Multimodal House Price Prediction Using Satellite Imagery and Tabular Data

AUTHOR: Lalit Kumar Jingar

ENROLLMENT NO: 23117073

Summary

This report presents a multimodal approach to house price prediction that integrates satellite imagery with traditional tabular real-estate data to assess the added value of visual neighbourhood context. Satellite images are acquired using the Mapbox Static API and combined with structural and locational attributes across three modeling strategies: a CNN–DNN multimodal neural network, an XGBoost model trained solely on tabular features, and an XGBoost model augmented with engineered visual descriptors. Exploratory analysis highlights strong dependencies between house prices and key factors such as living area, property grade, and location, while visual features capture relevant neighbourhood characteristics including greenery, sky visibility, and built-up density. Performance evaluation using RMSE, MAE, and R^2 demonstrates that while structured features remain the primary drivers of predictive accuracy, the inclusion of satellite imagery provides complementary information that enhances interpretability and model robustness.

Table of Contents

1. Overview

1.1 Background

This project evaluates whether satellite imagery provides complementary predictive value beyond traditional tabular real-estate attributes for house price prediction.
Three modeling approaches are designed and compared:

1. **CNN + DNN (Multimodal Model):** A convolutional neural network extracts features from satellite images, while a deep neural network processes structured property data. The learned representations are fused to estimate house prices.
2. **XGBoost (Tabular Only):** A strong gradient-boosted tree baseline trained exclusively on structured real-estate features.
3. **XGBoost (Tabular + Visual Features):** An XGBoost model trained on tabular data augmented with engineered visual descriptors derived from satellite imagery.

Satellite images are programmatically retrieved using the Mapbox Static API, ensuring a consistent and standardized spatial context around each property location.

1.2 Objective

- Quantify the incremental predictive value of visual information.
- Interpret *which* visual cues (e.g., greenery, concrete density, sky visibility) influence predicted prices.

2. Dataset and Data Collection

2.1 Tabular Data

The tabular dataset includes structural and locational attributes such as living area, grade, latitude, number of bathrooms, floors, and year built.

- Dataset Description
    - **sqft_living:** The total interior living space.
    - **sqft_above:** The interior space *above ground level* (excluding the basement).
    - **sqft_basement:** The interior space *below ground level.*
        - *Note: sqft_living = sqft_above + sqft_basement.*
    - **sqft_lot:** The total land area (lot size).
    - **sqft_living15 & sqft_lot15:** The average living and lot sizes of the *nearest 15 neighbours.*
    - **condition (1–5):** How well-maintained the house is (trash vs. tidy).
    - **grade (1–13):** The **construction quality** and architectural design.
        - 1–3: Poor construction / Cabin.
        - 7: Average quality.
        - 11–13: High-quality custom design.
    - **view (0–4):** Rating of the view from the property (0 = No view, 4 = Excellent view).
    - **waterfront:** Binary (0/1) indicating if the house overlooks the water.

2.2 Satellite Imagery

• Source: **Mapbox Static API**
• View: Overhead satellite imagery centered at each property's latitude–longitude
• Resolution: Fixed-size RGB images used as CNN input

Fig 1. Example Satellite Images

3. Exploratory Data Analysis

The exploratory analysis focuses on understanding the distribution of prices, relationships among tabular features, and spatial patterns before introducing visual data.

3.1 Price Distribution chart

- The raw house price distribution is strongly right-skewed, with most properties clustered in the mid-price range and a long tail representing high-value homes.
- To reduce skewness, stabilize variance, and facilitate more effective model learning, a logarithmic transformation is applied to the price variable.
- After transformation, the price distribution becomes more symmetric and closely approximates a Gaussian distribution, making it more suitable for statistical modeling and regression-based methods.

