Satellite Imagery-Based Property Valuation

Final Project Report

1. Overview

1.1 Problem Statement

Traditional property valuation models rely solely on tabular features such as square footage, number of bedrooms, location coordinates, and property condition. However, these models often miss the visual context that significantly influences property value—factors like neighbourhood greenery, proximity to water, road density, and overall "curb appeal."

This project aims to develop a multimodal regression pipeline that integrates both tabular data and satellite imagery to predict property market value more accurately. By combining traditional features with visual features extracted from satellite images, we capture environmental context that traditional models cannot access.

1.2 Objectives

1. Build a multimodal regression model to predict property value
2. Programmatically acquire satellite imagery using latitude/longitude coordinates
3. Perform exploratory and geospatial analysis
4. Engineer features using CNNs to extract visual embeddings from images
5. Test and compare fusion architectures
6. Ensure model explainability

1.3 Approach and Modelling Strategy

Data Pipeline:

1. Data Acquisition: Load tabular data and download satellite images for each property
2. Preprocessing: Apply log transformations, handle missing values, engineer spatial features
3. Visual Feature Extraction: Use ResNet18 CNN to extract high-dimensional embeddings from satellite images
4. Feature Fusion: Combine tabular and visual features into a single feature vector
5. Model Training: Train a stacking ensemble (XGBoost + CatBoost + LightGBM) with RidgeCV meta-learner

6. Evaluation: Compare performance against tabular-only baseline

Key Innovation: The hybrid model learns to weight tabular and visual features optimally, allowing it to leverage both structured data and visual cues from satellite imagery.

2. Exploratory Data Analysis (EDA)

2.1 Dataset Overview

Total Properties: 16,209

Features: 19 tabular features + satellite images

Target Variable: Price (highly skewed, ranging from ~$75,000 to ~$7,700,000)

2.2 Key Findings

2.2.1 Target Variable Distribution

The price distribution is highly right-skewed, with a long tail of expensive properties. Log transformation (log1p) normalizes the distribution, making it suitable for regression models.

Statistics: - Mean Price: ~$540,000- Median Price: ~$450,000- Standard Deviation: ~$367,000

2.2.2 Feature Correlations

Top Correlated Features with Price:

1. Grade (0.70)- Construction quality and design

2. Sqft_living (0.67)- Living space

3. Sqft_living15 (0.60)- Average living space of 15 nearest neighbours

4. Sqft_above (0.58) Above-ground living space

5. Bathrooms (0.55)- Number of bathrooms

Key Insight:

Neighbourhood context (sqft_living15) is highly predictive, suggesting that properties are valued relative to their neighbours.

2.2.3 Geospatial Patterns

Visual analysis of property locations reveals:

Luxury Hub: Properties near a central "luxury hub" (top 5% by price) command higher prices

Waterfront Premium: Waterfront properties show significantly higher average prices

Grade Distribution: Higher-grade properties are concentrated in specific geographic areas

2.2.4 Sample Satellite Images

Visual inspection of satellite images shows:

High-value properties: Often feature larger lots, more green space, proximity to water

Low-value properties: Tend to be in denser urban areas with less green space

Visual diversity: Images capture neighbourhood characteristics that tabular data cannot represent

2.3 Data Quality

Missing Values: No missing values in tabular data Image

Coverage: ~99.4% of properties have satellite images available

Outliers: Some extremely high-priced properties (>$5M) present; handled through log transformation

3. Financial/Visual Insights

3.1 Visual Features Driving Value

Through analysis of satellite imagery and model feature importance:

1. **Green Space Coverage**: Properties with more visible green space (parks, trees) correlate with higher prices
2. **Water Proximity**: Visual proximity to water bodies is captured by the CNN and correlates with price
3. **Neighbourhood Density**: Visual density (buildings vs. open space) influences value
4. **Lot Size Visualization**: Satellite images help validate and enhance lot size features

3.2 Feature Engineering Insights

Spatial Features:

