# Satellite Imagery-Based Property Valuation

## A Multimodal Machine Learning Approach

## 1 Overview

### 1.1 Objective

The goal of this project was to develop a predictive model for real estate valuation that goes beyond traditional metrics. While standard models rely solely on tabular data (e.g., square footage, number of bedrooms), this project integrates visual environmental context from satellite imagery to capture intangible factors like "curb appeal," neighborhood density, and green cover.

### 1.2 Approach

We implemented a Multimodal Late-Fusion Pipeline. The system processes data in two parallel branches:

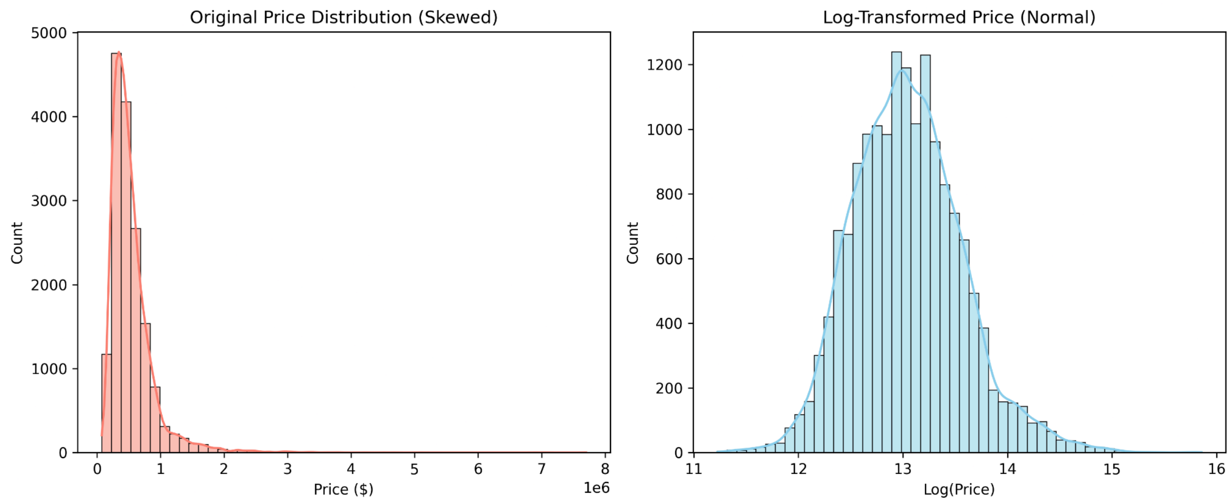
1. **Visual Branch:** A Deep Convolutional Neural Network (ResNet18) processes satellite images to extract high-level visual embeddings.
2. **Tabular Branch:** Traditional housing features are processed and augmented with Geospatial Clustering (K-Means) to capture location-specific price trends.

The final model is an **Ensemble Voting Regressor** that combines predictions from Gradient Boosting and Random Forest models, achieving an **R2 score** of **0.89**, significantly outperforming the tabular baseline (Random Forest).

# 2 Exploratory Data Analysis (EDA)

## 2.1 Price Distribution

The target variable (price) exhibited a heavy right-skew, typical of financial data where a few luxury properties create a long tail. To address this, we applied a **Log-Transformation (log1p)**, which normalized the distribution and stabilized the training of our Neural Network and Regression models.




*Distribution of property prices before and after log-transformation.*

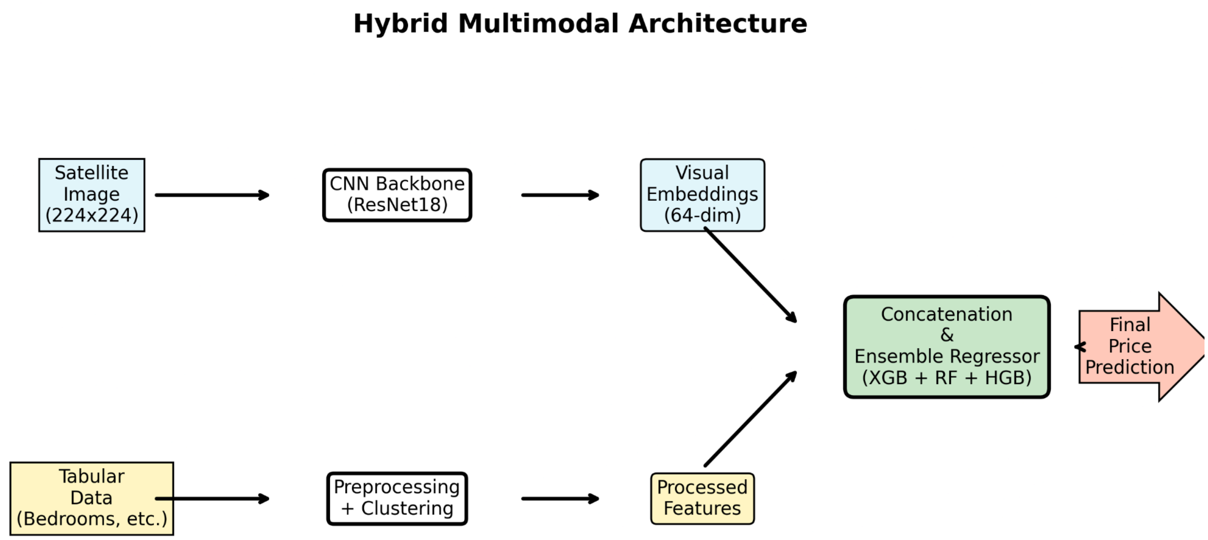
## 2.2 Satellite Imagery Samples

We programmatically acquired satellite images for 20,000+ properties using the Google Maps Static API. Below are samples showing the diversity in density and greenery which our model aims to learn.