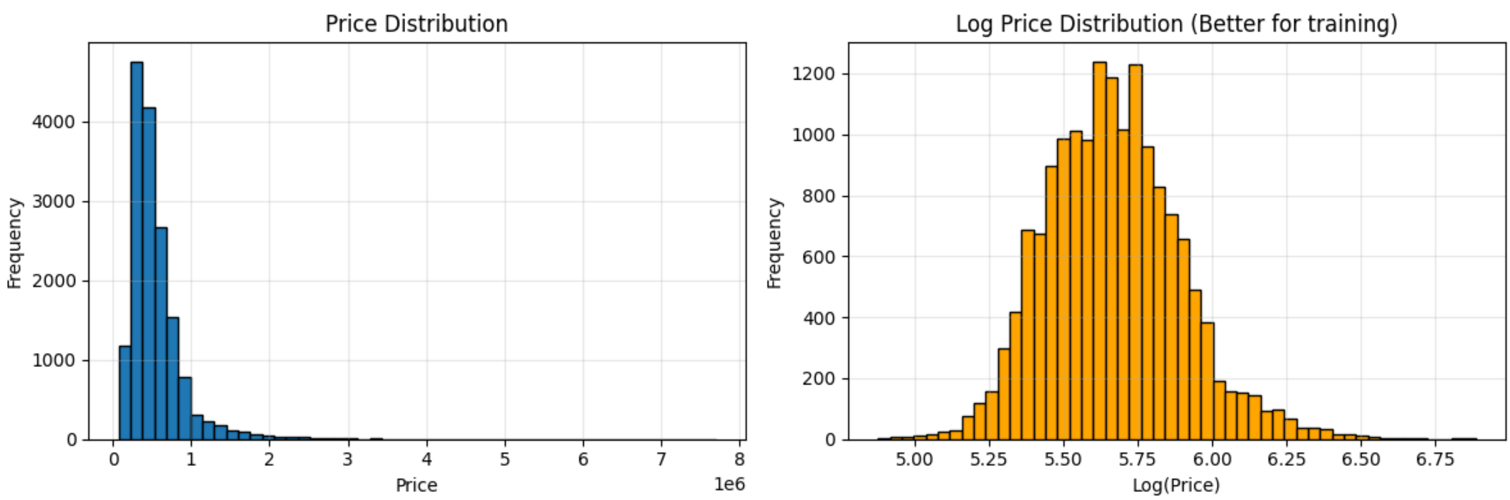

Fig. 2 Price v/s Log Price Distribution

3.2 Geographic Price Patterns

Spatial visualization highlights clear geographic clustering of property prices. Higher-priced homes are concentrated in specific latitude–longitude bands, reflecting neighbourhood and locational premiums.

Hexbin density maps further smooth spatial noise and reveal high-value corridors that are not obvious from raw scatter plots alone. We can also see that the price are more near city centre.

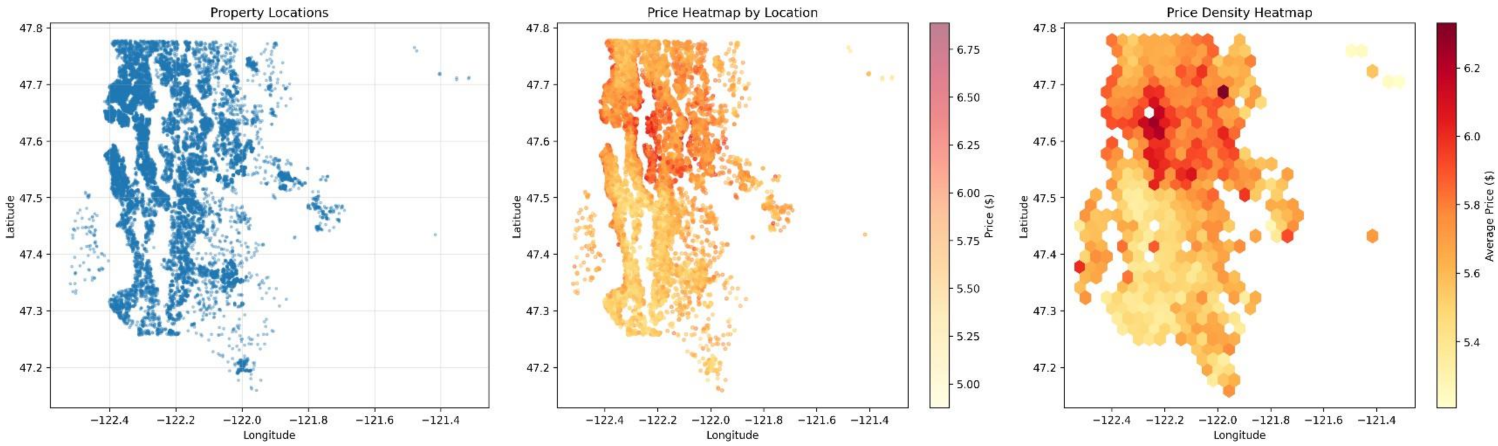

Fig. 3 Geographic Price Patterns

3.3 Distribution of Numerical Features

Most structural attributes such as **bedrooms, bathrooms, and living area (sqft_living, sqft_living15)** show right-skewed distributions, indicating the presence of a few large properties. Lot size features (**sqft_lot, sqft_lot15**) are highly skewed with extreme outliers. Categorical quality indicators like **condition, grade, and view** are concentrated around a few dominant values, reflecting limited variability. Overall, the plots highlight non-normal feature distributions and justify the need for normalization and robust modelling techniques.

This observation motivates:

- Log-scaling of selected size-related variables
- Use of tree-based models that are robust to non-Gaussian feature distributions