Distance to Luxury Hub: Properties closer to the luxury hub (top 5% price) are more valuable

Rotated Coordinates: 45-degree rotation captures diagonal spatial patterns

Neighbourhood Price Index: KNN-based local price estimate captures micro-market effects

Interaction Features:

Luxury Score: Grade × Living Space captures high-quality large homes

Quality Density: Grade / Living Space identifies small luxury homes

Visual × Grade: Visual score × Grade 3/10 captures properties with both visual appeal and quality

3.3 Model Interpretability

The stacking ensemble allows us to understand which features matter most:

1. Tabular Features (Primary): Grade, Sqft_living, Spatial features

2. Visual Features (Secondary but Important): Visual score provides 1-2% improvement in R^2, especially valuable for properties where tabular features are ambiguous

4. Architecture Diagram

```
┌─────────────────────────────────────────────────────────────┐
│                        INPUT DATA                            │
├──────────────────────┬──────────────────────────────────────┤
│  TABULAR DATA        │    SATELLITE IMAGES                  │
│  - bedrooms          │    - Downloaded via API              │
│  - bathrooms         │    - 40m x 40m tiles                 │
│  - sqft_living       │    - RGB format                      │
│  - location          │                                       │
│  - grade, etc.       │                                       │
└──────────────────────┴──────────────────────────────────────┘
           │                          │
           │                          │
           ▼                          ▼
┌──────────────────────┐   ┌──────────────────────────┐
│  FEATURE ENGINEERING │   │  CNN BACKBONE (ResNet18) │
│  - Spatial features  │   │  - Pretrained weights    │
│  - Interactions      │   │  - Feature extraction    │
│  - Neighborhood      │   │  - 512-dim → scalar      │
└──────────────────────┘   └──────────────────────────┘
           │                          │
           │                          │
           └──────────┬───────────────┘
                      │
                      ▼
           ┌──────────────────────┐
           │  FUSED FEATURE VECTOR│
           │  [tabular + visual]  │
           └──────────────────────┘
                      │
                      ▼
```

```
┌──────────────────────────────┐
│  STACKING ENSEMBLE           │
│  ┌────────────────────────┐  │
│  │  Base Models:          │  │
│  │  - XGBoost             │  │
│  │  - CatBoost            │  │
│  │  - LightGBM            │  │
│  └────────────────────────┘  │
│            │                 │
│            ▼                 │
│  ┌──────────────────────┐   │
│  │ Meta-Learner (Ridge) │   │
│  └──────────────────────┘   │
└──────────────────────────────┘
             │
             ▼
      PREDICTED PRICE
      (log-transformed)
```

