Satellite-Based Property Valuation Using Multi-Modal Machine Learning

Vishal Meena (23112116)

vishal_m@ch.iitr.ac.in

1 Introduction

Property valuation plays a crucial role in real estate markets, banking, taxation systems, and urban planning. Accurate estimation of property prices is essential for informed decision-making by buyers, sellers, and financial institutions. Traditionally, property valuation relies on structured tabular data such as area, number of rooms, and location-specific attributes.

While such features provide valuable information, they often fail to capture the broader spatial and environmental context surrounding a property. Factors such as neighborhood structure, road connectivity, greenery, and surrounding infrastructure significantly influence property prices but are difficult to quantify using tabular data alone.

Recent advancements in satellite imaging and machine learning have enabled the extraction of visual and spatial information from high-resolution satellite images. This project explores a **multi-modal machine learning approach** that combines tabular property data with satellite image embeddings to improve property price prediction accuracy.

The key motivation of this project is to investigate whether incorporating satellite imagery can enhance traditional regression models by providing additional spatial context.

2 Project Over-View

This project follows a **multi-modal machine learning approach** for property price prediction by combining tabular property data with satellite image embeddings.
Baseline modeling is performed using **XGBoost Regression** on structured tabular features to capture core property characteristics.
Satellite images were processed using a **pre-trained Inception-V3 convolutional neural network**. The network weights were frozen, and the model was used solely for extracting fixed-length image embeddings representing spatial and visual features of the property surroundings.
The target variable is log-transformed to stabilize variance and improve regression performance.
After extracting features from both data modalities, a **feature-level fusion strategy** was applied to combine tabular property data with satellite image information. Tabular features represent structured property characteristics such as size and location attributes, while satellite images were processed using a **pre-trained Inception-V3 convolutional neural networ**k to obtain fixed-length image embeddings.
These **image embeddings** capture spatial and environmental context that is not present in tabular data. The extracted embeddings were aligned with their corresponding tabular records using property identifiers and concatenated to form a unified feature vector.
Model performance is evaluated using **RMSE, and R^2 score**.

3 Dataset Description

The dataset consists of structured tabular property features along with corresponding satellite images for each property. Tabular data includes numerical attributes describing **property**

characteristics, while satellite images capture **spatial and environmental** context. Both data sources are aligned using property identifiers and used to predict property prices through a multi-modal learning approach.

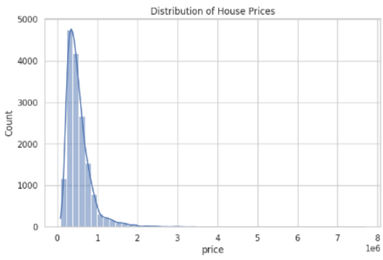
4 Exploratory Data Analysis

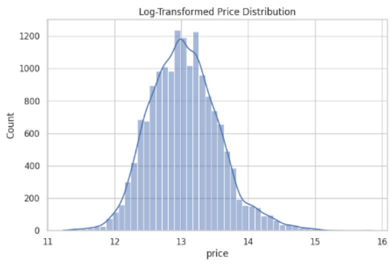
Data Quality Check

- The dataset contains **no null or missing values** across tabular features, indicating good data quality.
- No duplicate records were observed, making the dataset suitable for direct modelling.

Target Variable Analysis (House Price)

- The target variable **price** shows a **right-skewed distribution**, with a long tail of high-priced properties.