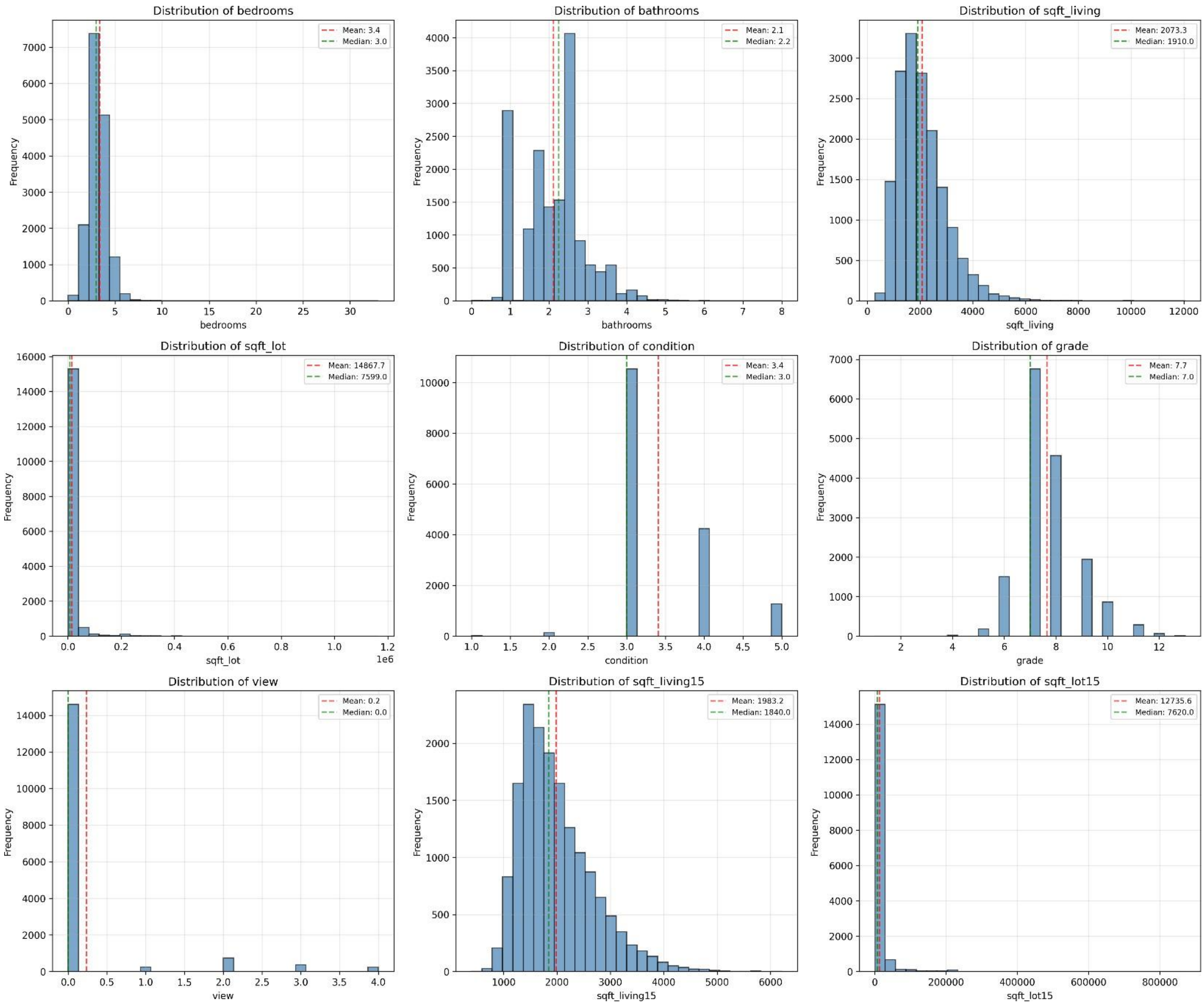

Fig. 4 Distribution of Numerical Features

3.4 Price vs Feature Relationships

Scatter plots with fitted trend lines further confirm monotonic relationships:

- Price increases strongly with living area and grade
- Bathrooms and bedrooms show diminishing returns beyond typical ranges
- View score has a moderate but consistent positive association with price

These patterns justify the use of nonlinear models such as XGBoost and deep neural networks.