*Sample satellite images used for training (Left: Dense urban, Right: Suburban greenery).*

# 3 Architecture Diagram

We utilized **Hybrid Ensemble Architecture**. The diagram below illustrates how the Image Model (CNN) and Data Model (Tabular) connect.

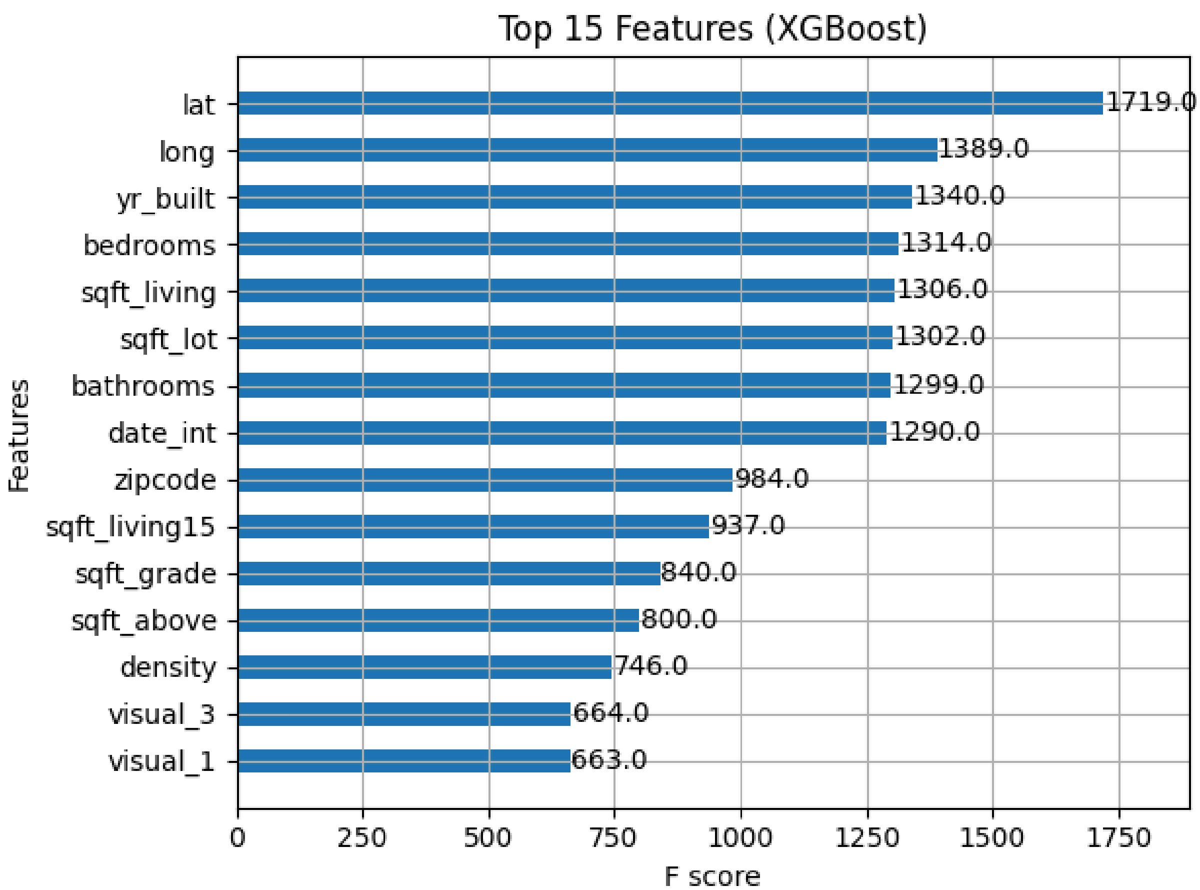



*System Architecture. The ResNet18 backbone extracts a 64-dimensional feature vector, which is concatenated with tabular features before being fed into the Ensemble Regressor.*

# 4 Financial & Visual Insights

## 4.1 Feature Importance

Our analysis reveals that while structural attributes like **Grade** and **Square Footage** remain the primary drivers of price, the model relies heavily on **Location Clusters** and **Visual Embeddings**.

