SATELLITE IMAGERY BASED PROPERTY VALUATION

Eshika Suresh Katekhaye

24114036

CSE (2nd yr)

IIT Roorkee

1. Overview

Accurate property valuation plays a central role in real estate analytics, influencing decisions related to buying, selling, investment, and taxation. Traditionally, property prices are estimated using structured attributes such as property size, number of rooms, construction quality, and geographic location. While these factors capture key property - level characteristics, they often provide only a partial view of the broader neighborhood environment in which a property is located.

With recent advances in computer vision and the growing availability of geospatial data, there has been increasing interest in incorporating visual information, particularly satellite imagery, into property valuation models. Satellite images offer an overhead perspective of neighborhoods, capturing elements such as road connectivity, surrounding infrastructure, building density, green cover, and proximity to water bodies. These environmental features are intuitively linked to perceived neighborhood quality and may influence market value in ways that are not always explicitly represented in tabular datasets.

Figure 1 - Example of satellite image illustrating residential property surroundings, including road access, nearby buildings, and green cover.

This project examines whether such a visual context can meaningfully enhance property price prediction when combined with structured housing data. A **baseline - first modeling strategy** is adopted, in which structured attributes are first modeled independently using classical regression techniques to establish a strong predictive benchmark. Satellite imagery is then incorporated through **CNN - based feature extraction** and a **late fusion approach**, enabling a systematic evaluation of whether visual neighborhood information provides complementary predictive value.

The primary objective of this study is to assess the effectiveness of multimodal learning for property valuation while maintaining transparency and interpretability. By directly comparing tabular only and multimodal models under consistent experimental conditions, the project aims to offer practical insights into when visual data can enhance predictive performance and when it may add limited value despite its intuitive appeal.

2. Dataset Description and Exploratory Data Analysis (EDA)

2.1 Tabular Data

The structured data used in this project comes from the **King County House Sales Dataset**, which contains historical housing transactions from King County, USA. The prediction target is the final sale price of each property.

The dataset includes a diverse set of features, such as:

- Number of bedrooms and bathrooms
- Living area and lot size
- Construction grade and condition
- View quality and waterfront access
- Geographic coordinates (latitude and longitude)

Together, these features capture both **intrinsic property characteristics** and **location-based value drivers**, making the dataset well-suited for regression-based valuation tasks.

2.2 Exploratory Data Analysis

Initial exploration of the dataset reveals a right-skewed price distribution, indicating the presence of high-value outliers. Most properties fall within a moderate price range, while a smaller fraction represents premium listings.

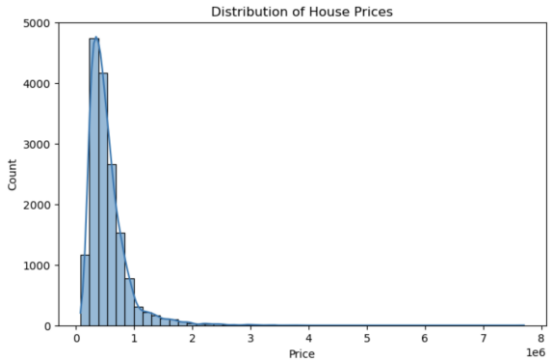

Figure 2.2.1 - Price distribution histogram

Satellite images were programmatically retrieved using open-access satellite imagery services based on property latitude and longitude coordinates. These images capture overhead views of neighborhoods, providing contextual information beyond individual property characteristics.

(a) *(b)* *(c)*

Figure 2.2.2 - Satellite images from different residential neighborhoods illustrating variations in urban density, road structure, and green cover. (These examples demonstrate the type of neighborhood-level visual context captured by satellite imagery in this study.)
(a) Low-density residential area with high green cover and sparse built structures.
(b) Moderately dense residential neighborhood with planned housing layouts and balanced vegetation.
(c) Higher-density residential neighborhood with closely spaced buildings and limited green cover.

Satellite images were programmatically retrieved using open-access satellite imagery services based on property latitude and longitude coordinates. These images capture overhead views of neighborhoods, providing contextual information beyond individual property characteristics.