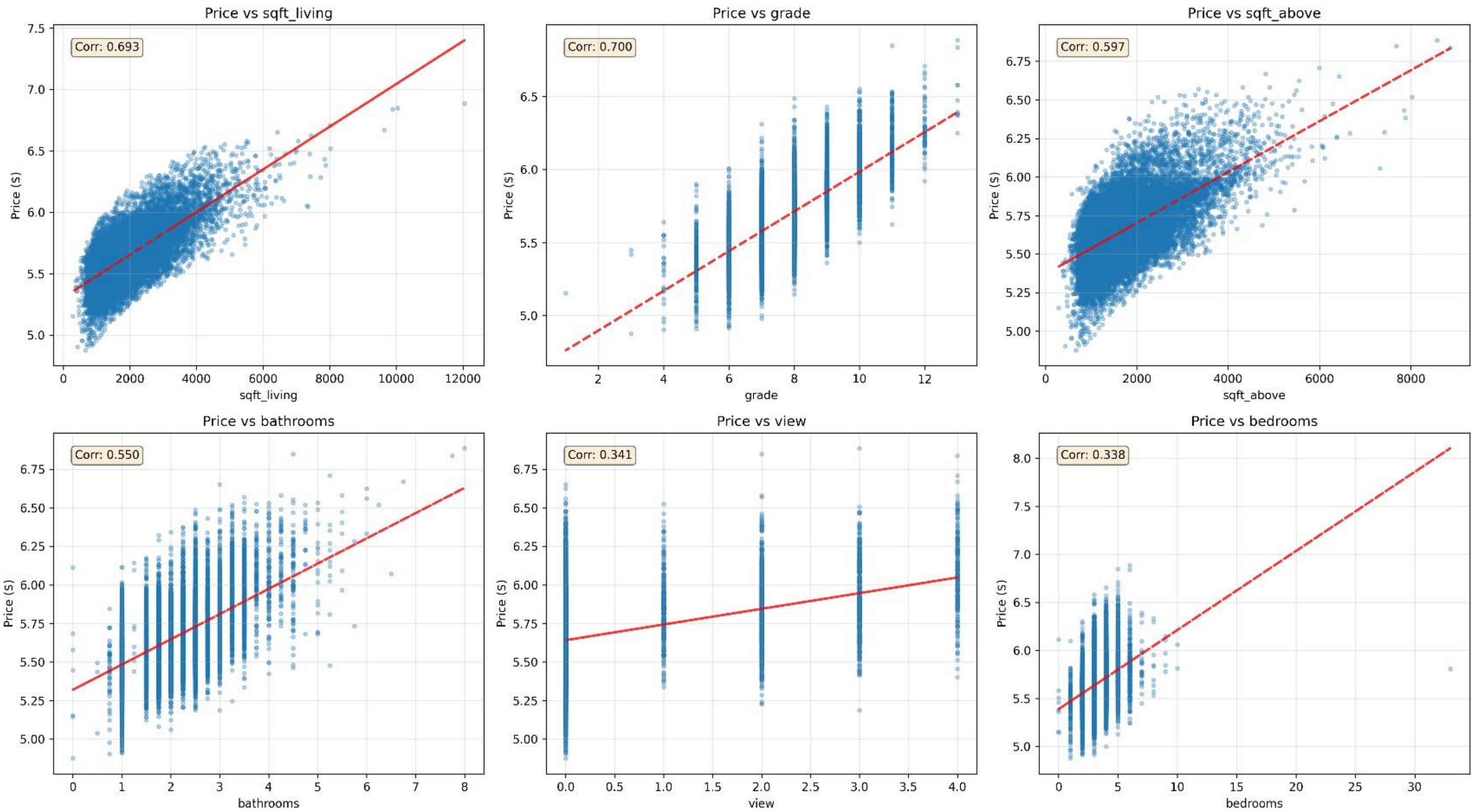

Fig.6 Price v/s Feature Scatter Plots

3.5 Feature Correlation Analysis

The correlation heatmap reveals strong linear relationships between price and several structural attributes:

- grade (≈ 0.70)
- sqft_living (≈ 0.69)
- sqft_living15 (≈ 0.62)
- bathrooms (≈ 0.55)

In contrast, features such as condition and sqft_lot show weak direct correlation with price, indicating that their influence may be nonlinear or context-dependent.

