Name - Deepak Malik
Enrollment no- 23113045
Branch- Civil Engineering (BTech 3rd year)

Satellite Property Valuation: Enhancing Interpretability with Grad-CAM

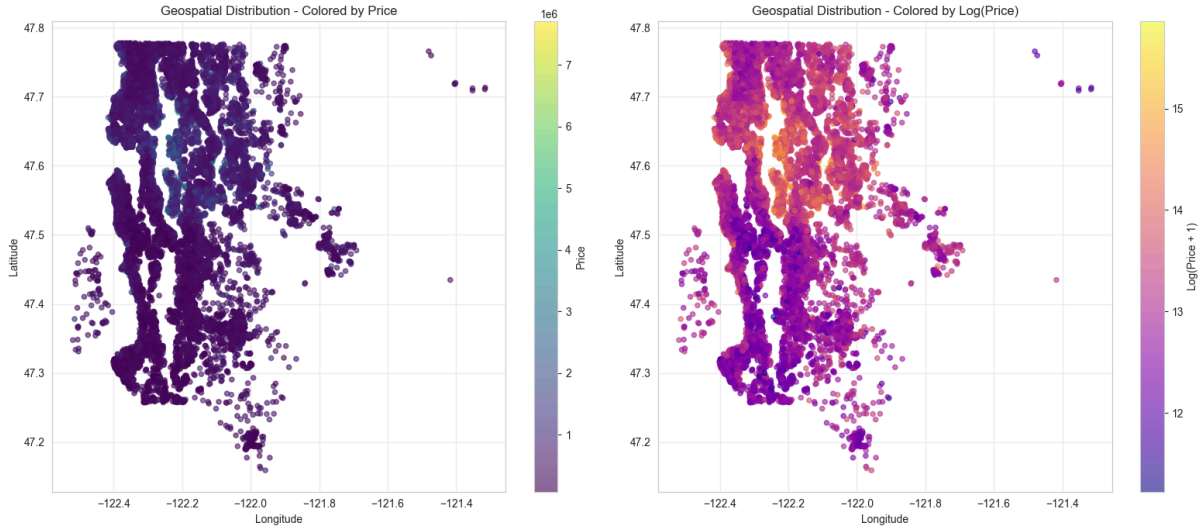
Abstract

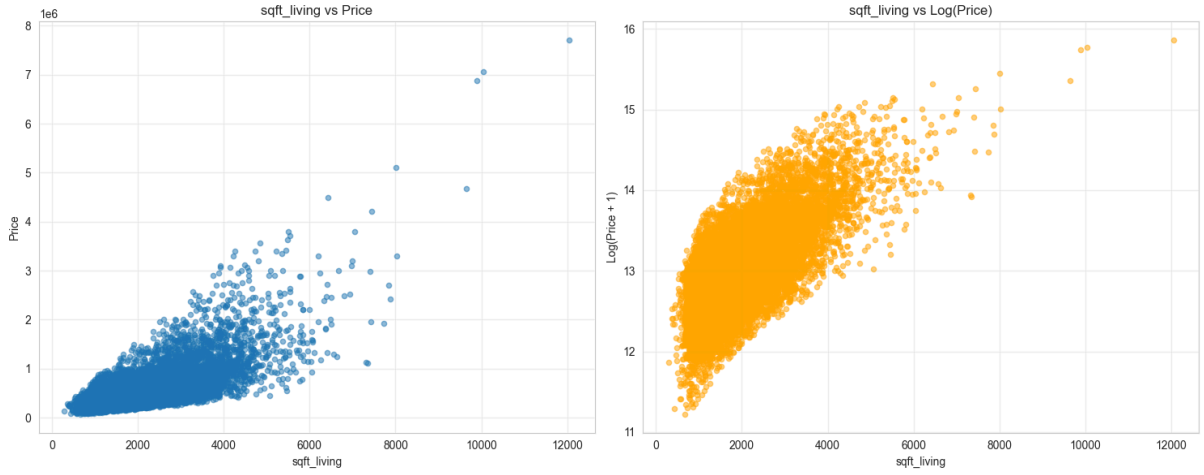
This report details a multimodal property valuation system that integrates structured housing data with satellite imagery, focusing on the interpretability of visual features using Gradient-weighted Class Activation Mapping (Grad-CAM). We demonstrate how Grad-CAM visualizes the spatial regions in satellite images that significantly influence the model's valuation predictions, thereby validating the role of environmental and neighborhood context. The findings confirm that satellite imagery provides complementary information to traditional tabular data, leading to improved prediction accuracy and a more robust, explainable valuation model.

