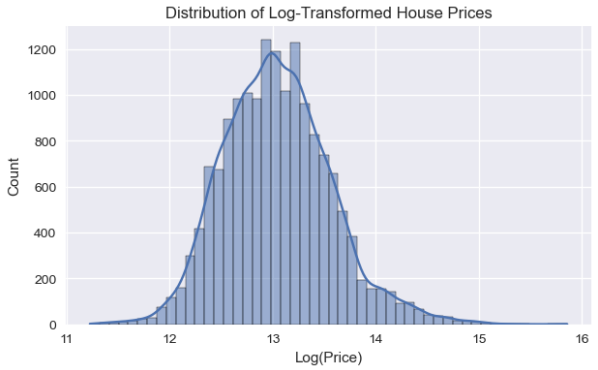

Figure 4: Distribution of log-transformed property prices.

To stabilize variance and reduce the impact of outliers, a logarithmic transformation $\log(1 + \text{price})$ is applied to the target variable. All subsequent models are trained using the log-transformed price.

5.2 Relationship Between Key Features and Price

- Living area (`sqft_living`) shows a strong positive relationship with property price. Larger homes are consistently associated with higher market values.

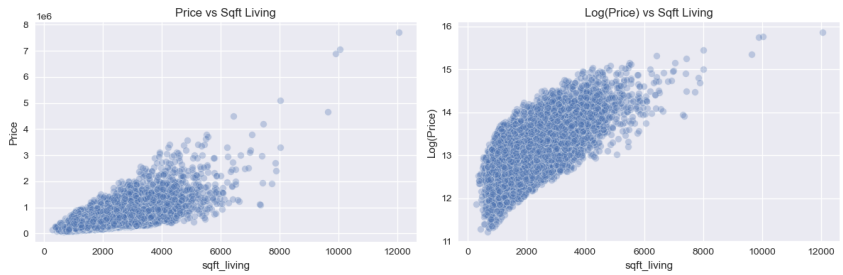

Figure 5: Relationship between living area and property price.

- Construction quality (`grade`) exhibits a clear monotonic relationship with price. Higher quality construction corresponds to higher property values.

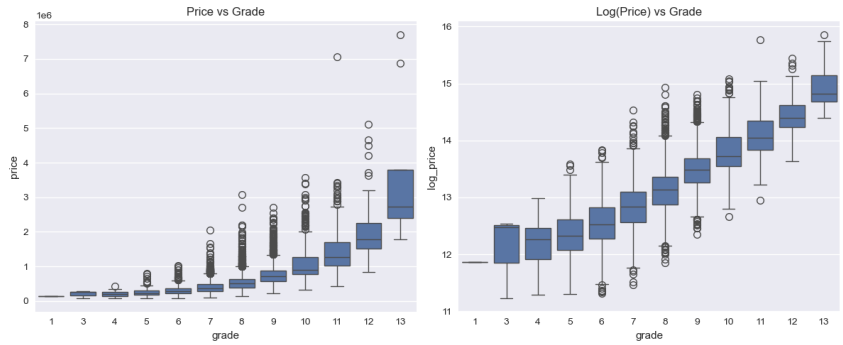

Figure 6: Property price variation across construction quality grades.

- Waterfront properties command a significant price premium compared to non-waterfront properties.

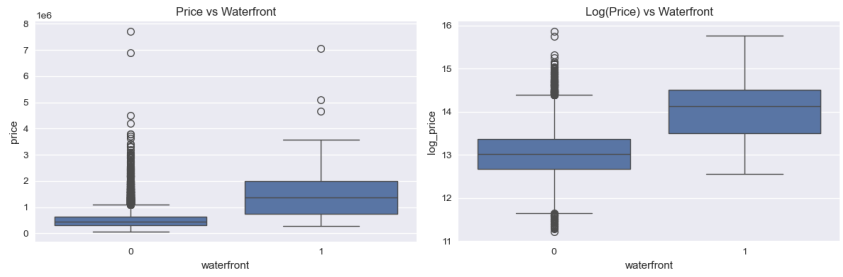

Figure 7: Impact of waterfront presence on property price.