3. Methodology

This section describes the modeling pipeline used to predict property prices using structured housing attributes and satellite imagery. The approach follows a baseline-first strategy, where tabular models are first evaluated independently, followed by multimodal extensions incorporating visual features.

3.1 Tabular Regression Models

Let x_i denote the structured feature vector for the i-th property and y_i its corresponding sale price. The objective is to learn a function $f(\cdot)$ such that: $y_i = f(x_i)$

Three regression models were evaluated using tabular features: Linear Regression, Random Forest Regressor (RF) , and Gradient Boosting Regressor (GBR). Linear Regression models price as a weighted sum of features, but its assumption of linearity limits its ability to capture complex interactions.

Tree-based models overcome this limitation by modeling non-linear relationships. Random Forest aggregates predictions from multiple decision trees, while Gradient Boosting builds trees sequentially to iteratively reduce prediction error. Due to its strong performance on structured data, **Gradient Boosting Regressor is selected as the primary baseline model**.

3.2 Satellite Image Feature Extraction

Satellite images corresponding to each property were programmatically retrieved using geographic coordinates (latitude and longitude) from open-access satellite imagery services. Since raw images cannot be directly used in regression models, a pretrained Convolutional Neural Network (CNN) was employed for feature extraction.

Let I_i denote the satellite image of property i. The CNN maps the image to a fixed length embedding:

$$z_i = g(I_i)$$

where $z_i \in R\ 512$ represents visual patterns such as vegetation coverage, road layout, and building density. The CNN is used solely as a feature extractor and is not fine-tuned during training.

3.3 Multimodal Fusion Strategy

To evaluate whether satellite imagery provides complementary information beyond structured features, a **late fusion approach** was adopted. Let x_i represent the tabular features and z_i the corresponding image embedding. The fused feature vector is defined as:

$$u_i = [x_i \ // \ z_i]$$

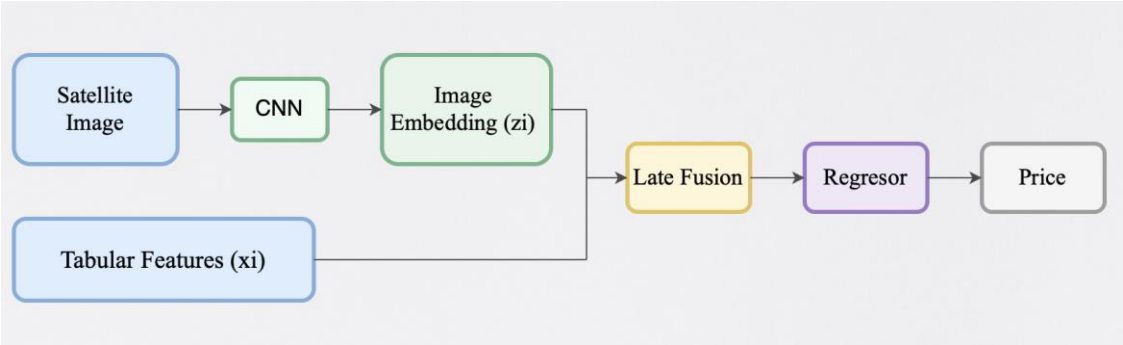

Figure 3.3 - Multimodal architecture illustrating late fusion of structured housing attributes with CNN-extracted satellite image embeddings for property price prediction.

This concatenated representation is then used as input to a regression model for price prediction. Late fusion allows the model to independently weigh structured and visual information without imposing assumptions about their interactions at the representation level.

Two tree-based multimodal models were evaluated: Random Forest with image embeddings (RF + CNN) and Gradient Boosting with image embeddings (GBR + CNN). In addition, a linear multimodal configuration using PCA - reduced image features with Ridge Regression were explored to assess whether dimensionality reduction and regularization could mitigate overfitting.

Comparative performance analysis of tabular-only and multimodal models is presented in Section 5.

4. Experimental Setup

The dataset was split into training and validation subsets using a fixed random seed to ensure reproducibility. Model performance was evaluated using:

- **R^2 (Coefficient of Determination)** - primary metric, as it is scale-invariant
- **RMSE (Root Mean Squared Error)** - reported for completeness, but interpreted cautiously due to scale sensitivity

5. Results

5.1 Tabular-Only Model Performance

Structured housing attributes were first evaluated independently to establish a strong baseline.