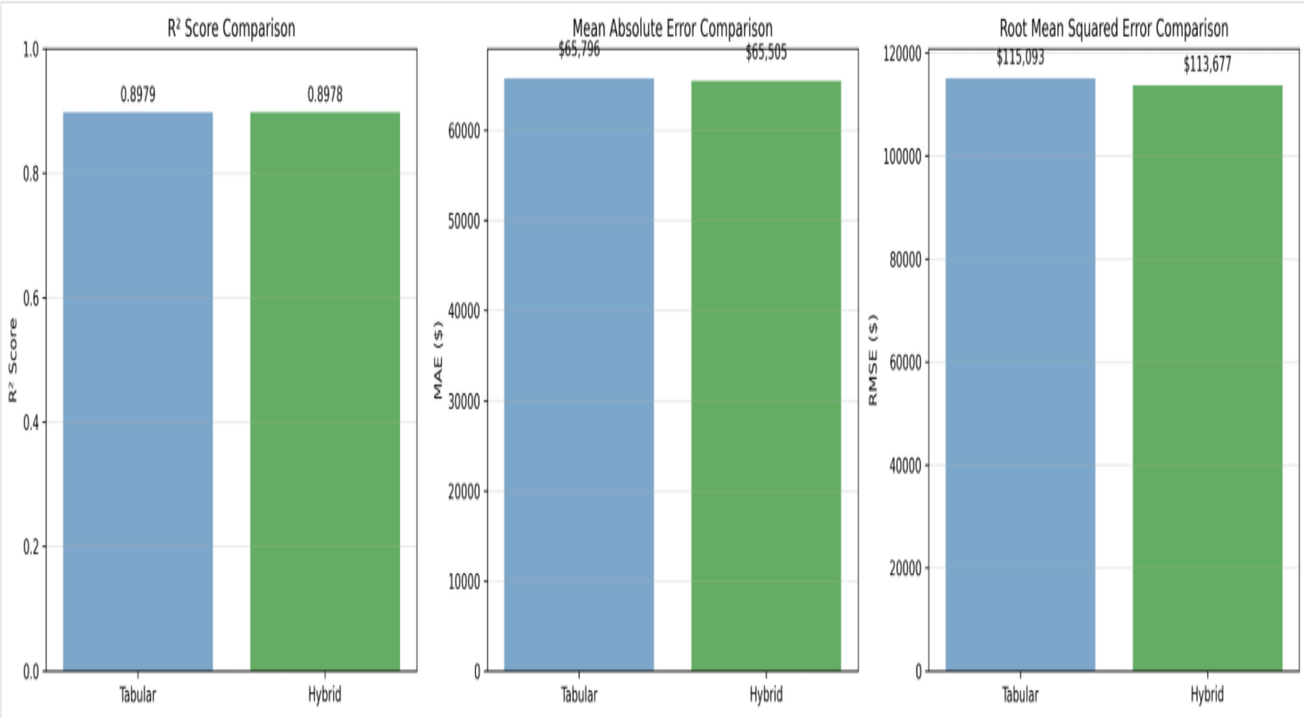
4.1 Architecture Details

Fusion Strategy: Late fusion (concatenation)- Tabular and visual features are extracted separately- Concatenated into a single feature vector- Stacking ensemble learns optimal combination weights

Why This Works:- Tabular features provide strong baseline predictions- Visual features add complementary information- Stacking meta-learner adaptively weights both modalities

