PROJECT REPORT

CDC x YHILLS

Satellite Imagery-Based Property Valuation

SUBMITTED BY:
Kalpesh Jagtap
23113078
B.Tech 3rd year
Civil Engineering

OVERVIEW

The objective of this project is to predict house prices by leveraging both structured tabular data and unstructured satellite image data. The core idea is to evaluate whether visual neighbourhood context can improve prediction performance beyond traditional tabular-only models.

The modelling strategy follows a two-stage approach. First, several strong tabular only regression models were developed to establish a reliable baseline. These included Random Forest, Ridge Regression, XGBoost, and LightGBM. Among these, LightGBM demonstrated the best performance due to its ability to capture complex nonlinear relationships and handle feature interactions efficiently. This model was selected as the primary tabular baseline.

In parallel, a multimodal deep learning framework was designed to integrate satellite images with tabular features. Tabular inputs were processed through a fully connected neural network, while satellite images were processed using a pretrained CNN (EfficientNetV2) acting as a feature extractor. The learned representations from both branches were concatenated and passed through dense layers to produce the final prediction.

To ensure stable optimization, the target variable was scaled during training and inverse transformed during evaluation. Performance was assessed using RMSE and R^2 scores to enable fair comparison between tabular-only and multimodal models.

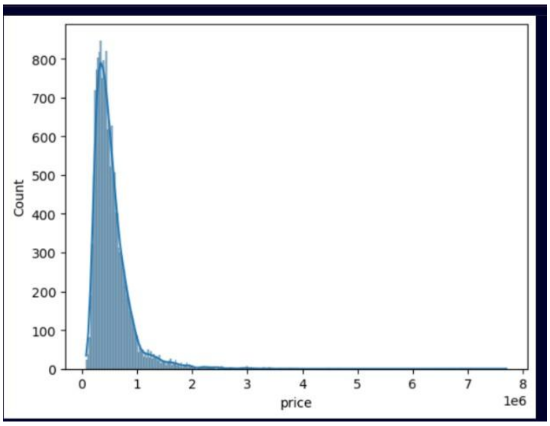
This approach allows for a systematic evaluation of model complexity versus performance gain, while also examining the practical contribution of visual features in real estate price prediction.

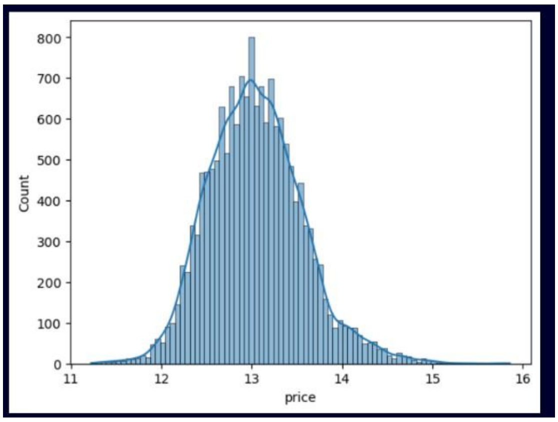
Exploratory Data Analysis (EDA)

The purpose of this section is to understand the **distribution of house prices** and to visually inspect **satellite images** in order to identify patterns that may influence property value. EDA helps justify modelling choices such as log transformation and the use of image based features.

1)Price Distribution Analysis

To understand the nature of the target variable, we first analyze the distribution of house prices.