- View quality (`view`) is positively associated with property price. Higher view ratings correspond to higher property values.

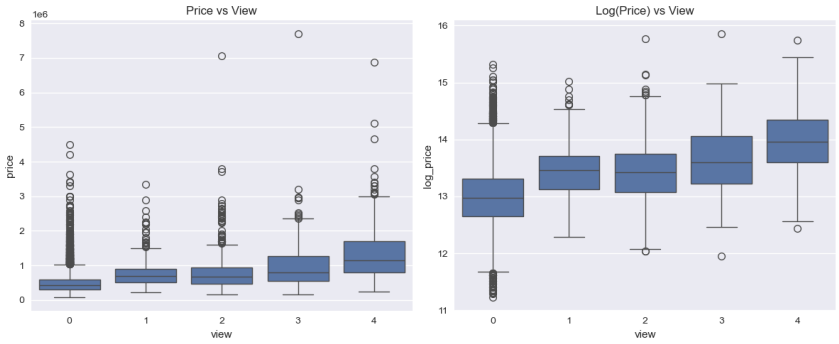

Figure 8: Relationship between view quality and property price.

- Property condition has a positive but weaker association with price compared to size and construction quality.

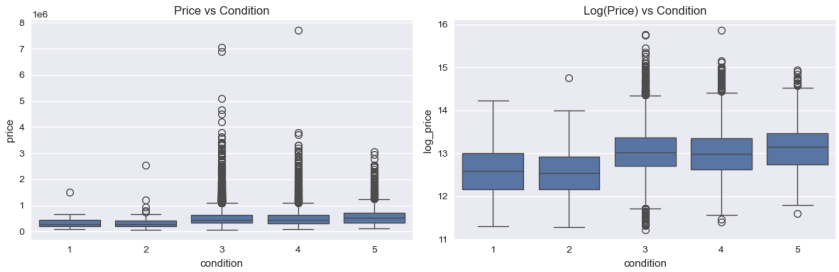

Figure 9: Property price variation across condition levels.

5.3 Feature Correlation Analysis

Correlation analysis highlights strong associations between property price and several structural and quality-related features, including `sqft_living`, `sqft_above`, number of bathrooms, and construction quality (`grade`). In contrast, lot size and renovation year show weaker correlations with price.

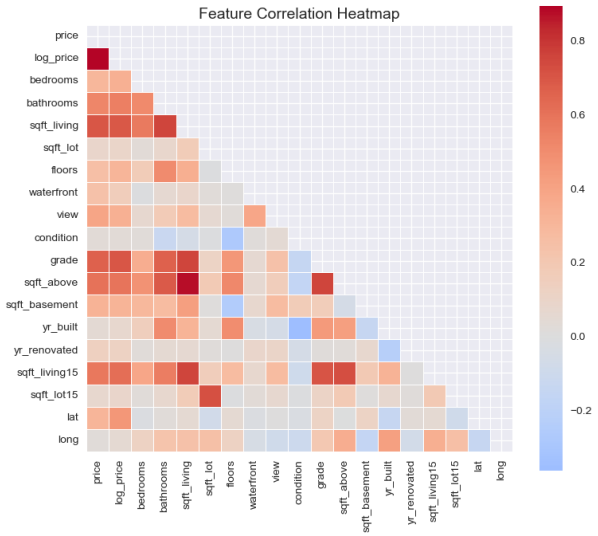

Figure 10: Correlation heatmap of key numerical features.

High intercorrelation among size-related variables (e.g., `sqft_living`, `sqft_above`, and `sqft_living15`) indicates multicollinearity, motivating the use of tree-based models that can effectively handle correlated features.

5.4 Geographic Distribution of Properties

Mapping property locations using latitude and longitude reveals clear geographic clustering of property prices. High-priced properties are concentrated near water bodies and premium residential zones, while lower-priced properties are more uniformly distributed inland.

