Satellite Imagery Based Property Valuation using Multimodal Regression

Name: Ashi Agrawal
Enrollment Number: 24118012
Date: 5 Jan 2026

---

1. Introduction

The objective of this project is to improve property price prediction by incorporating satellite imagery along with conventional tabular features. Satellite images provide a visual representation of the neighborhood and help capture spatial and environmental context. By combining image-based features extracted using a convolutional neural network (CNN) with tabular housing data, a multimodal regression framework is developed and evaluated.

---

2. Dataset Description

The dataset consists of residential property records containing structured attributes such as number of bedrooms, bathrooms, living area, lot size, construction grade, and geographical coordinates (latitude and longitude). The target variable for prediction is the market price of the property.

Satellite images corresponding to each property location were programmatically acquired using the latitude and longitude values. These images capture neighborhood-level characteristics that are not directly available in the tabular dataset.

---

3. Data Preprocessing and Feature Engineering

Data preprocessing involved removing duplicate entries and dropping non-essential columns such as identifiers and timestamps that do not contribute to prediction. The cleaned dataset was then used for feature engineering.

Several derived features were created to enhance the predictive power of the model, including:

- Property age calculated from the year of construction
- Basement area ratio relative to total living area
- Difference between property living area and neighborhood average
- A binary indicator for high-grade construction

These engineered features help incorporate domain knowledge and improve model interpretability.

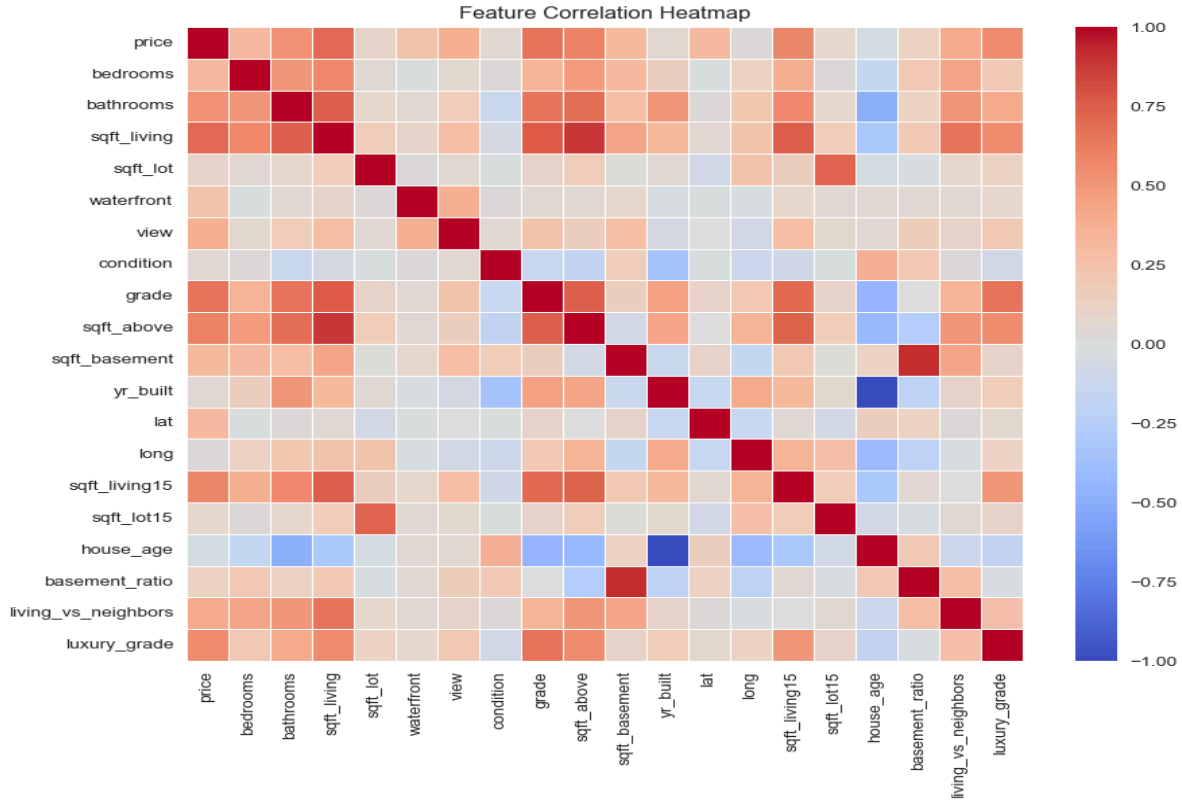
---

4. Exploratory Data Analysis (EDA)

Exploratory data analysis was performed to understand relationships between features and the target variable. Distribution plots, scatter plots, and correlation heatmaps were used to analyze trends and dependencies in the data.