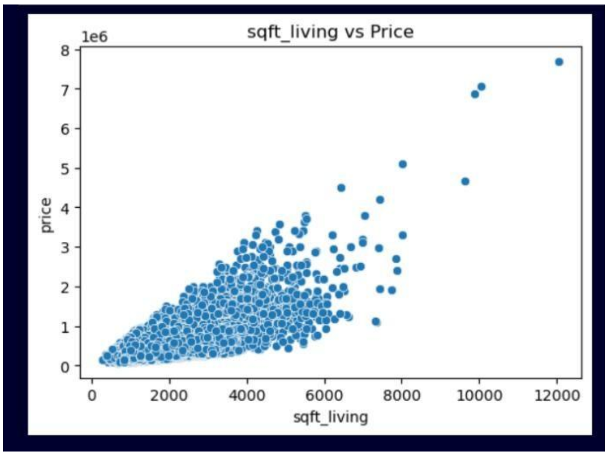




The distribution of house prices is highly right-skewed, with a small number of properties having very high prices. Such skewness can negatively impact regression model performance.
To address this, a logarithmic transformation is applied to the target variable. After transformation, the price distribution becomes more symmetric, which helps stabilize training and improves convergence for machine learning models.

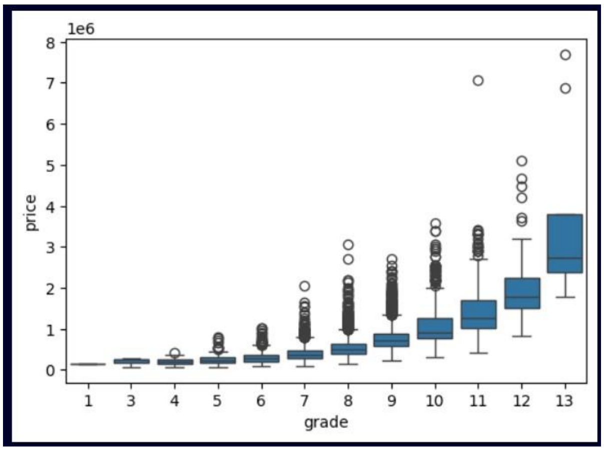
2)Relationship between target Variable and Key feature

i) sqft_living



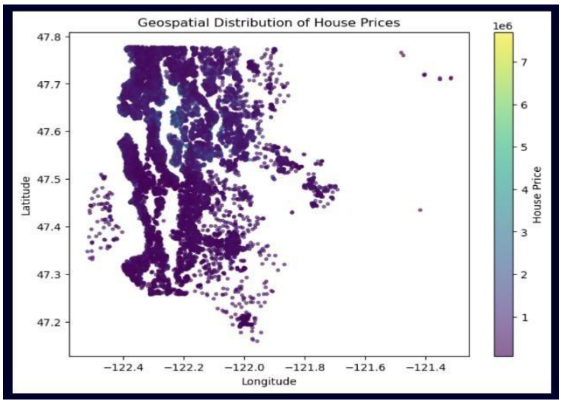
The scatter plot shows a clear positive relationship between living area (sqft_living) and house price, indicating that larger houses generally command higher prices. However, the relationship is not strictly linear. As the living area increases, the spread of prices also widens, suggesting the influence of additional factors such as location, neighbourhood quality, and surrounding environment.

ii) Grade



The box plot illustrates the relationship between property grade and house price. A strong monotonic increase in median price is observed as the grade improves, indicating that higher grade properties are consistently valued more. This suggests that the grade feature effectively captures overall construction quality, design, and finishing standards. Additionally, higher grades show a wider interquartile range and a larger number of high value outliers. This indicates greater price variability among premium properties, likely influenced by location, size, and surrounding amenities. Lower grade properties exhibit more compact price distributions with limited upside, reflecting more uniform valuation.

3)Geospatial Distribution of House Prices



The geospatial visualization highlights a strong dependence of house prices on location. Higher priced properties are concentrated in specific geographic clusters, while lower priced houses are more widely distributed. This spatial clustering indicates that location driven factors such as neighbourhood desirability, proximity to urban centers, waterfront access, and infrastructure play a significant role in determining property value

4) Visual Validation of Waterfront Effect

The geospatial visualization highlights a strong dependence of house prices on location. Higher priced properties are concentrated in specific geographic clusters, while lower priced houses are more widely distributed. This spatial clustering indicates that location driven factors such as neighborhood desirability, proximity to urban centers, waterfront access, and infrastructure play a significant role in determining property value.