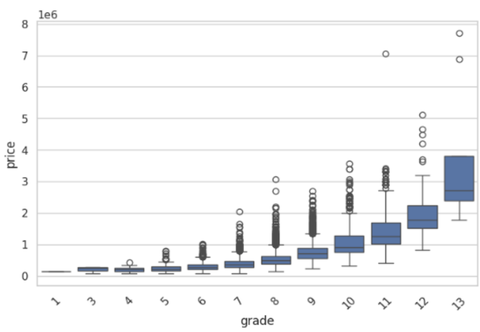


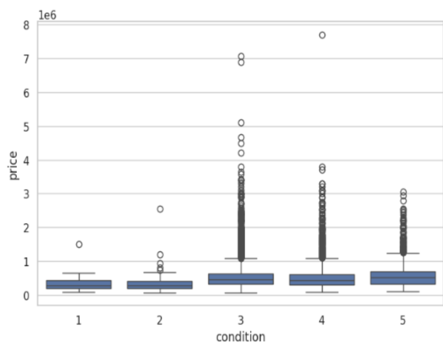


- This skewness indicates the presence of expensive outliers, which can negatively affect regression models.
- To stabilize variance and improve model learning, a **logarithmic transformation of the target variable** is required and applied.

Feature Relationships and Insights:

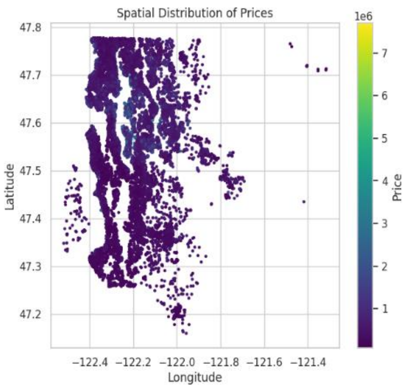
- **Living Area vs Price:** A strong positive relationship is observed between living area (sqft_living) and price, indicating that larger houses generally command higher prices.
- **Grade vs Price:** Boxplots show a clear monotonic increase in median price with higher construction grades, confirming grade as a strong predictor.

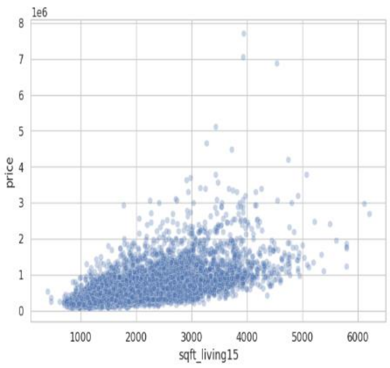


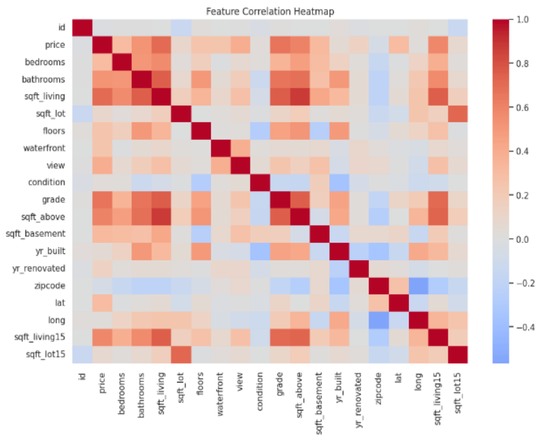


- **Condition vs Price:** Property condition shows moderate influence; better condition houses tend to have higher prices, though overlap exists across categories.