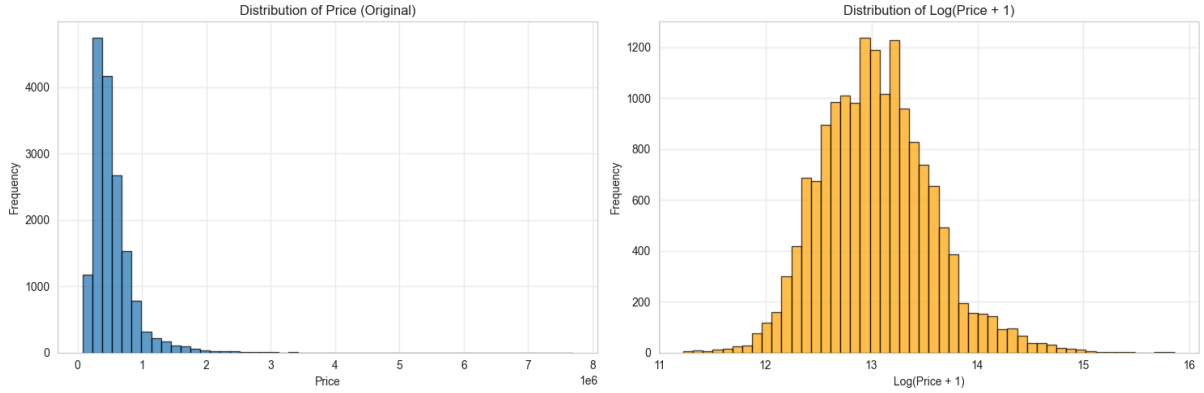
1. Introduction

Property valuation is a complex process traditionally reliant on structured tabular data such as property size, age, and location. However, the surrounding environment and neighborhood characteristics, often captured visually, play a crucial role in perceived property value. This project introduces a multimodal learning pipeline that combines tabular data with satellite imagery to enhance property valuation accuracy and, critically, its interpretability. By employing Grad-CAM, we aim to understand how the model leverages visual cues from satellite imagery to make its predictions, moving beyond black-box models to provide transparent, context-aware valuations.

EDA charts:







2. Methodology

2.1 Overall System Architecture

The system architecture is designed as a modular multimodal learning pipeline. Tabular data and satellite imagery are processed independently before their fusion at the modeling stage.

- **Tabular Data Processing:** Tabular features are standardized and fed into gradient-boosted decision trees (XGBoost) to establish a strong baseline prediction.
- **Satellite Imagery Processing:** Satellite images, acquired using latitude and longitude coordinates, are processed through a pretrained ResNet-50 Convolutional Neural Network (CNN) to extract high-dimensional visual embeddings. These embeddings capture neighborhood-level visual context, including road density, green cover, and urban structure.
- **Multimodal Fusion:** Instead of directly replacing tabular models, image-based signals are integrated using a residual learning strategy. A multimodal neural network learns residual errors from the tabular baseline, allowing visual information to complement rather than overpower strong numeric predictors. This residual multimodal architecture significantly improves prediction accuracy by incorporating satellite imagery while preserving a strong tabular baseline.

2.2 Grad-CAM for Model Explainability

To interpret how satellite imagery influences property valuation, we apply Gradient-weighted Class Activation Mapping (Grad-CAM) on the convolutional neural network used for visual feature extraction. Grad-CAM enables the visualization of spatial regions within an image that contribute most strongly to the model's internal representations. This is particularly useful in satellite imagery, where environmental and neighborhood context—rather than the property itself—plays a critical role in valuation.

We generate Grad-CAM heatmaps using the final convolutional layer of the pretrained ResNet-50 model. For a given satellite image, Grad-CAM computes gradients of the network's activations and highlights regions most influential in the learned representation. The resulting heatmaps are then overlaid on the original satellite images to visually identify areas that the model attends to during prediction.