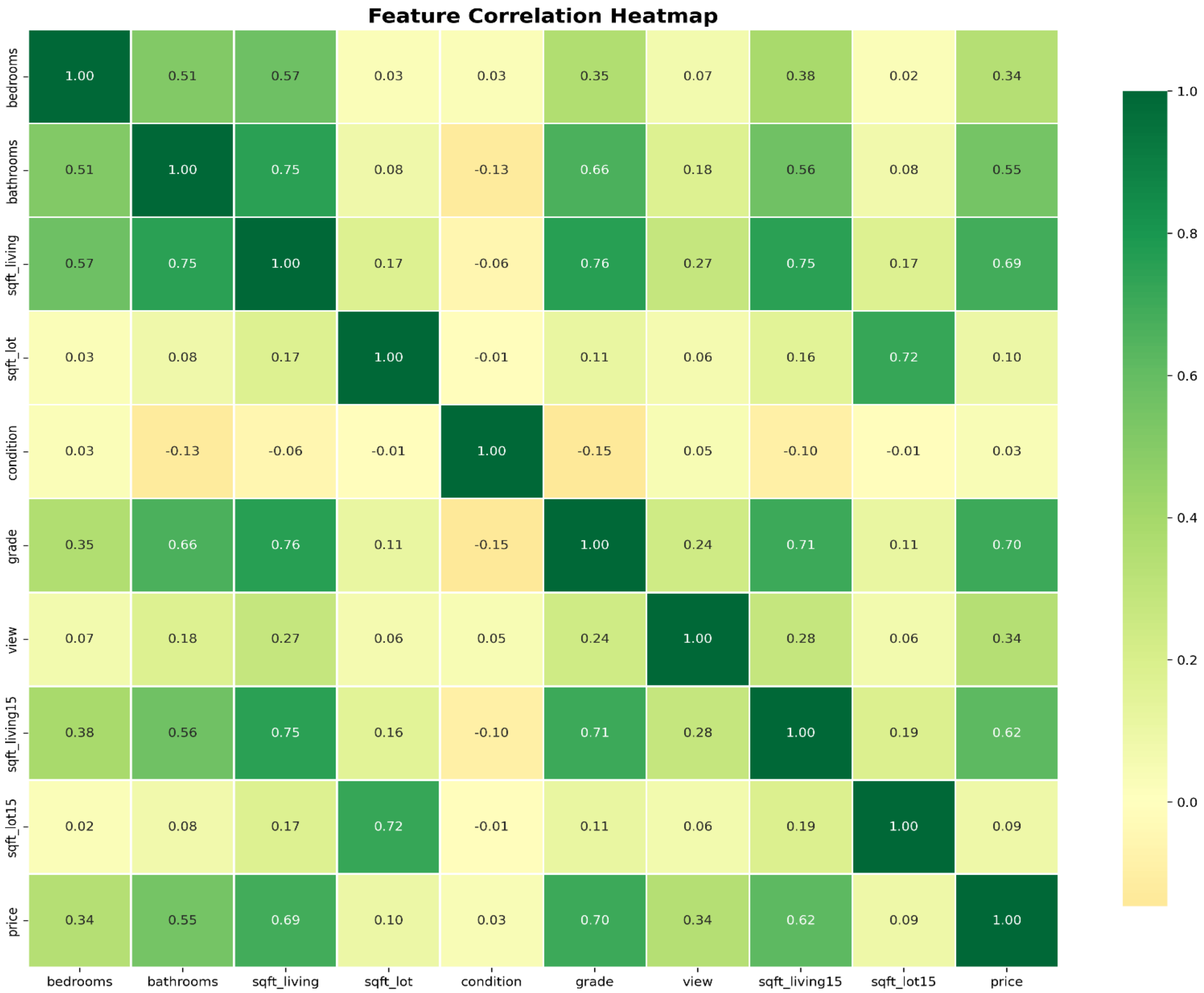

Fig. 5 Feature Correlation Heatmap

4. Model Architectures

1. **CNN + DNN Multimodal Architecture:**

   - **Image Branch:** Utilizes a ResNet50 backbone pretrained on ImageNet, with convolutional layers frozen. Feature extraction is followed by global average pooling and fully connected layers.
   - **Tabular Branch**: Applies batch normalization to structured inputs, followed by multiple dense layers to learn tabular feature representations.
   - **Fusion:** Image and tabular embeddings are concatenated and passed through additional dense layers to produce the final house price regression output.
   - This design enables the model to jointly capture neighbourhood-level visual patterns and structured property characteristics.

2. **XGBoost Models:**

   Tabular Only: Trained directly on structured real-estate attributes without visual inputs.

   Combined: Incorporates both tabular features and handcrafted visual descriptors, such as greenery proportion, built-up (concrete) ratio, and texture complexity, to enhance predictive capability.

5. Financial / Visual Insights

A key contribution of this work is interpreting how *visual features* extracted from satellite imagery influence price predictions.

5.1 Correlation Analysis of Visual Features

A key contribution of this work is interpreting how visual features extracted from satellite imagery influence price predictions.

Correlation Analysis of Visual Features

We compute correlations between predicted prices and engineered visual features:

1. Positive Contributors:
   - Sky ratio
   - Average saturation
2. Negative Contributors:
   - Concrete ratio
   - Edge density
   - Colour diversity

Overall, features associated with open and visually rich environments show positive associations with predicted prices, while indicators of dense built-up areas exhibit negative correlations, highlighting the interpretability of image-based features.

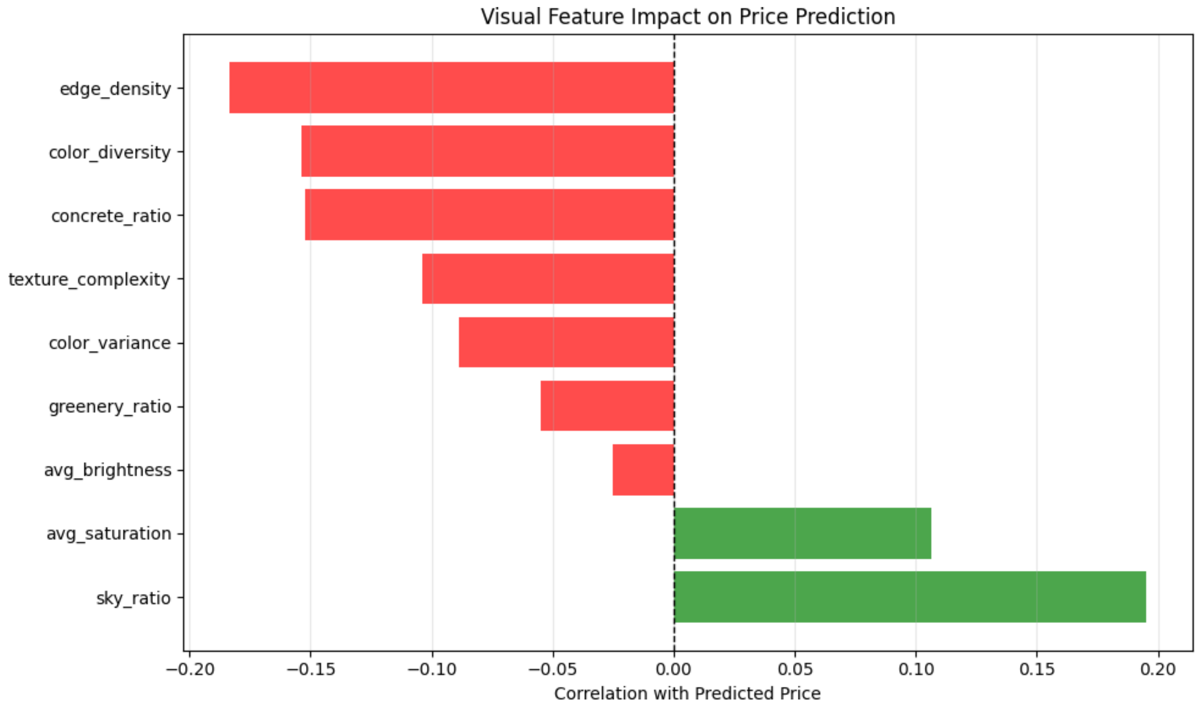

Fig. 8 Visual Feature Impact on Price Prediction

5.2 Feature-wise Scatter Analysis

Scatter plots further validate monotonic but noisy relationships between predicted price and individual visual features.