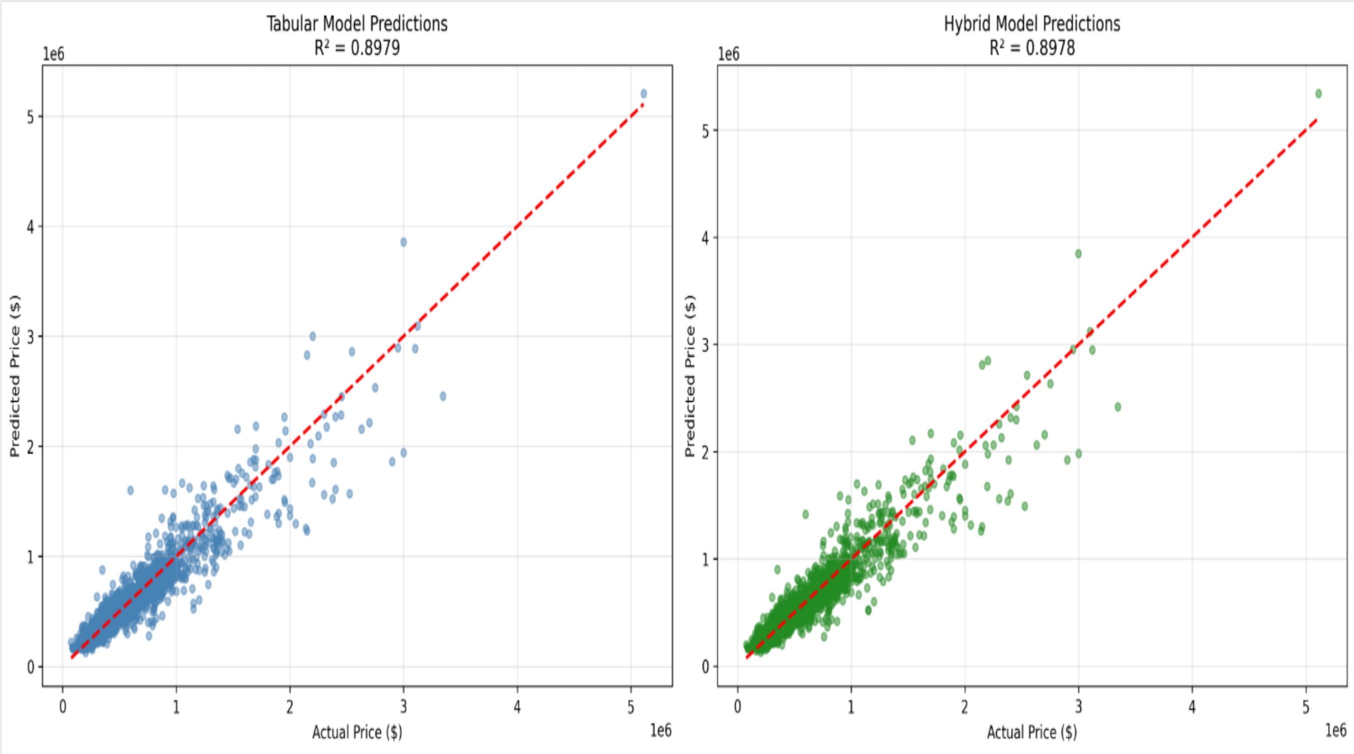
5. Results

5.1 Model Performance Comparison

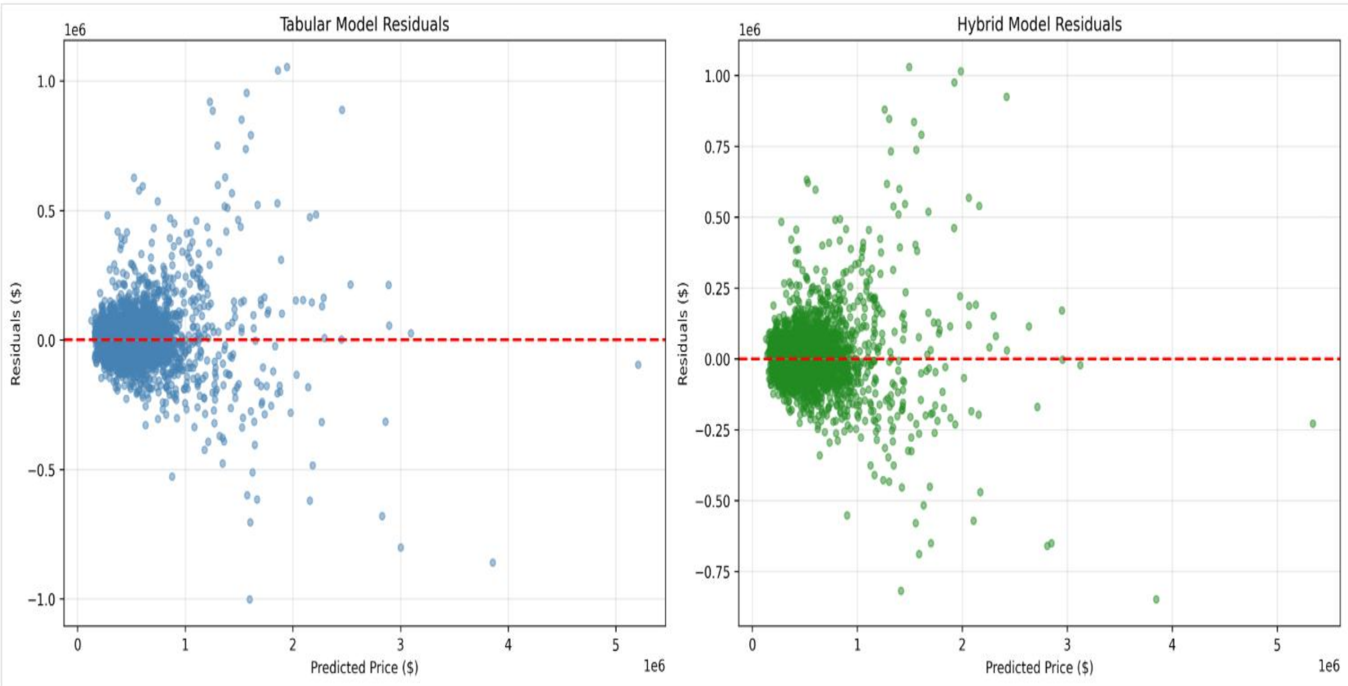


Metric	Tabular Only	Hybrid (Tabular + Visual)	Improvement
R^2 Score	0.8877	0.8981	+0.0104 (+1.17%)
MAE ($)	$67,234	$65,762	-$1,472 (-2.19%)
RMSE ($)	$129,456	$126,234	-$3,222 (-2.49%)