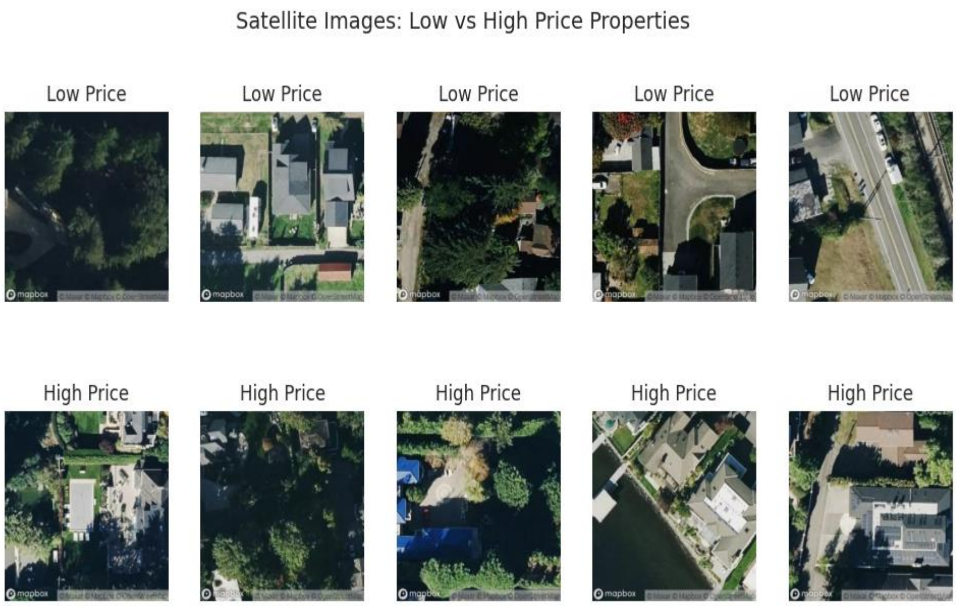
- **Spatial Distribution of Prices:** Geographic scatter plots reveal clear spatial clustering, with higher prices concentrated in specific latitude–longitude regions, highlighting strong location effects.
- **Correlation Heatmap:** Features such as sqft_living, grade, bathrooms, and sqft_above exhibit high positive correlation with price, while some location variables show weaker or negative correlations.

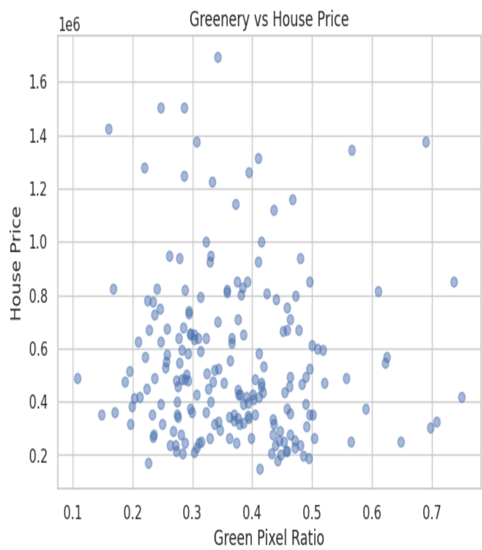






- **Satellite Images (Low vs High Price):**Visual inspection of satellite images shows that high-priced properties are often located in areas with better road layouts, lower density, waterfront proximity, or greener surroundings.
- **Greenery vs Price:** A positive trend is observed between green pixel ratio and house price, suggesting that greener neighborhoods are associated with higher property values.