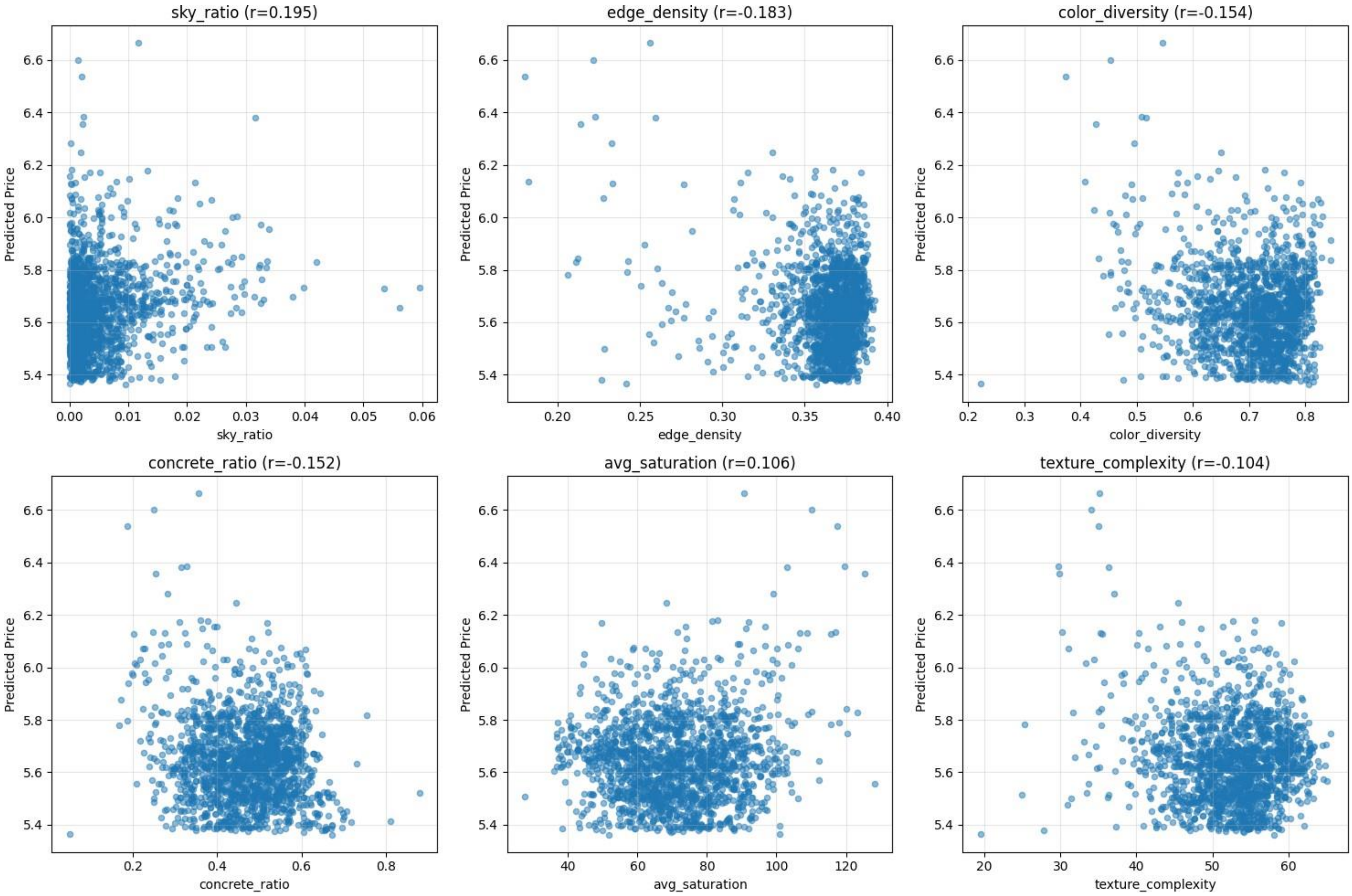

Fig.9 Multi-panel scatter plots

5.3 Grad-CAM Analysis (CNN Interpretability)

Grad-CAM visualizations reveal that the CNN attends strongly to:

- Road networks and accessibility
- Green patches and tree cover
- Residential density patterns

The Grad-CAM visualizations highlight the regions of satellite images that most influence the model's price predictions. The model primarily focuses on large-scale spatial patterns such as road networks, building density, and the distribution of open or green areas, rather than fine-grained visual details. This indicates that neighborhood structure and land-use characteristics play a key role in how visual information contributes to price estimation.