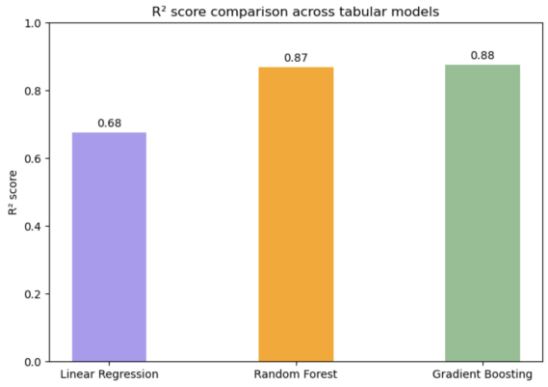

Figure 5.1.1 - R^2 score comparison across tabular models

Linear Regression achieves a moderate R^2 score, indicating limited ability to capture non-linear relationships in the data. In contrast, tree-based models significantly outperform the linear baseline. Random Forest exhibits strong explanatory power, while Gradient Boosting Regressor achieves the highest R^2 score, explaining approximately 88% of the variance in property prices.

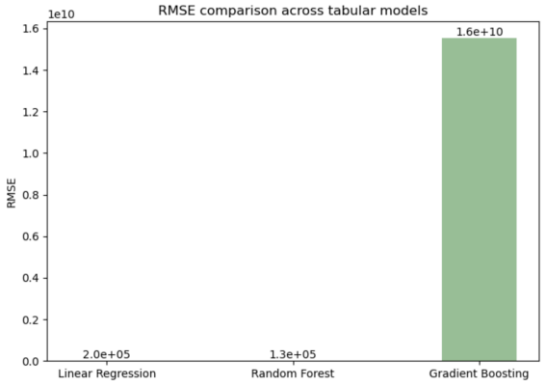

Figure 5.1.2 - RMSE comparison across tabular models

Although RMSE values vary substantially across models, these differences are influenced by target scaling and transformation effects. Since RMSE is scale-dependent, R^2 is used as the primary criterion for model selection. Based on this, **Gradient Boosting Regressor is selected as the strongest tabular baseline**.

5.2 Multimodal Model Performance (Tabular + Image embeddings)

To assess the contribution of satellite imagery, multimodal models were evaluated using CNN-based image embeddings combined with tabular features.

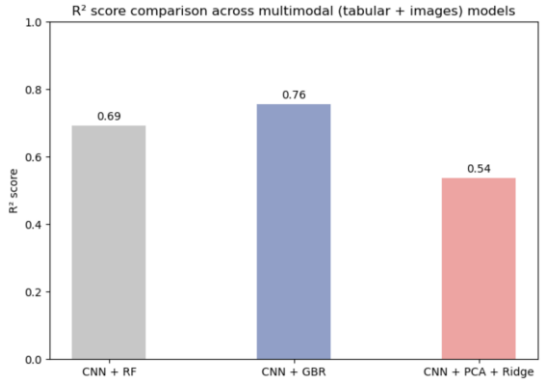

Figure 5.2.1 - R^2 score comparison across multimodal models

Among the multimodal configurations, the **CNN + Gradient Boosting Regressor** achieves the highest R^2 score, outperforming CNN + Random Forest and CNN + PCA + Ridge. However, its performance remains inferior to the tabular-only Gradient Boosting model reported in Section 5.1.

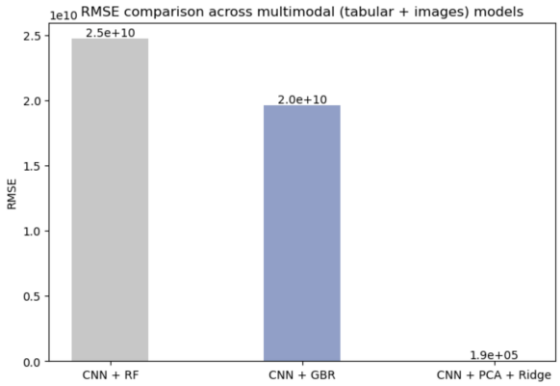

Figure 5.2.2 - RMSE comparison across multimodal models

RMSE values exhibit large magnitude differences due to scaling effects and inverse transformations applied during evaluation. As a result, RMSE comparisons across multimodal models are not directly reliable, and R^2 remains the principal metric for assessment.