[FIGURE 1: Price distribution]

[FIGURE 2: Living area vs price]



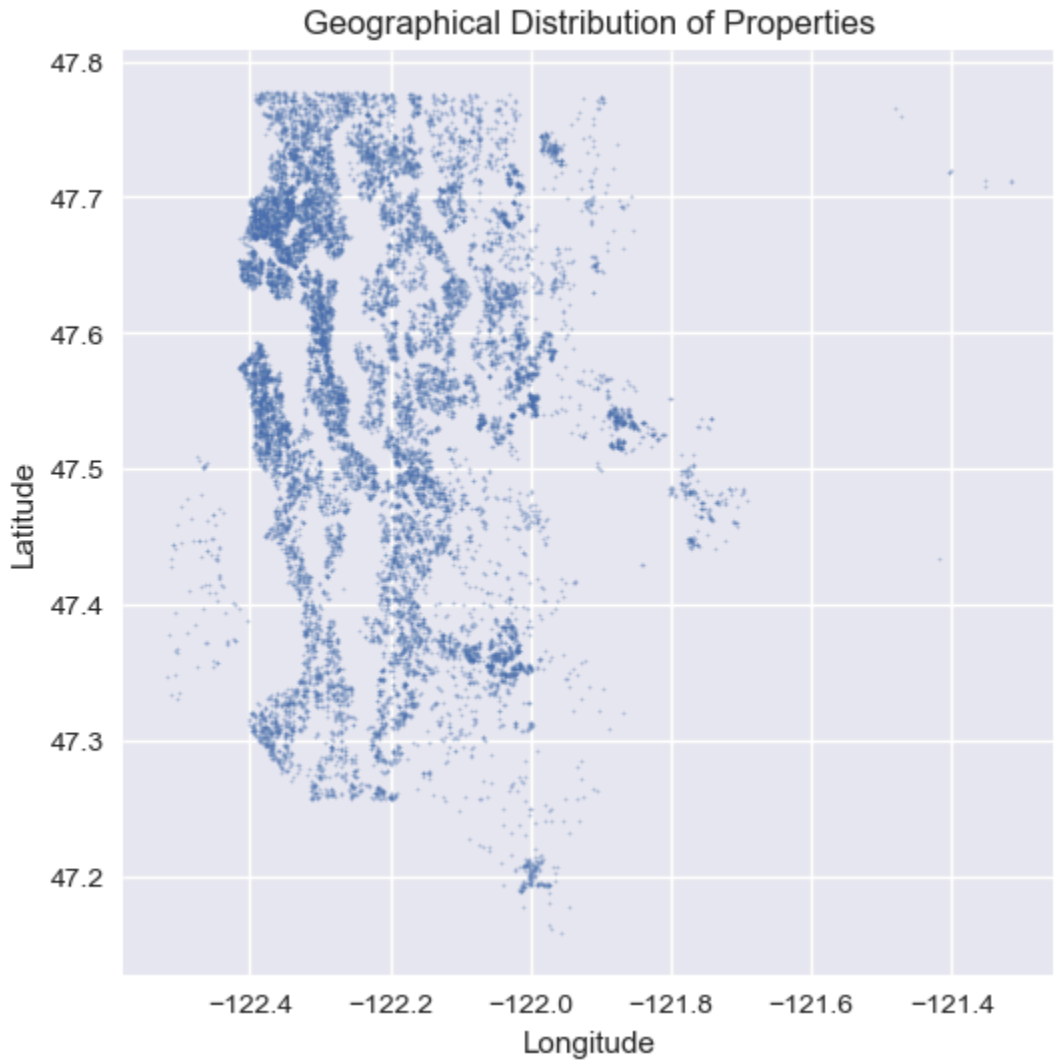
[FIGURE 3: Correlation heatmap]

The analysis showed strong relationships between property price and features such as living area, construction grade, and location-related attributes.

---

5. Geospatial Analysis

To study spatial patterns, a geospatial visualization was created using latitude and longitude values of the properties. This visualization highlights clustering of properties in urban regions and variations in density across locations.



The spatial distribution supports the motivation for using satellite imagery, as neighborhood context varies significantly across different regions.

---

6. Satellite Image Acquisition

Satellite images were programmatically fetched using a static maps API based on the geographical coordinates of each property. A fixed zoom level was used to ensure consistent neighborhood-scale context across all images. Image downloading was fully automated through a Python script to ensure reproducibility.

---

7. CNN-Based Image Feature Extraction

A pretrained ResNet-18 convolutional neural network was used as a feature extractor. The final classification layer was removed, and each satellite image was passed through the network to obtain a 512-dimensional embedding vector.

These embeddings act as compact numerical representations of visual neighborhood features such as road connectivity, surrounding infrastructure, and green cover.

---

8. Multimodal Regression Model

8.1 Tabular-Only Baseline

As a baseline, a Random Forest Regressor was trained using only the structured tabular features. This model establishes a reference performance level without any visual information.