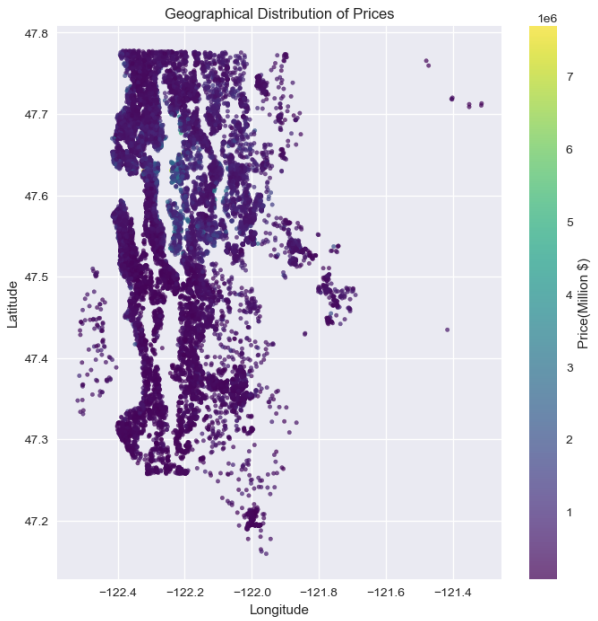

Figure 11: Geographic distribution of properties colored by property price.

This spatial clustering indicates that neighborhood-level context plays a significant role in property valuation and motivates the integration of satellite imagery.

5.5 EDA Takeaways

- Property prices are highly right-skewed, making log transformation necessary.
- Structural and quality-related features dominate price prediction.
- Waterfront presence and view quality introduce strong price premiums.
- Strong feature correlations justify the use of tree-based models.
- Property prices exhibit clear geographic clustering.
- Environmental and neighborhood context is likely to provide complementary signal.

6 Modeling Approach

6.1 Tabular Baseline Model

A baseline regression model was developed using structured tabular features to establish reference performance. The target variable (price) was log-transformed to reduce skewness, and numerical features were standardized prior to training.
A Gradient Boosted Decision Tree regressor (XGBoost) was used due to its strong performance on structured data, ability to capture non-linear relationships, and robustness to correlated features. This model serves as the primary baseline for comparison.

6.2 Image Feature Extraction

Satellite images were transformed into numerical representations using a pretrained Convolutional Neural Network (CNN). A pretrained ResNet architecture was used strictly as a fixed feature extractor, with the final classification layer removed to obtain fixed-length image embeddings.
All CNN weights were frozen during feature extraction to reduce overfitting and computational cost. The extracted embeddings encode neighborhood-level spatial and environmental patterns.

6.3 Image-Only Model

To evaluate whether satellite imagery contains standalone predictive signal, an image-only regression model was trained using the CNN-extracted image embeddings. The embeddings were used as inputs to a Gradient Boosted Decision Tree regressor (XGBoost), with the target variable being log-transformed property price.
This model is exploratory in nature and is used to assess the predictive contribution of visual features in isolation from structured tabular data.

6.4 Multimodal Fusion Model

A feature-level fusion strategy was adopted to combine structured tabular features with CNN-extracted image embeddings. A reduced subset of tabular features was selected for the fusion model to complement the visual representations and minimize redundancy. The concatenated feature vectors were passed to a Gradient Boosted Decision Tree regressor (XGBoost) for price prediction. This approach provides a stable and interpretable multimodal framework without relying on end-to-end neural fusion architectures.

6.5 Training and Evaluation

All models were trained to predict log-transformed property prices. Performance was evaluated using standard regression metrics, including R^2 and RMSE, on validation and held-out test sets. Strict separation between training and test data was maintained to avoid data leakage throughout the modeling pipeline.

7 Model Explainability

SHAP-Based Feature Attribution

Model predictions were interpreted using SHAP (SHapley Additive exPlanations) applied to the XGBoost regressors. SHAP quantifies the contribution of each feature to the predicted log-transformed property price.
The SHAP summary plot for the fusion model shows that structural and locational features such as `sqft_living`, `grade`, and geographic coordinates (`lat`, `long`) are the strongest predictors. Image-derived features appear lower in the ranking but contribute consistently, indicating that satellite imagery provides complementary signal rather than replacing tabular information.