Overall, the inclusion of satellite imagery does not lead to improved predictive performance compared to using structured data alone.

6. Financial and Visual Insights from Satellite Imagery

The following insights summarize how visual neighborhood characteristics, as observed from satellite imagery, relate to property valuation in the context of this study:

- **Green cover and open spaces:**
  Areas with visible vegetation, parks, and open land are generally associated with higher perceived neighborhood quality. Such environments are often linked to suburban or low-density residential zones, which tend to command higher property values.

- **Building density:**
  Regions with dense building clusters and limited open space typically reflect highly urbanized settings. While these areas may offer accessibility and infrastructure advantages, high density alone does not consistently translate to higher prices in the dataset.

- **Road networks and connectivity:**
  Satellite imagery highlights road layouts and transportation networks, which provide indirect

cues about accessibility. However, the financial impact of these visual patterns appears to be largely captured by geographic coordinates and neighborhood-level tabular features.

- **Visual homogeneity vs. variability:**
  Neighborhoods with uniform housing patterns exhibit less visual variability, while mixed-use or transitioning areas show heterogeneous structures. These visual differences are intuitive to human observers but do not strongly influence model predictions.

- **Limited incremental financial value:**
  Despite their interpretability, visual features extracted from satellite imagery contribute limited additional predictive value when combined with strong structured attributes such as living area, construction quality, and neighborhood averages.

Overall, these observations suggest that while satellite imagery provides meaningful contextual understanding of neighborhoods, its direct contribution to price prediction remains marginal in the presence of rich tabular data.

7. Visual Explainability Using Grad-CAM

Interpretability is an important consideration when working with visual data. Gradient-weighted Class Activation Mapping (Grad-CAM) is a commonly used technique for understanding how convolutional neural networks attend in different regions of an image.

In this study, satellite images were processed using a pretrained CNN solely for feature extraction, and regression models were trained independently on the extracted embeddings. As the model was not trained end-to-end, Grad-CAM visualizations were not directly generated. Nevertheless, Grad-CAM remains a useful conceptual tool for interpreting CNN-based representations and understanding the type of visual cues such models can capture.

These insights are consistent with the overall findings of this project: while satellite imagery provides interpretable contextual information, it does not significantly enhance predictive performance when strong structured housing features are already present.

8. Limitations and Future Work

A key limitation of this study is the relatively low spatial resolution of satellite imagery, which restricts the capture of fine-grained architectural details. Higher-resolution imagery or street-level views may provide richer visual information more closely aligned with property valuation.

Additionally, the CNN used in this project was pretrained on generic image datasets and not fine-tuned for real estate-specific imagery. Future work could explore end-to-end multimodal training or attention-based fusion techniques to better integrate visual and structured data.

9. Conclusion

This project investigated a multimodal regression framework for property price prediction by integrating structured housing attributes with satellite imagery. Experimental results demonstrate that **tabular data alone, when modeled using Gradient Boosting Regressor, provides strong predictive performance**.

While satellite imagery offers interpretable neighborhood-level context, its inclusion does not lead to improved prediction of accuracy in the presence of rich structured features. These findings highlight an important principle in applied machine learning: **additional data modalities do not guarantee better performance unless they provide complementary and predictive information**.

References

1. Harlfoxem. *House Sales Prediction Dataset*. Kaggle. https://www.kaggle.com/datasets/harlfoxem/housesalesprediction
2. Friedman, J. H. (2001). *Greedy function approximation: A gradient boosting machine*. Annals of Statistics, 29(5), 1189–1232.
3. Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
5. Selvaraju, R. R., Cogswell, M., Das, A., et al. (2017). *Grad-CAM: Visual explanations from deep networks via gradient-based localization*. Proceedings of the IEEE International Conference on Computer Vision (ICCV).
6. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.