8.2 Multimodal Model

To build the multimodal model, image embeddings were concatenated with tabular features. The combined feature vector was then used to train the same regression model, allowing a fair comparison between tabular-only and multimodal approaches.

---

9. Model Performance Comparison

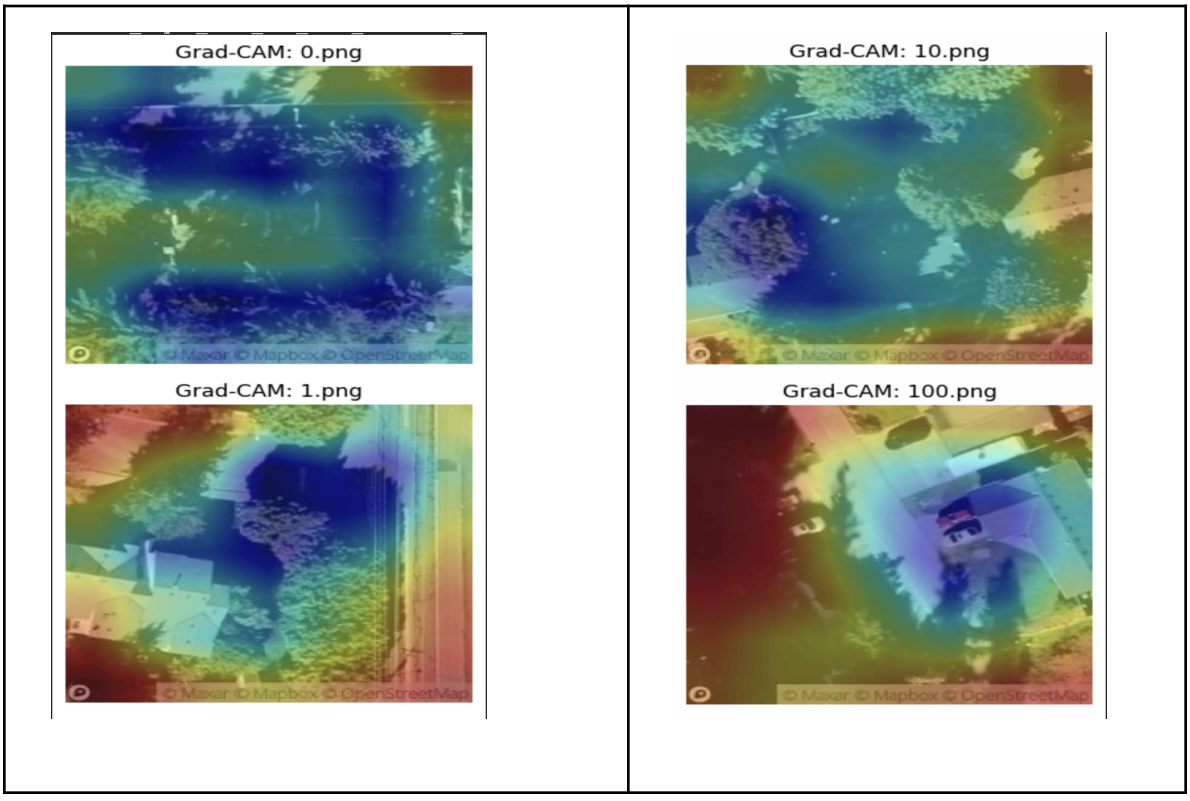
Model performance was evaluated using Root Mean Squared Error (RMSE) and R^2 score on a validation set.

Model	RMSE	R^2
Tabular Only	130984.8027	0.863
Tabular + Satellite Images	148720.2242	0.824

The tabular-only model achieved better quantitative performance than the multimodal model on the validation set. This is likely because structured features such as living area, construction grade, and location already capture strong predictive signals. In contrast, the satellite image embeddings introduce high-dimensional visual features that may contain noise or redundant information not directly aligned with the price prediction task. Additionally, the CNN feature extractor was pretrained on a generic image dataset and not fine-tuned for real estate valuation, which may limit the usefulness of the extracted visual features. These factors together explain the reduced R^2 observed in the multimodal setting.

---

10. Explainability Using Grad-CAM



Grad-CAM was applied to the CNN feature extractor to visualize salient regions in satellite imagery. The heatmaps show attention to road networks, built-up areas, and green spaces, indicating that neighborhood context is captured by the visual features.

---

11. Conclusion

This project investigated a multimodal approach to property valuation by combining structured housing data with satellite imagery. While the tabular-only model achieved stronger validation performance, the multimodal framework demonstrates how neighborhood-level visual context can be integrated into traditional valuation pipelines. The results highlight that incorporating image-based features requires careful feature alignment and task-specific tuning to be effective. Overall, the project provides a practical and balanced view of the potential and challenges of multimodal learning for real estate analytics.

---

12. Future Work

Future improvements could include the use of higher-resolution satellite imagery, temporal satellite data, or end-to-end deep learning models that jointly learn visual and tabular representations.