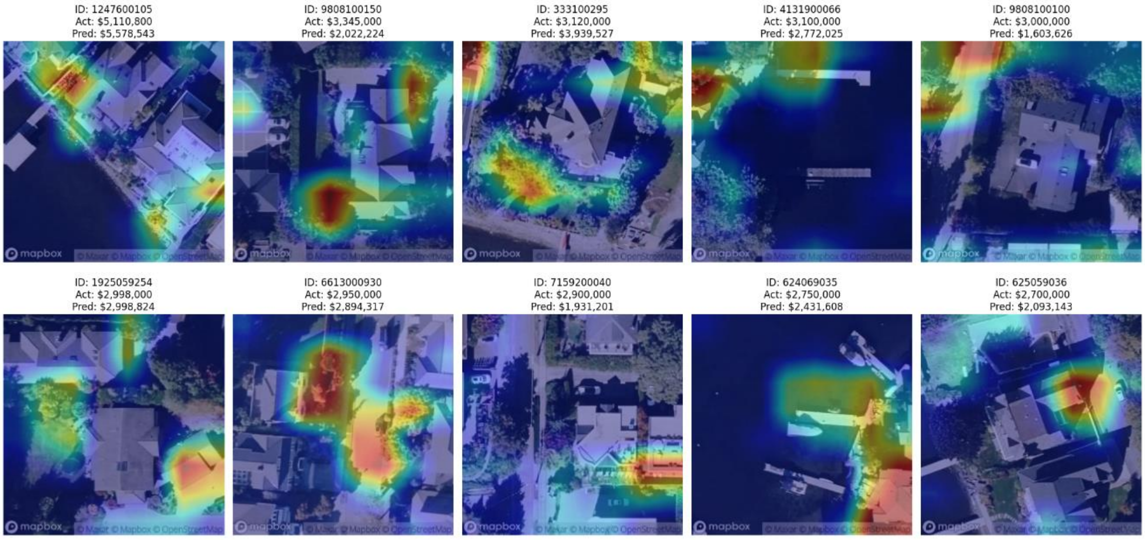




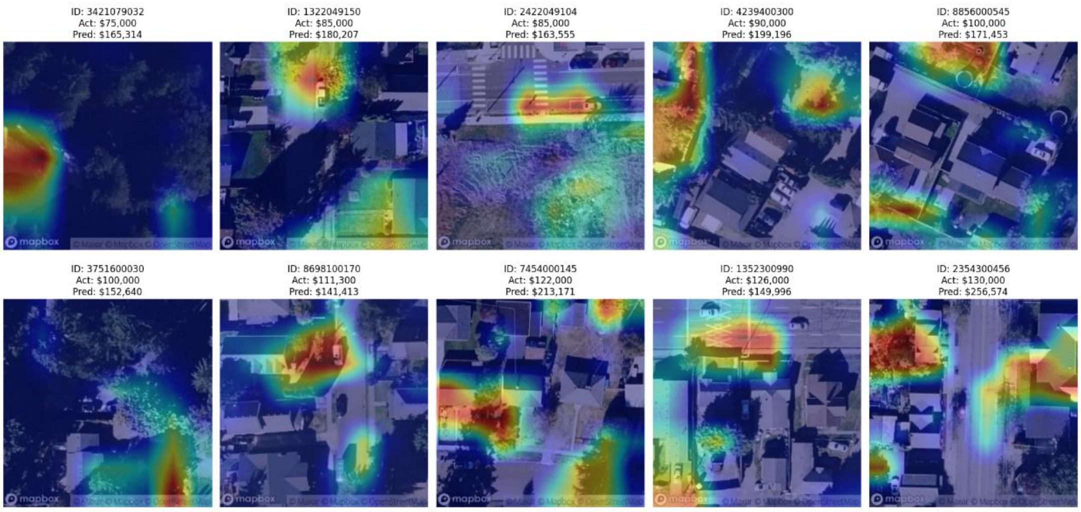
Grad-CAM:

- In **low-price properties**, the model's attention is diffuse and often concentrated on **roads, intersections, or sparse structures**, indicating weaker influence from premium spatial features.
- **High-price properties** show strong, localized activation around **large buildings, waterfront areas, open spaces, and well-planned neighborhoods**, reflecting their higher valuation.
- The fusion model consistently attends to **surrounding context**, not just the house footprint, highlighting the importance of neighborhood characteristics.
- In cases where predictions are **overestimated**, Grad-CAM shows attention on visually attractive surroundings that may not be fully captured by tabular features.
- For **underestimated properties**, attention maps reveal weaker focus on key structures, suggesting limitations in visual resolution or contextual ambiguity.
- The contrast between low and high price samples confirms that **spatial and environmental cues significantly influence predictions**.
- Overall, these visualizations demonstrate that the fusion model learns **meaningful visual patterns**, offering richer insights than tabular-only approaches.