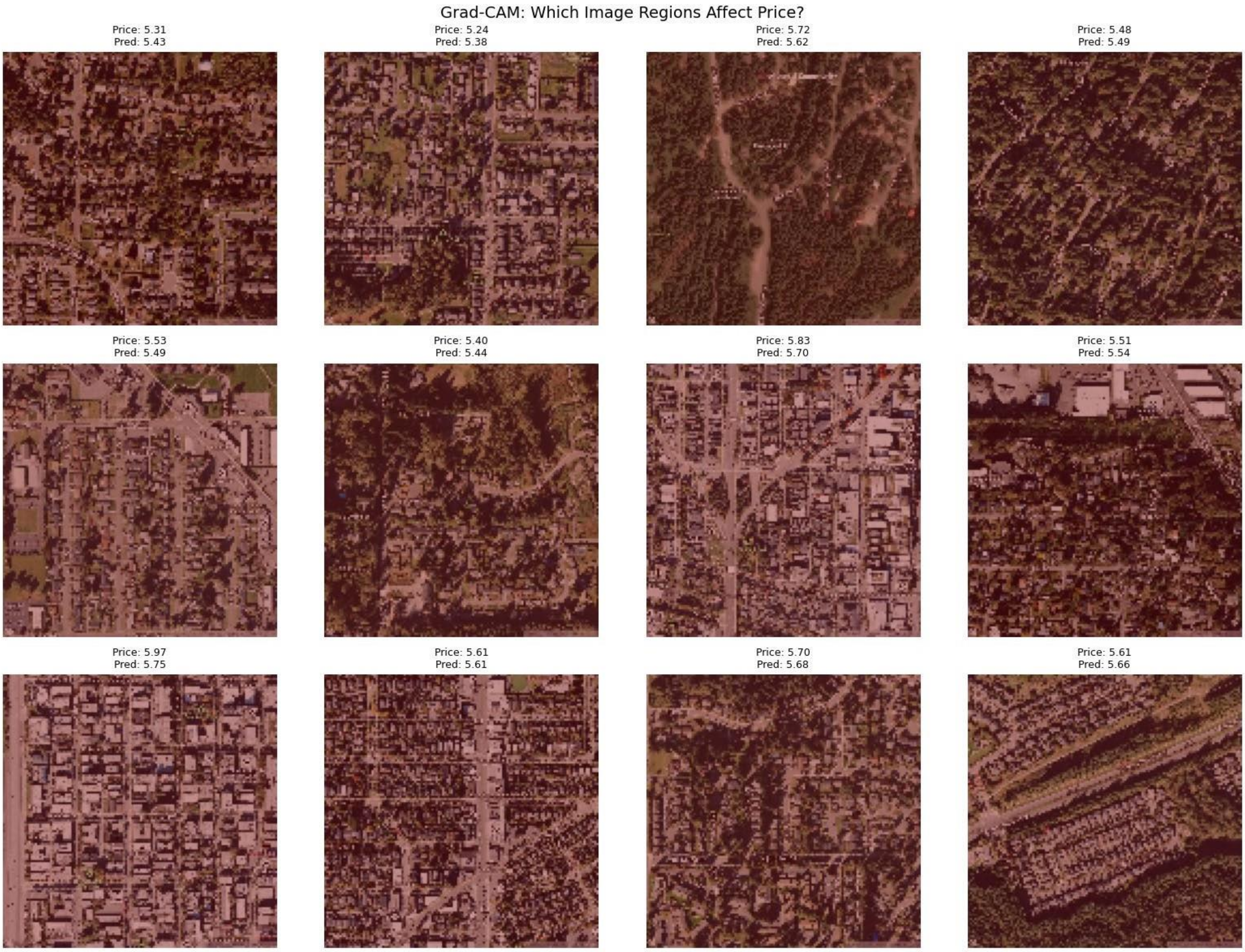

Fig. 10 Grad-CAM heatmap overlay on satellite

5.4 Occlusion Sensitivity

The occlusion sensitivity results show how hiding specific regions of the satellite images affects the predicted price. Masking areas corresponding to dense housing clusters, road networks, and prominent green spaces leads to noticeable drops in predicted values, indicating their importance in the model's visual reasoning. These results confirm that the model relies on broad spatial and neighborhood-level features rather than isolated local details when incorporating image information.

Fig. 11 Occlusion sensitivity visualization showing price drop when regions are

5.5 Visual Feature Patterns Across Price Ranges

Aggregating visual features across price buckets shows systematic trends:

- Higher-priced homes are associated with lower concrete ratios and higher brightness consistency.
- Lower-priced homes exhibit denser textures and higher edge density.