5) Visual Neighborhood Density Comparison

The figure compares satellite images from high density and low-density neighbourhoods to highlight visual differences in urban structure. High density regions exhibit closely packed buildings, limited open spaces, and dense road networks, indicating more urbanized environments. In contrast, low-density areas show larger plots, increased green cover, and more open surroundings, reflecting suburban or semirural settings.

These visual characteristics have a direct impact on property valuation. High density neighbourhoods often benefit from better connectivity and proximity to urban amenities, while low-density regions may command premium prices due to privacy, greenery, and lower congestion. Such patterns are difficult to quantify using tabular features alone but are effectively captured through satellite imagery.

6) Curb Appeal & Visual Context vs Price

The comparison between high price and low-price property locations reveals clear visual patterns in the surrounding environment. High priced properties are frequently associated with waterfront proximity, dense green cover, well planned layouts, and organized residential clusters. These areas often exhibit larger plots, recreational spaces, and scenic surroundings, all of which contribute to higher perceived and market value.

Financial and Visual Insights

This section analyses how visual characteristics extracted from satellite images influence house prices and provides financial intuition behind these patterns. By comparing high price and low-price regions, we identify key environmental and structural features that contribute to property valuation.

1. Green Cover vs. Built-Up Areas

Properties surrounded by dense tree cover, parks, and open green spaces consistently correspond to higher prices. Greenery enhances aesthetic appeal, improves air quality, reduces noise pollution, and provides recreational value, all of which increase buyer willingness to pay. From a financial perspective, green neighbourhoods are perceived as premium residential zones and tend to retain long-term value.

2. Waterfront and Natural Features

Satellite images show that properties located near water bodies such as lakes or rivers command significantly higher prices. Waterfront access is a scarce resource and is strongly associated with luxury housing segments. These visual cues are powerful indicators of premium valuation but are difficult to capture using tabular data alone.

3. Neighbourhood Density and Layout

High-value areas typically exhibit planned layouts, wider roads, organized housing blocks, and visible open spaces. These features reflect better urban planning and infrastructure investment. Conversely, irregular layouts with narrow roads and

high structural density often indicate lower-priced neighbourhoods.
This density related information is implicitly learned by the CNN through spatial patterns in satellite imagery.

4. Infrastructure and Negative Externalities

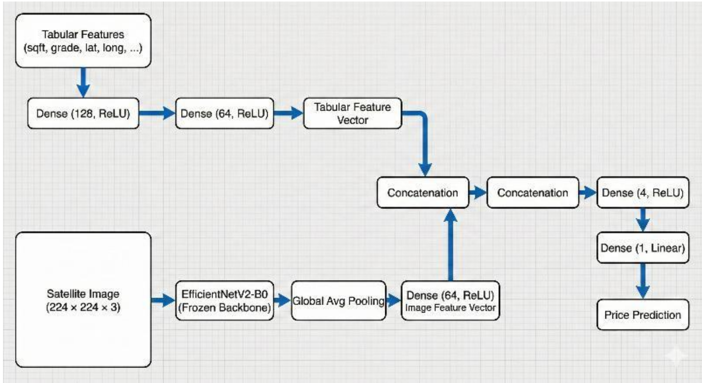
Low-priced properties are frequently observed near highways, construction zones, and industrial areas. While such locations may offer connectivity, they introduce noise, pollution, and reduced visual appeal, which negatively affect property value. These negative externalities are visually detectable but rarely available as structured features.

5. Implications for Modelling

These financial and visual observations highlight why satellite imagery adds value beyond tabular data. CNN based models can automatically learn complex spatial patterns such as vegetation density, land usage, and neighbourhood planning, enabling the multimodal model to better capture real-world valuation drivers.

Model Architecture and Design

I implemented a multimodal deep learning architecture that combines tabular property data with satellite image data to predict house prices. The model is designed to jointly learn structured numerical relationships and visual neighbourhood patterns that influence property valuation.