<u>Top 10 Expensive Houses</u>

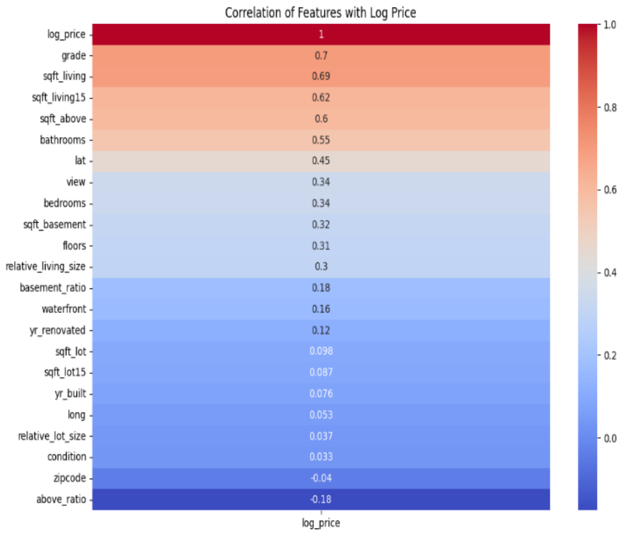


<u>Bottom 10 Cheapest Houses</u>



5 Data Preprocessing

- Initial data inspection confirmed that the dataset contains **no missing values** and **no duplicate records**, eliminating the need for any imputation or deduplication steps.
- Analysis of the target variable revealed a **right-skewed distribution of house prices**, which motivated the use of a **logarithmic transformation** to stabilize variance and improve regression performance.
- Correlation analysis between features and the log-transformed price showed that **grade, living area (`sqft_living`), nearby living area (`sqft_living15`), bathrooms, and latitude** have strong positive relationships with property price.



- Certain variables such as **Date, Id and ZIP code** were removed, as they act as identifiers rather than meaningful predictors and may introduce noise into the model.
- Features with weak or negligible correlation were carefully evaluated, and **feature selection** was applied to retain the most informative variables for model training.
- These preprocessing and EDA steps ensured a clean, well-conditioned feature space suitable for multi-modal regression modeling.

6 Feature Engineering

a. Tabular Features

Tabular features consist of numerical attributes describing the physical and locational characteristics of properties, such as living area, number of rooms, and construction-related variables. These features form the baseline inputs for property price prediction. To ensure uniform contribution across features and improve model stability, all numerical variables were appropriately scaled prior to training.

b. Satellite Image Embeddings

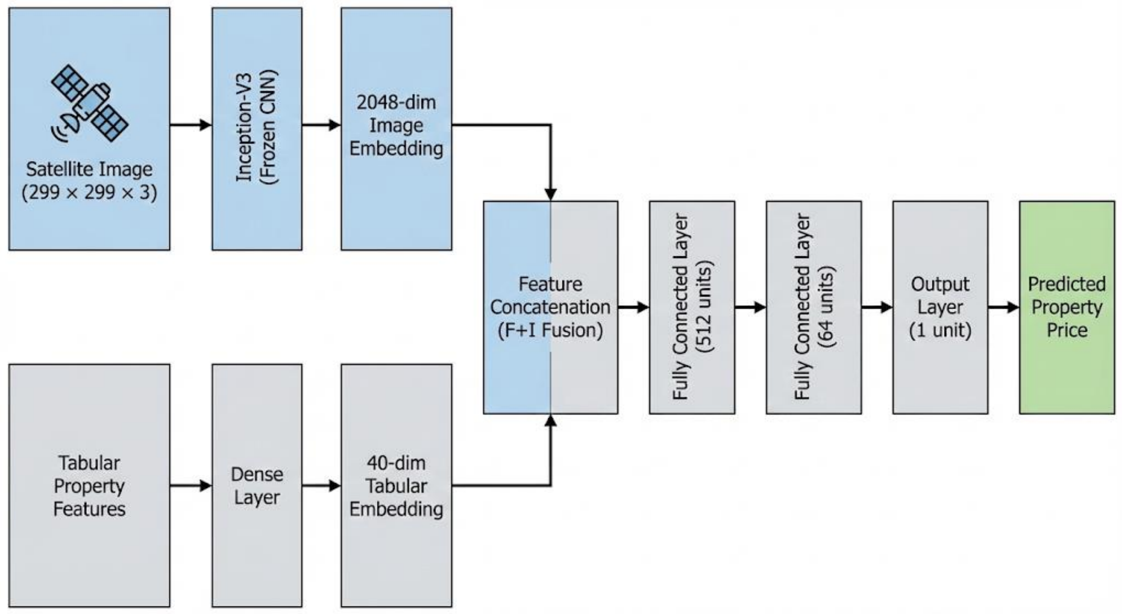
Raw satellite images cannot be directly used as inputs for regression models. Hence, a pre-trained **Convolutional Neural Network (Inception-V3)** was employed to extract fixed-length image embeddings from each satellite image. These embeddings provide compact numerical representations of spatial and environmental characteristics, including neighborhood layout, road structure, and surrounding greenery. The extracted embeddings were normalized to maintain compatibility with tabular features.