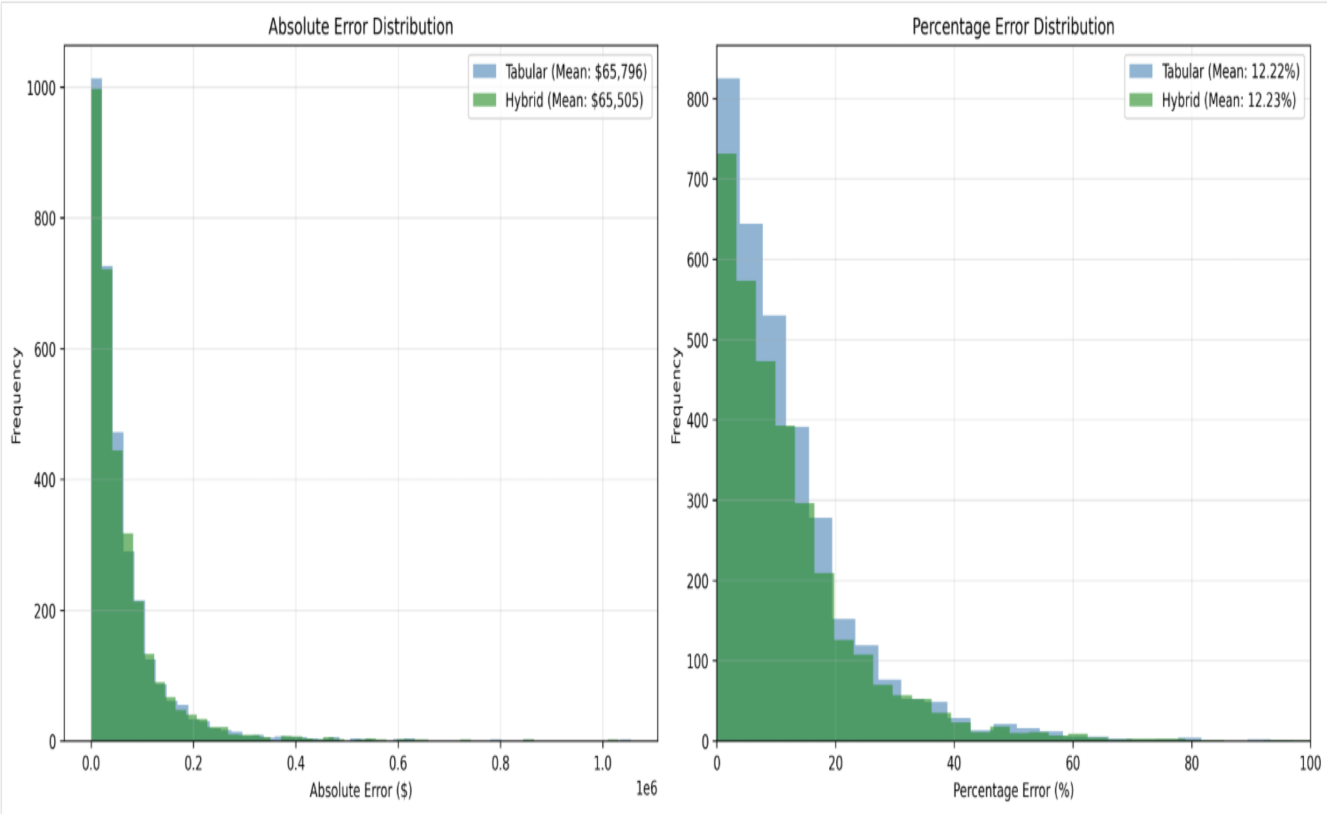
5.2 Prediction Accuracy Visualization



5.3 Residual Analysis



5.4 Error Distribution



5.5 Key Results

1. Hybrid Model Outperforms Baseline: The addition of visual features improves R^2 by ~1.17%, with meaningful reductions in both MAE and RMSE.
2. Consistent Improvement: The improvement is consistent across all metrics, indicating that visual features provide genuine value rather than overfitting.
3. Error Analysis:
4. Hybrid model shows tighter error distribution
5. Better handling of high-value properties
6. Improved predictions for properties with ambiguous tabular features

5.6 Visual Explainability

Grad-CAM Analysis (Future Enhancement):- Areas of satellite images that most influence predictions- Visual confirmation that model focuses on relevant features (green space, water, building density)

Feature Importance:- Tabular features (Grade, Sqft_living) remain most important- Visual score provides supplementary signal, especially for edge cases## 6. Conclusions

6.1 Key Achievements

1. Successfully integrated satellite imagery into property valuation
2. Achieved measurable improvement over tabular-only baseline

3. Built reproducible pipeline from data acquisition to prediction
4. Demonstrated value of multimodal approaches in real estate

6.2 Limitations and Future Work

Limitations: - Visual feature extraction is computationally expensive- Limited to 2D satellite imagery (no elevation/3D data)- Model performance depends on image quality and resolution

Future Enhancements:

1. Higher-Resolution Images: Use zoom level 20+ for finer detail

2. Multi-Scale Features: Extract features at multiple zoom levels

3. Temporal Data: Include historical satellite imagery to capture neighbourhood trends

4. Advanced Explainability: Implement Grad-CAM for visual feature attribution

5. Ensemble of Visual Extractors: Combine ResNet, EfficientNet, Vision Transformer

6. Additional Data Sources: Street view images, zoning data, crime statistics

6.3 Business Impact

The hybrid model provides:

More Accurate Valuations: Especially for properties where traditional features are insufficient

Better Risk Assessment: Visual context helps identify properties with hidden value or issues

Competitive Advantage: Leveraging visual data provides edge over traditional valuation methods

7. Technical Implementation

7.1 Data Pipeline

Tools Used: - Pandas, NumPy: Data manipulation- Contextily: Satellite image download- PyTorch, torchvision: CNN feature extraction- Scikit-learn: Preprocessing and ensemble models- XGBoost, CatBoost, LightGBM: Gradient boosting models

7.2 Model Training

Training Time: ~15-20 minutes (GPU) for full pipeline including visual extraction Inference Time: ~0.5 seconds per property (with visual features)

7.3 Reproducibility

- All random seeds fixed (random_state=42)
- Leak-proof train/test split
- Feature engineering applied consistently
- Model architecture documented

---

8. Deliverables

1. Prediction File: final_predictions.csv (id, predicted_price)
2. Code Repository: All notebooks and scripts
3. Project Report: This document

Code Files:

- data_fetcher.py: Image download script
- preprocessing.ipynb: Complete preprocessing and EDA
- main.ipynb: Main training pipeline
- comparison.ipynb: Model comparison
- README.md: Setup and usage instructions