Tabular Data Branch

The tabular branch processes structured property features such as living area, grade, location coordinates, and other numerical attributes.

- Input features are first standardized to ensure stable optimization.
- The tabular input is passed through a two-layer fully connected neural network:
  - Dense layer with 128 neurons (ReLU)
  - Dense layer with 64 neurons (ReLU)

This branch learns non-linear interactions among property attributes and outputs a compact tabular feature embedding.

Image Data Branch (CNN)

The image branch processes satellite images corresponding to each property location.

- Images are resized to 224 × 224 and normalized to the $[0, 1]$range.
- A pretrained EfficientNetV2-B0 backbone (ImageNet weights) is used as a fixed feature extractor.
- The backbone is frozen to prevent overfitting and reduce training cost.
- Global Average Pooling is applied, followed by a Dense layer (64 neurons, ReLU).

This branch extracts high-level visual features such as green cover, building density, road layout, and proximity to water or infrastructure.

Multimodal Fusion and Prediction

The feature vectors from the tabular model and image model are concatenated to form a unified representation. This fused feature vector is passed through:

- A Dense layer (64 neurons, ReLU)
- A final linear output layer producing the scaled house price prediction

During training, the target variable is standardized to optimize RMSE effectively. Predictions are later inverse-transformed to obtain real-world price values.

Results and Performance Comparison

This section compares the performance of tabular-only models with the proposed multimodal (tabular + satellite image) model to evaluate the contribution of visual features.

1. Tabular Data Only Models

Multiple regression models were trained using only tabular features to establish a strong baseline. The following models were evaluated using RMSE and R^2 score:

Model	RMSE	R^2 Score
Random Forest	0.1936	0.8642
Ridge Regression	0.2500	0.7735
XGBoost	0.1742	0.8901
LightGBM	**0.1672**	**0.8987**

Among all models, LightGBM achieved the lowest RMSE and highest R^2 score, indicating superior performance in capturing non-linear relationships within the tabular data. As a result, LightGBM was selected as the best-performing tabular baseline.

2. Multimodal Model (Tabular + Satellite Images)

A multimodal deep learning model was then trained by combining tabular features with satellite image features extracted using a pretrained CNN. Two training strategies were evaluated:

(a) RMSE-Optimized Training

- Test RMSE (scaled): 52.74
- Test R^2 Score: −2876.31

This configuration performed poorly, indicating severe instability and misalignment between target scaling and optimization.

(b) MAE-Optimized Training

- Test MAE: 112,738.25
- Test R^2 Score: 0.6858

While MAE optimization improved stability compared to RMSE optimization, the overall performance still remained significantly worse than the tabular-only LightGBM model.

3. Comparative Analysis

Despite incorporating satellite imagery, the multimodal model did not outperform the tabular-only approach. This suggests that, for the given dataset and setup, tabular features alone were sufficient to explain most of the variance in house prices.

4. Reasons for Multimodal Underperformance

Based on prior experience and empirical observation, the following factors likely contributed to the weaker multimodal results:

Weak Image-Target Signal

The satellite images used in this project were obtained through a free public imagery API, which imposed limitations on image resolution, zoom level, and update frequency. As a result, many fine-grained visual details that could strongly influence property valuation such as building condition, road quality, and small-scale neighbourhood features were not clearly distinguishable. Although multiple image configurations and zoom levels were explored, the available free sources consistently produced images with limited discriminative power. More advanced satellite imagery providers and street-level data sources are typically paid services, and access constraints prevented their use in this project. Consequently, the visual signal extracted from images was relatively weak and less directly correlated with house prices.

Dominance of Strong Tabular Predictors

The tabular dataset contained several highly informative features, including living area, property grade, and geographic location, which already explained a significant portion of price variability. These features inherently capture structural quality, size, and macro-location effects, leaving limited residual variance for image features to explain.