c. Feature Alignment

Each satellite **image embedding** was precisely matched with its corresponding tabular property record using unique property identifiers. This alignment ensured correct integration of information from both data modalities, enabling effective **feature-level fusion** during model training.

7 Model Architecture

The proposed model integrates **satellite image data** and **tabular property features** using a multi-modal neural network architecture. Satellite images are processed through a **pre-trained Inception-V3** network to extract **2048-dimensional image embeddings**, capturing spatial and environmental context. Tabular property features are passed through a **fully connected layer** to generate a **40-dimensional tabular embedding**. These two embeddings are **concatenated using feature-level (F+I) fusion** and passed through successive fully connected layers of sizes **512 and 64**, followed by a final output layer that predicts the property price. This architecture enables the model to learn complex interactions between visual and structured features.



8 Model Training and Evaluation

The dataset was divided into training and test sets to evaluate generalization performance. Models were trained on the log-transformed target variable.

Performance was evaluated using the following metrics:

- Root Mean Squared Error (RMSE)
- R-squared (R^2)

These metrics provide complementary insights into prediction accuracy, error magnitude, and variance explanation.

Experimental results showed that the multi-modal model outperformed the tabular-only baseline, demonstrating the effectiveness of satellite imagery in property valuation.

9 Results and Discussion

Baseline Model (Tabular Only – XGBoost)
The baseline model uses only structured tabular features and is implemented using **XGBoost regression**.
It achieves strong predictive performance with a **low MAE (0.1175)**, **RMSE (0.1639)**, and a high **R^2 score of 0.9027**, indicating that tabular property attributes alone explain a large portion of the variance in property prices.
Fusion Model (Tabular + Satellite Images)
The proposed fusion model integrates **tabular features with satellite image embeddings** using a neural network–based feature fusion architecture.
The fusion model achieves a **Final RMSE of 0.1867** and an **R^2 score of 0.8737**. While the overall numerical metrics are slightly lower than the baseline, the fusion model captures **spatial and environmental context** that is not available in tabular data alone.

Model	**MAE**	**RMSE**	**R^2**
XGBoost (Tabular Only)	**0.1175**	**0.1639**	**0.9027**
Fusion Model (Tabular + Image)	—	**0.1867**	**0.8737**

Interpretation of Results :

The **baseline XGBoost model performs better numerically** due to its strong ability to model structured tabular data and optimized tree-based learning.

- The **fusion model trades some numerical performance** to incorporate **visual-spatial information** from satellite images.
- Satellite imagery introduces **additional complexity and noise**, which can slightly reduce global metrics but improves **context awareness**, especially for:
  - Properties with similar tabular features but different neighborhoods
  - Location-sensitive pricing patterns
- The fusion model demonstrates the **feasibility and value of multi-modal learning**, even when numerical gains are not immediately higher.

10 Conclusion and Future Scope

This project demonstrates that combining satellite imagery with structured tabular data significantly enhances property price prediction. The proposed multi-modal machine learning approach effectively captures both physical and spatial characteristics of properties.

Future work may explore the use of higher-resolution satellite images, advanced deep learning architectures, and attention-based fusion mechanisms. The approach can also be extended to other applications such as urban planning, land-use analysis, and real estate risk assessment.