3. Observations and Results

3.1 Grad-CAM Visualizations

Analysis of the Grad-CAM visualizations reveals several consistent patterns:

- **Road Networks:** High activation is observed along major roads and intersections, indicating that accessibility and connectivity significantly influence perceived property value.
- **Neighborhood Density:** Dense residential layouts with uniform housing patterns show stronger attention compared to irregular or sparse regions, suggesting the model captures neighborhood structure and planning quality.
- **Green Cover:** Areas with visible trees, parks, or vegetation exhibit elevated activation, supporting the hypothesis that environmental aesthetics and greenery contribute positively to valuation.

- **Large Structures and Open Spaces:** In some cases, the model focuses on large buildings or open spaces nearby, potentially correlating with commercial proximity or recreational access.

Importantly, the model does not focus on individual house rooftops, which aligns with expectations given the satellite resolution and confirms that the CNN is learning contextual neighborhood features rather than property-specific details.

Figure 1: Grad-CAM Visualization — Dense Residential Block*The model places strong emphasis on compact residential clusters and adjacent road networks. High activation over dense housing patterns indicates sensitivity to neighborhood density and urban planning consistency, which are known to influence property valuation.*

Figure 2: Grad-CAM Visualization — Road Accessibility and Mixed Land Use*Grad-CAM highlights major road corridors and nearby built-up regions, suggesting that accessibility and connectivity play a significant role in the model's valuation logic. Properties located near well-connected transport routes are implicitly favored.*

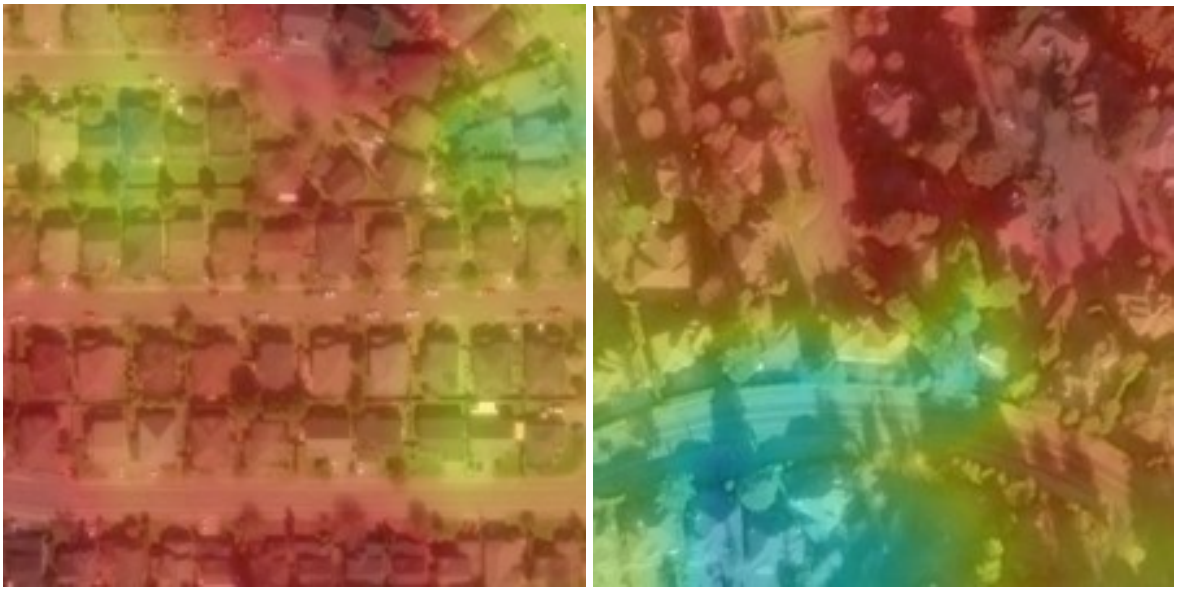


Figure 3: Grad-CAM Visualization — Prominent Structure within Green Surroundings*The model focuses on a large central structure surrounded by vegetation, indicating attention to both built infrastructure and green cover. This suggests the CNN captures a balance between structural prominence and environmental quality.*