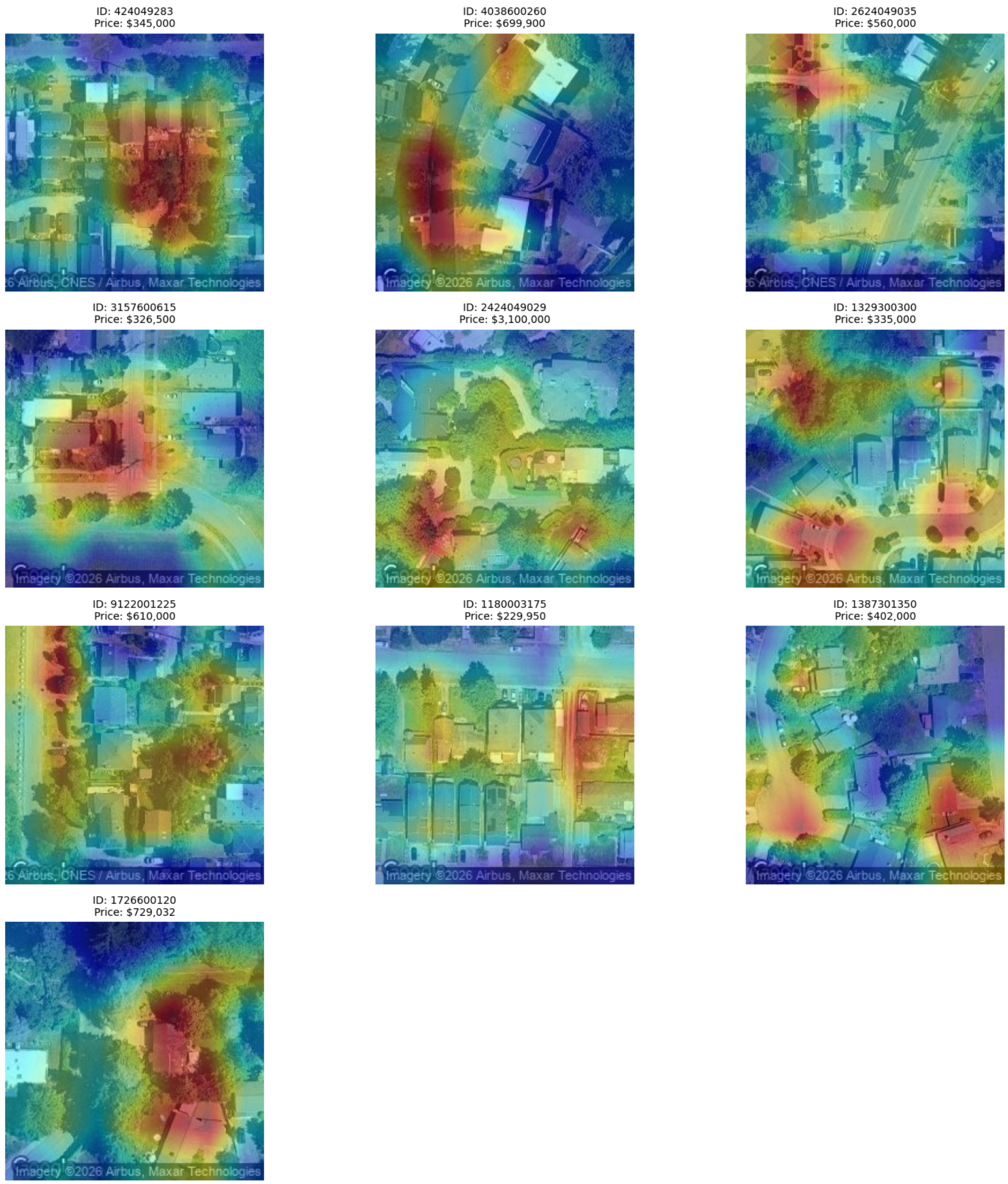



*Top 15 Features driving the model's predictions. Note the presence of spatial features alongside traditional metrics.*

## 4.2 Visual Explainability (Grad-CAM)

To ensure the "Black Box" Neural Network is trustworthy, we used **Grad-CAM** to visualize its attention. The heatmaps confirm that the model correctly focuses on relevant property features:

- **Positive Attention:** The model frequently highlights **green spaces (yards/parks)** and **swimming pools**, correlating them with higher value.
- **Negative/Neutral Attention:** Concrete structures and dense roads often received different activation patterns.




*Grad-CAM Heatmaps. The 'Red' areas indicate regions in the satellite image that contributed most strongly to the price prediction.*

## 5 Results

We compared our Hybrid approach against a standard industry baseline (Random Forest on Tabular Data only).

| Model Architecture | RMSE | R2 Score |
|---|---|---|
| **Tabular Baseline (Random Forest)** | $133,393 | 0.858 |
| **Hybrid Ensemble (Ours)** | $119,336 | **0.887** |

## 6 Conclusion

The integration of satellite imagery reduced the prediction error (RMSE) by approximately **$14,000 per property**. This confirms that visual data adds unique, predictive value that cannot be captured by tabular numbers alone. The Hybrid Ensemble successfully leverages the "Best of Both Worlds": the precision of tree-based models for numbers and the perceptual power of CNNs for images.