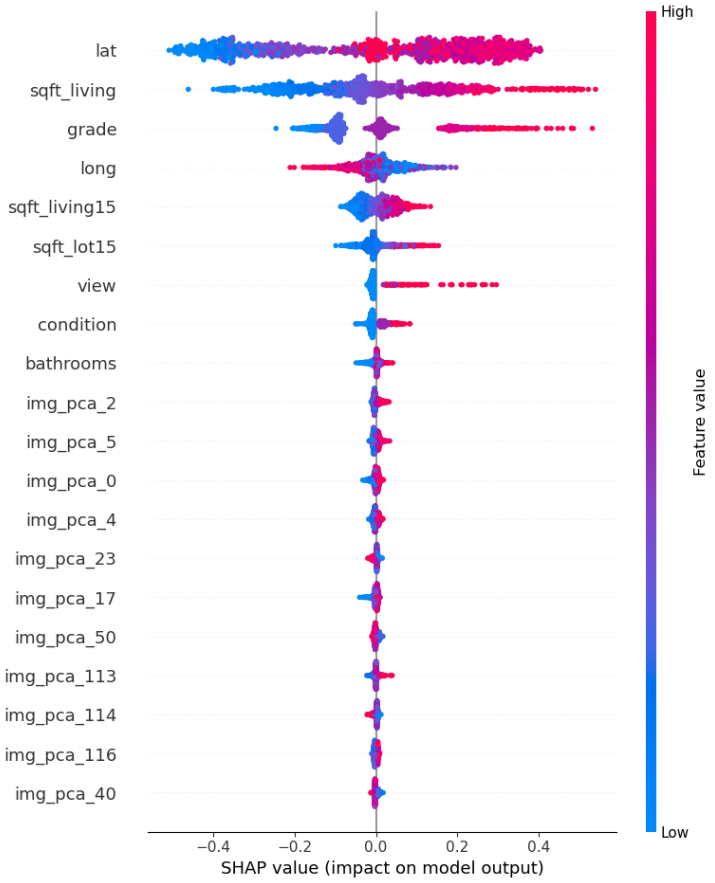

Figure 12: SHAP summary plot for the fusion model.

SHAP dependence plots for image-derived neighborhood features reveal non-linear effects. Built-up density shows positive contribution at low values and diminishing or negative impact at higher densities. Green cover exhibits an optimal range, where moderate vegetation coverage contributes positively while extreme values reduce predicted price. Road density similarly shows diminishing returns, with moderate connectivity being most beneficial.

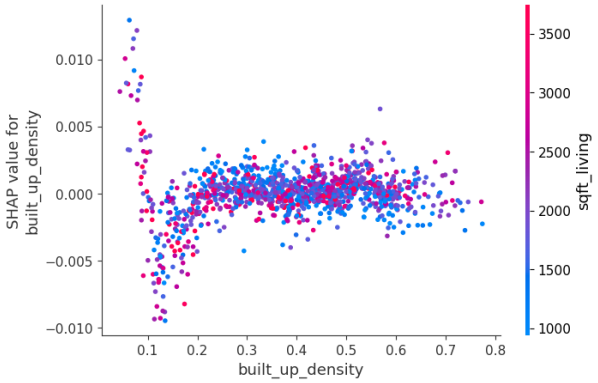

Figure 13: SHAP dependence plot for built-up density.

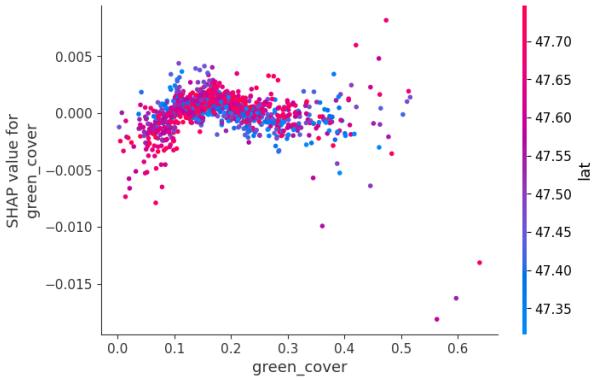

Figure 14: SHAP dependence plot for green cover.