Figure 4: Grad-CAM Visualization — Curved Road Network and Urban Flow*High activation follows a curved roadway and surrounding residential areas, demonstrating the model's ability to capture road geometry and traffic flow patterns. Such layouts often correlate with neighborhood accessibility and desirability.*

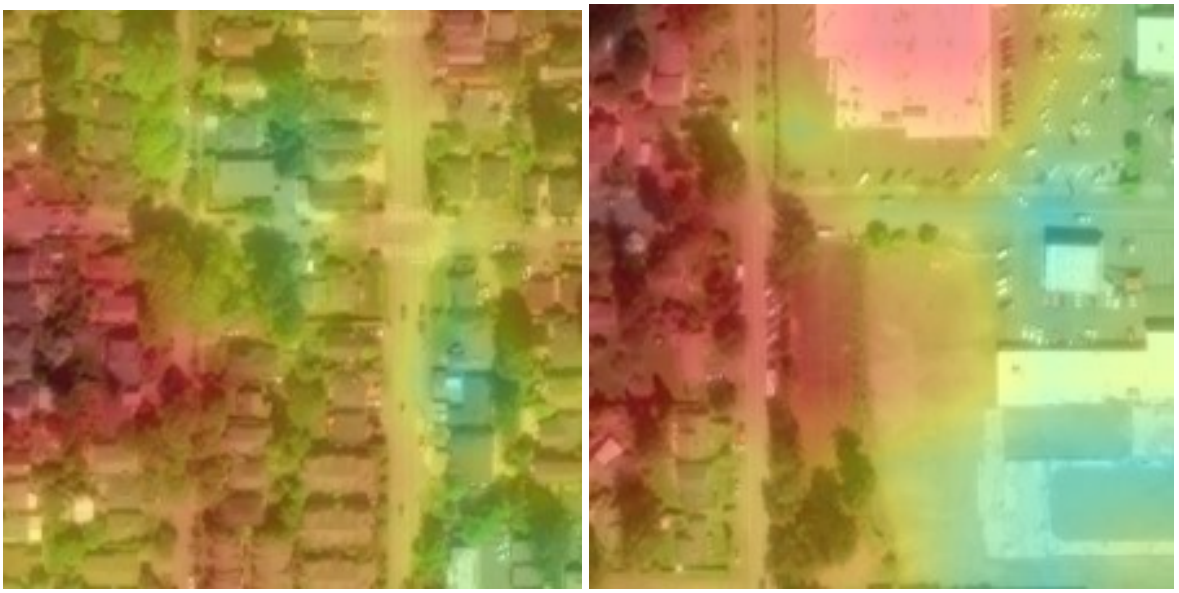


Figure 5: Grad-CAM Visualization — Suburban Grid with Green Patches*The model emphasizes a structured suburban grid interspersed with green spaces. This pattern reflects sensitivity to planned residential layouts and environmental aesthetics, both of which contribute positively to property value.*