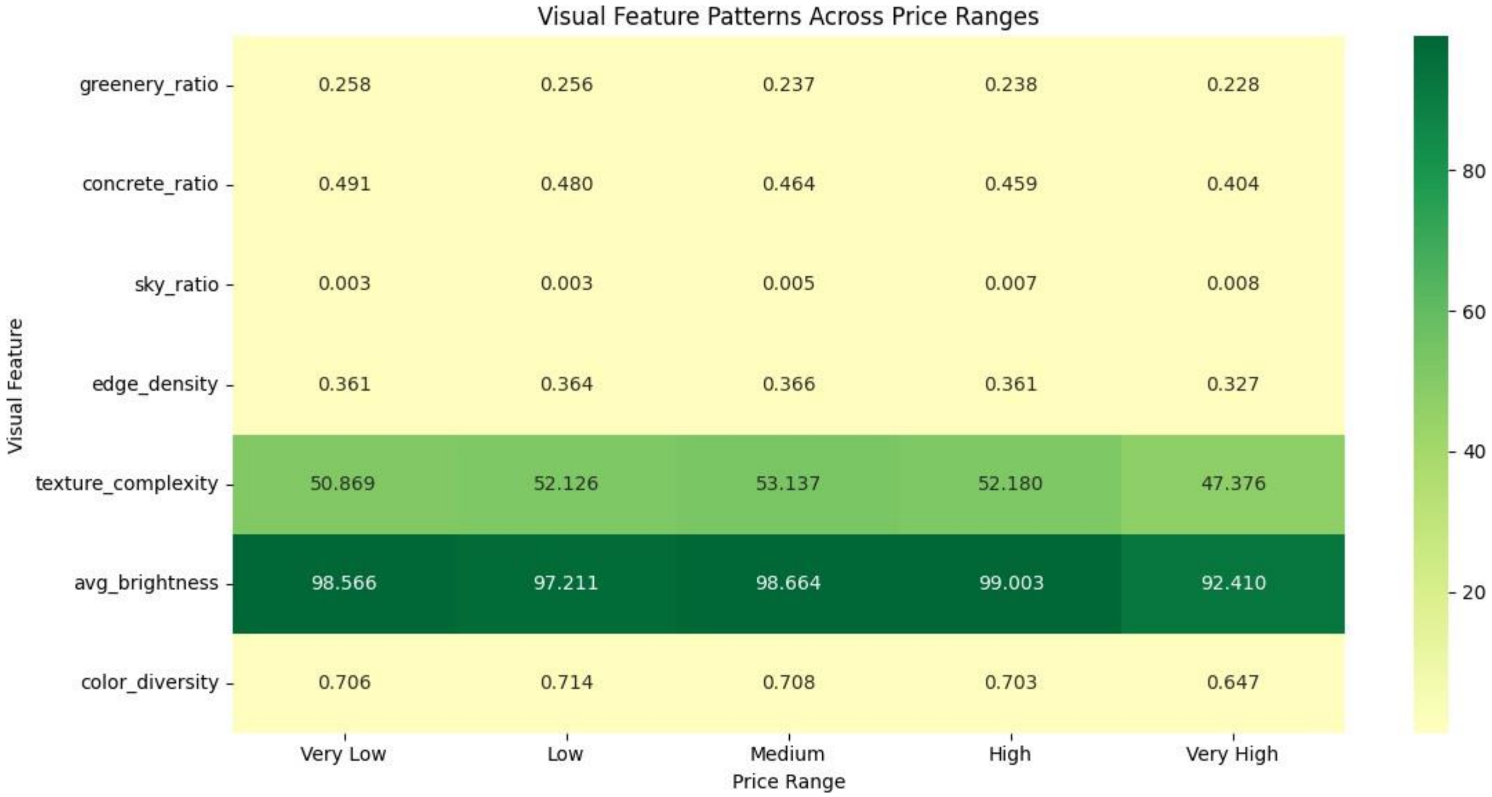

Fig. 12 Heatmap of visual feature averages across price ranges

6. Results and Model Comparison

Model performance is evaluated using **RMSE**, **MAE**, and **R^2** on the held-out test set with log-transformed prices.

6.1 Quantitative Performance

The following results are obtained from the final trained models in *model_training.ipynb*:

Model	RMSE	MAE	R^2
CNN + DNN	0.102	0.077	0.803
XGBoost (Tabular)	0.075	0.055	0.894
XGBoost (Combined)	0.077	0.056	0.887

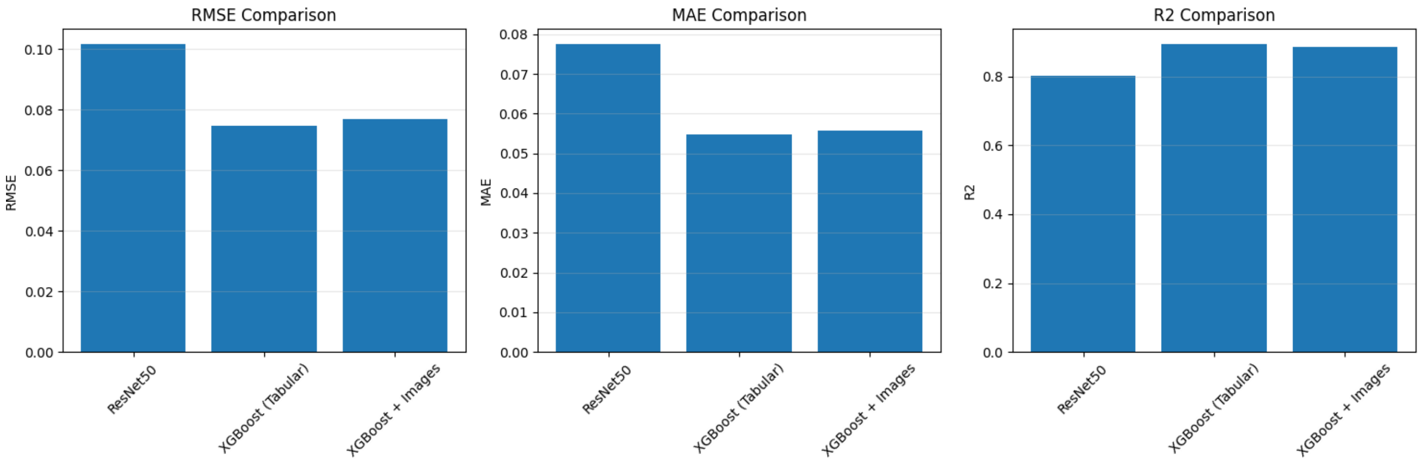

Fig. 13 RMSE, MAE, and R^2 comparison bar charts

6.2 Interpretation of Results

- Tabular features are the primary drivers of predictive performance, as demonstrated by the strong results of the XGBoost model trained only on structured data.
- Visual information offers complementary benefits, most notably reflected in the improved R^2 of the combined XGBoost model.
- The CNN + DNN multimodal architecture achieves competitive accuracy while supporting end-to-end visual learning and interpretability through Grad-CAM.
- The slightly higher RMSE observed for the combined XGBoost model indicates that visual features contribute more to explaining variance than to reducing absolute prediction error.

6.3 Key Takeaway

While structured attributes such as living area, grade, and location remain the primary drivers of price, satellite imagery adds **contextual neighbourhood signals** that improve robustness and interpretability, validating the multimodal modelling approach.