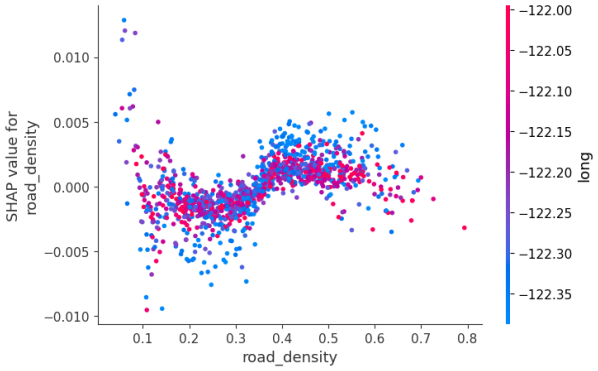

Figure 15: SHAP dependence plot for road density.

Grad-CAM Visual Interpretation

Grad-CAM was applied to the convolutional layers of the pretrained CNN used for image feature extraction. Grad-CAM highlights image regions that contribute most strongly to the extracted image embeddings.

For lower-priced properties, Grad-CAM attention is more diffuse and concentrated around dense road networks and surrounding structures. In contrast, higher-priced properties exhibit focused activation on building footprints, nearby greenery, and open spaces. This indicates that the CNN captures meaningful neighborhood-level visual patterns related to property valuation.

Grad-CAM is used solely for interpretability of visual feature extraction. Final price predictions are generated by the downstream XGBoost regressor.

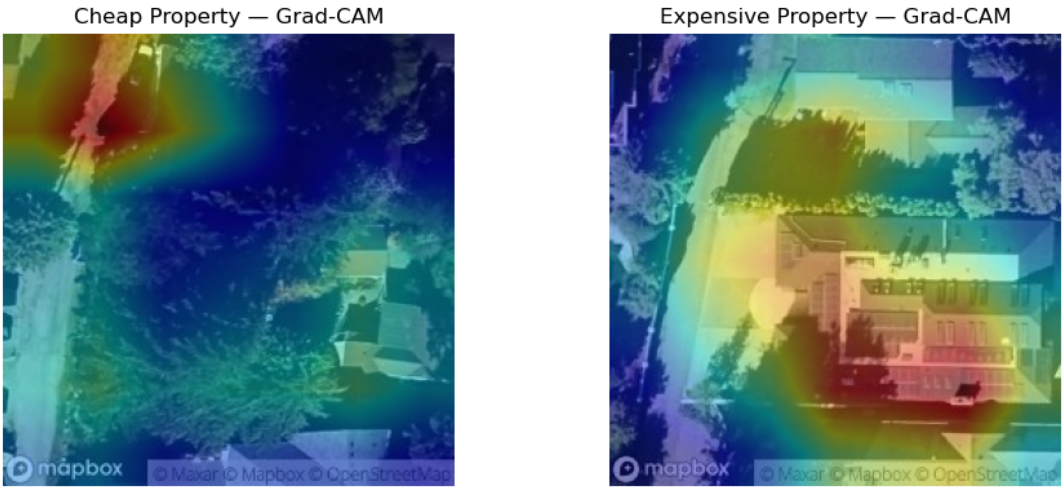

Figure 16: Grad-CAM visualizations for low-priced and high-priced properties.

8 Results and Model Comparison

8.1 Evaluation Metrics

Model performance was evaluated using standard regression metrics, including R^2 and Root Mean Squared Error (RMSE). All metrics were computed on log-transformed property prices. Higher R^2 and lower RMSE indicate better predictive performance.

8.2 Model Performance Comparison

Three modeling approaches were evaluated: a tabular-only baseline, an image-only model, and a satellite–tabular fusion model.
The tabular-only model achieved strong predictive performance, confirming that structured housing attributes contain substantial signal for property valuation. The image-only model showed weaker performance, indicating that satellite imagery alone is insufficient to accurately predict property prices.
The fusion model achieved the best performance among all approaches. Despite using a reduced set of tabular features, the fusion model outperformed the tabular-only baseline, demonstrating that satellite imagery provides complementary neighborhood- level information that improves valuation accuracy.