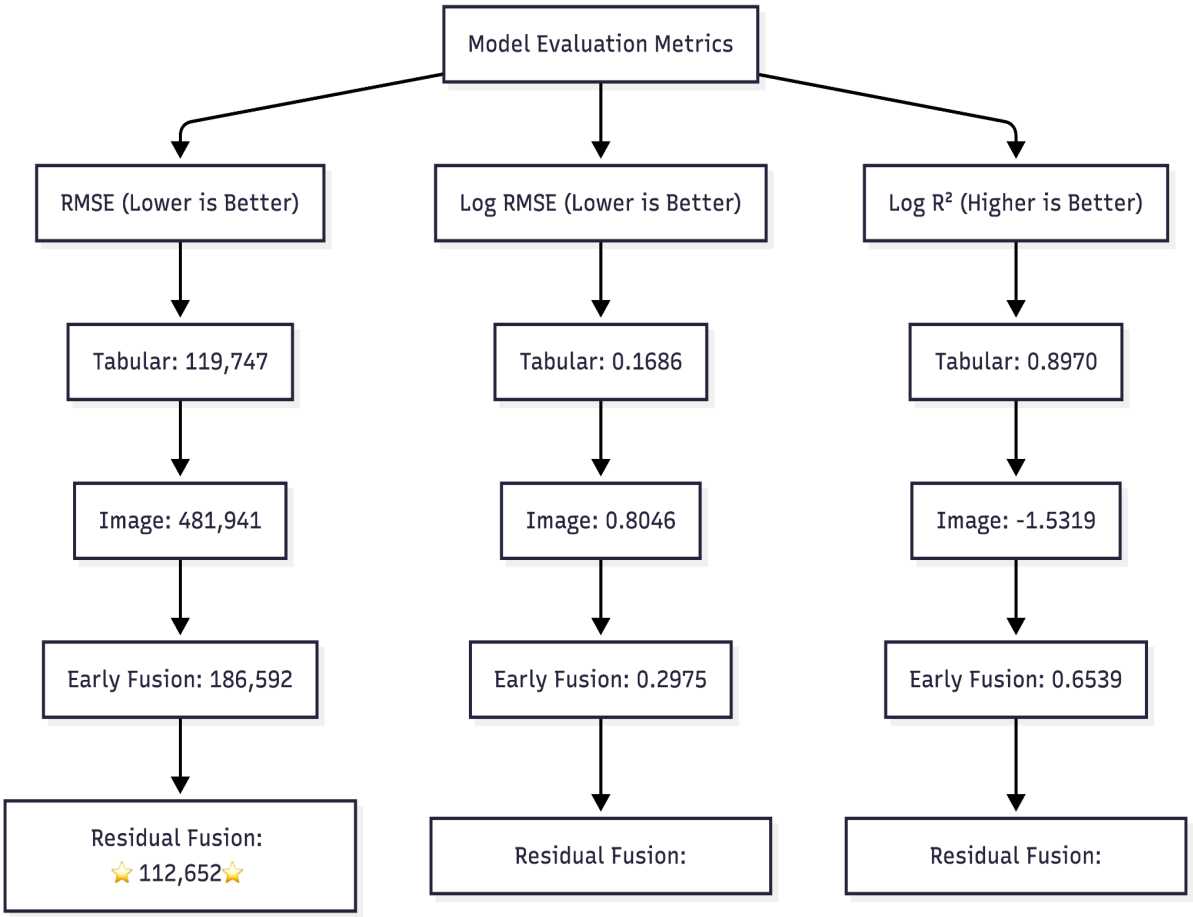
Figure 6: Grad-CAM Visualization — Uniform Housing Layout*Activation is distributed across a uniform housing block, indicating that the model captures consistency in neighborhood design rather than individual buildings.*

3.2 Model Comparison

Three modeling approaches were evaluated:

1. **Tabular-only regression:** Using XGBoost.
2. **Image-only regression:** Using CNN embeddings.
3. **Multimodal fusion:** Using residual learning.

