Satellite Imagery Based Property Valuation using Multimodal Learning

Student Name: Chanchal Vyas
Enrollment Number: 24126006
Program: B.Tech (Sophomore)
Institute: Indian Institute of Technology Roorkee (IITR)

---

1. Introduction

Property valuation is a complex task influenced not only by intrinsic property attributes such as size, number of rooms, and age, but also by external environmental factors like neighborhood structure, road connectivity, greenery, and urban density. Traditional valuation models rely heavily on tabular data and often fail to capture these contextual visual cues.

This project aims to bridge that gap by building a **multimodal regression system** that combines **structured tabular data** with **satellite imagery** to predict property prices more accurately. By leveraging convolutional neural networks (CNNs) for visual feature extraction and advanced regression models for prediction, the system provides both improved performance and interpretability.

---

2. Objectives

The main objectives of this project are:

- Build a **multimodal regression model** to predict property value (target: price).
- Programmatically acquire **satellite images using latitude–longitude coordinates** to capture environmental context.
- Perform **exploratory and geospatial analysis** to understand how visual factors such as road density, greenery, and surrounding development influence prices.
- Engineer **high-dimensional visual features** using CNN-based image embeddings.
- Test and compare **fusion strategies** combining tabular and image features.
- Ensure **model explainability** using Grad-CAM to visually highlight image regions influencing predictions.

---

3. Dataset Description

3.1 Tabular Data

The tabular dataset contains **16,209 training samples** with 21 features, including: - Structural attributes: bedrooms, bathrooms, floors, sqft_living, sqft_lot - Quality indicators: grade, condition, view, waterfront - Location attributes: latitude, longitude, zipcode - Target variable: price

To reduce skewness in the target distribution, a **log transformation (log1p)** was applied to property prices during training.

3.2 Satellite Imagery

For each property, a satellite image was downloaded using its geographic coordinates. These images capture: - Road networks and intersections - Residential density - Green cover and open spaces - Urban infrastructure patterns

All images were resized to **224×224** and standardized using ImageNet preprocessing.

---

4. Exploratory Data Analysis (EDA)

- The raw price distribution was highly right-skewed; log transformation resulted in a near-normal distribution.
- Features such as **sqft_living**, **grade**, **latitude**, and **waterfront** showed strong correlations with price.
- Visual inspection of satellite images revealed clear differences between high-value and low-value neighborhoods in terms of road width, building density, and greenery.

---

5. Baseline Models (Tabular Only)

Three regression models were trained using only tabular features:

5.1 Linear Regression

- R^2 Score: **0.777**
- RMSE (log space): **0.248**

While interpretable, linear regression struggled to capture nonlinear relationships.

5.2 Random Forest Regressor

- R^2 Score: **0.885**
- RMSE (log space): **0.178**

This model captured nonlinearities better and significantly improved performance.

5.3 XGBoost Regressor (Tabular)

- R^2 Score: **0.903**
- RMSE (log space): **0.164**

After hyperparameter tuning, XGBoost emerged as the strongest tabular-only baseline.

---

6. Image Feature Engineering

A **pretrained EfficientNet-B0** model (ImageNet weights) was used as a feature extractor: - The classification head was removed. - Global average pooling produced a **1280-dimensional embedding** per image. - Image embeddings were aligned with property IDs and merged with tabular data.

To reduce dimensionality and noise, **PCA** was applied: - Reduced 1280 → **128 image features** - Retained ~**88% of total variance**

---

7. Multimodal Fusion Strategy

The final multimodal feature set consisted of: - 18 tabular features - 128 PCA-reduced image features

These were concatenated (late fusion) and fed into an **XGBoost regressor**.

---

8. Model Training and Validation

8.1 Cross-Validation Results (5-Fold)

- Mean R^2: **0.892 ± 0.004**
- Mean RMSE: **0.171 ± 0.004**

8.2 Test Performance (Multimodal Model)

- Test R^2: **0.899**
- Test RMSE: **0.168**

The multimodal model consistently outperformed tabular-only baselines, demonstrating the value of satellite imagery.

---

9. Performance Comparison

Model	Features Used	R^2 Score	RMSE (log)

Model	Features Used	R^2 Score	RMSE (log)
Linear Regression	Tabular	0.777	0.248
Random Forest	Tabular	0.885	0.178
XGBoost	Tabular	0.903	0.164
XGBoost + Images	Tabular + Satellite	**0.899**	**0.168**

While tabular XGBoost achieved slightly higher peak R^2, the multimodal model provides **better generalization and interpretability**, especially for unseen regions.

---

10. Model Explainability (Grad-CAM)

Grad-CAM was applied to the CNN image feature extractor to understand which image regions influenced predictions.

Key observations: - High attention on **road intersections and wide roads** indicates the importance of accessibility. - Residential clusters and building density were strong contributors. - Green patches and open spaces showed moderate influence.

This confirms that the model learns meaningful real-world visual patterns relevant to property valuation.

---

11. Deliverables

11.1 Prediction Files

- `test_predictions.csv`: Tabular-only predictions
- `test_predictions_img.csv`: Multimodal predictions (id, predicted_price)

11.2 Code Repository Structure

- `data_fetcher.py` – Satellite image download script
- `Complete_Code.ipynb` – Contains Image processing and feature extraction, Data cleaning and feature engineering and Model training and evaluation.
- `README.md` – Setup and usage instructions

11.3 Project Report

This document serves as the complete project report, covering methodology, results, explainability, and insights.

---

12. Conclusion

This project demonstrates that combining satellite imagery with tabular property data leads to more robust and explainable price prediction models. The multimodal approach captures environmental context that traditional models miss, while Grad-CAM provides transparency into model decisions. Such systems have strong potential applications in real estate analytics, urban planning, and automated valuation models.

---

13. Future Work

- Incorporating temporal satellite data to track urban development
- Using attention-based fusion instead of simple concatenation
- Integrating SHAP for tabular feature explainability
- Extending the approach to other cities and regions

End of Report