Table 2: Performance comparison of tabular, image-only, and multimodal fusion models.

Model Type	Data Modality	Features Used	Model	RMSE (log-price)	R^2
Tabular-Only	Structured housing data	Tabular features	XGBoost	0.1654	0.9009
Image-Only	Satellite imagery	ResNet-18_image embeddings (512-D)	XGBoost	0.4579	0.2318
Multimodal Fusion	Tabular+ Satellite imagery	Reduced_tabular features+ PCA-reduced_image embeddings	XGBoost	0.1864	0.8693

8.3 Discussion of Results

The results indicate that while traditional tabular features remain the dominant drivers of property valuation, satellite imagery contributes meaningful additional information when combined appropriately. Image-only models are unable to capture property value accurately in isolation; however, visual features significantly enhance predictions when fused with structured data.

Key Takeaway:
Satellite imagery does not replace tabular housing data, but it enhances property valuation models by capturing neighborhood-level and environmental context that is difficult to encode using structured features alone.

9 Financial and Economic Insights

- The strong performance of the tabular-only model indicates that traditional structural and locational attributes (e.g., `sqft_living`, `grade`, latitude and longitude) capture a large portion of property value and remain essential for automated valuation systems.
- The image-only model shows limited predictive power, suggesting that satellite imagery alone is insufficient for accurate price estimation and cannot replace structured housing data in financial valuation tasks.
- The multimodal fusion model achieves improved performance over the tabular baseline, demonstrating that satellite imagery provides complementary economic signal rather than redundant information.
- SHAP analysis shows that image-derived neighborhood features such as built-up density, green cover, and road density have measurable contributions to predicted prices, confirming that environmental context carries implicit economic value.

- Grad-CAM visualizations indicate that the model focuses on meaningful spatial patterns such as open spaces, vegetation, road connectivity, and building layout, aligning with known real estate valuation factors like accessibility and neighborhood quality.
- From a financial perspective, incorporating satellite imagery can reduce valuation error in areas where properties with similar internal characteristics differ significantly in surrounding environment.

10 Limitations

- Satellite imagery was collected for a stratified subset of properties rather than the full dataset due to API and computational constraints. As a result, the fusion model does not utilize imagery for all available records.
- The CNN backbone was used strictly as a frozen feature extractor. While this improves stability and reduces overfitting, it limits the model's ability to learn task-specific visual features tailored to property valuation.
- Image embeddings were reduced using PCA before fusion, which may discard some fine-grained visual information relevant to pricing.
- The image-only model demonstrates limited predictive performance, indicating that satellite imagery alone cannot capture important internal property attributes such as size, quality, and condition.
- The study relies on static satellite imagery and does not account for temporal changes in neighborhoods, infrastructure development, or seasonal variation.

11 Future Work

- End-to-end multimodal training could be explored by fine-tuning the CNN jointly with the regression model to learn valuation-specific visual features.
- Higher-resolution imagery or multi-scale image inputs could improve the representation of fine-grained neighborhood characteristics.
- Temporal satellite imagery could be used to study neighborhood evolution and its impact on long-term property price trends.
- More advanced fusion strategies, including attention-based or late-fusion architectures, could be evaluated and compared with the current feature-level fusion approach.

12 Conclusion

- This project demonstrated an end-to-end multimodal pipeline for residential property valuation by integrating structured housing data with satellite imagery.
- A strong tabular-only baseline using XGBoost achieved high predictive performance, confirming the importance of structural, quality, and location-based features in property price estimation.
- Satellite imagery, when used in isolation, showed limited predictive power, indicating that visual context alone is insufficient for accurate valuation.
- The multimodal fusion model outperformed the tabular baseline despite using a reduced set of tabular features, demonstrating that satellite imagery provides complementary neighborhood-level information.
- SHAP-based analysis confirmed that image-derived features such as built-up density, green cover, and road density contribute meaningfully to price predictions, while Grad-CAM visualizations validated that the CNN focuses on semantically relevant spatial regions.
- Overall, the results highlight the value of multimodal learning for building more context-aware and robust property valuation models without replacing existing tabular-based systems.