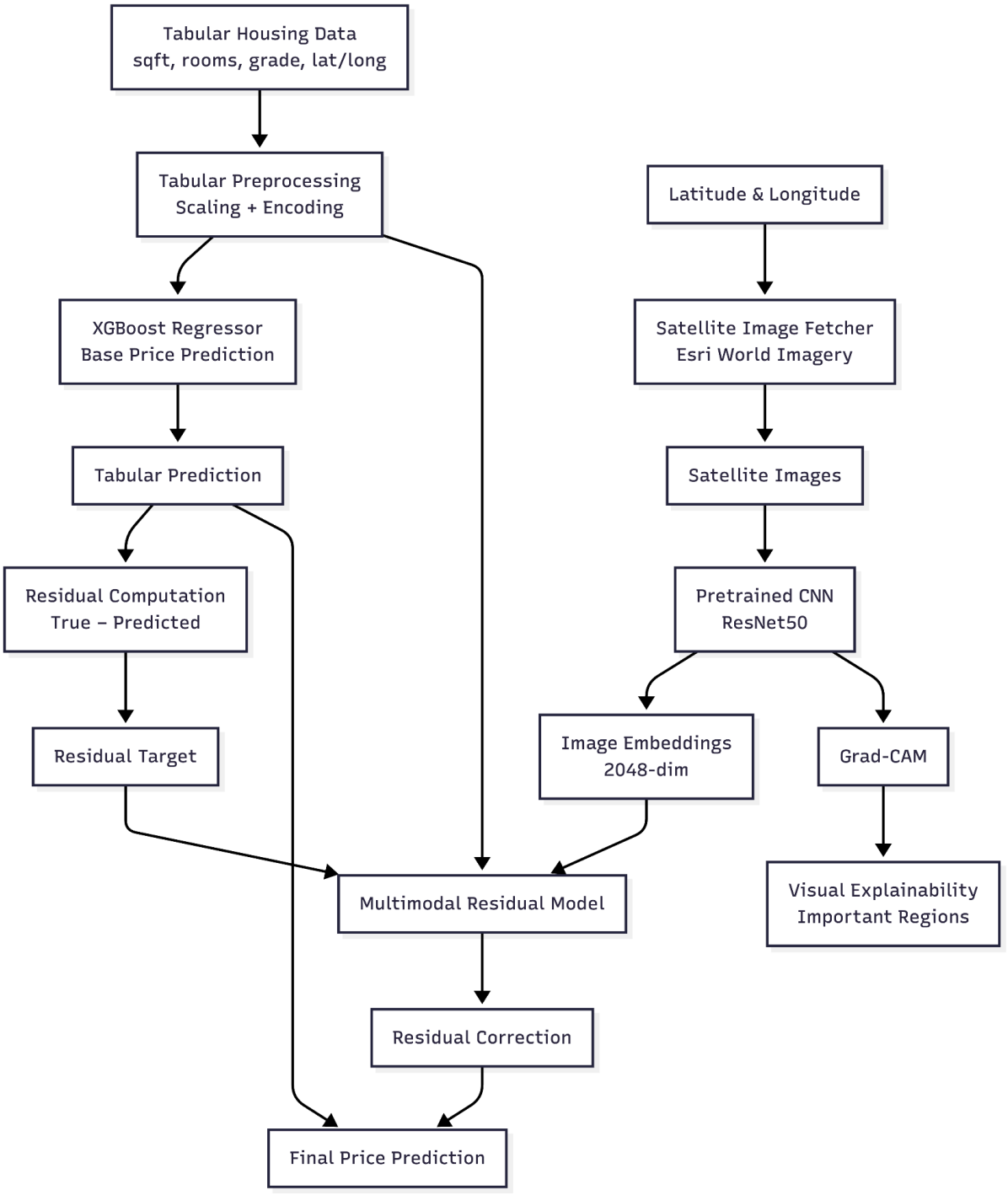
The image-only model performed poorly, confirming that satellite imagery alone is insufficient for accurate valuation. However, when integrated as a residual signal on top of the tabular baseline, the multimodal model achieved a noticeable improvement. This demonstrates that visual context provides complementary information not captured by structured features alone.



Performance comparison of tabular, image only, and multimodal models on the validation set.
The residual multimodal architecture significantly improves prediction accuracy by incorporating satellite imagery while preserving a strong tabular baseline.

4. Implications

These visual explanations validate the role of satellite imagery as a complementary signal to tabular data. While traditional features capture intrinsic property attributes (e.g., size, condition, grade), satellite imagery contributes extrinsic environmental context, such as accessibility, greenery, and urban layout. This explainability analysis provides confidence that the multimodal model's predictions are grounded in interpretable and economically meaningful visual patterns rather than spurious correlations.



Multimodal architecture for satellite imagery–based property valuation.

A strong tabular XGBoost model predicts base prices, while a CNN-based multimodal residual model learns image-driven corrections. Grad-CAM provides visual explainability by highlighting influential spatial regions.

5. Limitations and Future Work

Several limitations remain in the current implementation:

- **Resolution:** Satellite imagery resolution restricts the model to neighborhood-level interpretation rather than property-specific details.
- **Data Availability:** Only a subset of available images was used due to API and bandwidth constraints.

Future work could include:

- Using higher-resolution imagery or multiple zoom levels.
- Incorporating temporal satellite data.
- Exploring attention-based fusion architectures.

6. Conclusion

This project demonstrates a complete, end-to-end multimodal valuation system that integrates structured housing data with satellite imagery. By combining strong tabular baselines, deep visual embeddings, and explainability tools such as Grad-CAM, the pipeline moves beyond traditional regression models and provides interpretable, context-aware property valuations. The methodology is robust, extensible, and suitable for real-world deployment in real estate analytics.