Satellite Imagery-Based Property Price Prediction

CDC X Yhills Competition 2025-26

Validation R^2	**0.8182**
RMSE	**$143,495**
MAPE	**17.28%**
Training Samples	**12,967**

Multimodal Deep Learning Approach combining EfficientNet-V2-S CNN with Tabular Feature Engineering

Table of Contents

1. Overview & Approach — Problem definition and modeling strategy

2. Exploratory Data Analysis — Price distribution and feature insights

3. Visual Feature Analysis — Satellite imagery insights via Grad-CAM

4. Model Architecture — Multimodal fusion design

5. Results & Comparison — Tabular vs Tabular+Image performance

1. Overview & Approach

Problem Statement

Predicting real estate prices is a complex task influenced by both quantitative property attributes (bedrooms, square footage, condition) and qualitative environmental factors (neighborhood quality, proximity to amenities, greenery). Traditional tabular models capture structural features but miss visual context that significantly impacts property value. This project addresses this gap by developing a multimodal deep learning system that combines satellite imagery with tabular property data.

Dataset:

Split	Samples	Usage
Training	12,967	Model optimization
Validation	3,242	Hyperparameter tuning
Test	5,404	Final predictions
Total	21,613	—

Modelling Strategy

- **Multimodal Fusion:** Combine satellite imagery (224×224, zoom level 20) with 21 engineered tabular features through late fusion architecture.
- **Transfer Learning:** Use EfficientNet-V2-S pretrained on ImageNet with BatchNorm-only unfreezing (1.4% trainable parameters) for efficient domain adaptation.
- **Log-Price Prediction:** Transform highly skewed price distribution (skewness 4.03) to log-scale (skewness 0.41) for stable training.
- **Attention Mechanisms:** Apply CBAM (Convolutional Block Attention Module) to help the model focus on value-relevant image regions.
- **Multi-Scale Features:** Use FPN-Lite (Feature Pyramid Network) to fuse spatial information from multiple backbone stages.

2. Exploratory Data Analysis

2.1 Price Distribution

The target variable (property price) exhibits significant right-skewness, typical of real estate markets where a few luxury properties command prices far above the median. This distribution motivated our log-transformation approach for model training.

Statistic	Original Price	Log(Price)
Min	$75,000	11.23
Max	$7,700,000	15.86
Mean	$537,470	13.05
Median	$450,000	13.02
Std Dev	$360,304	0.53
Skewness	4.03	0.41

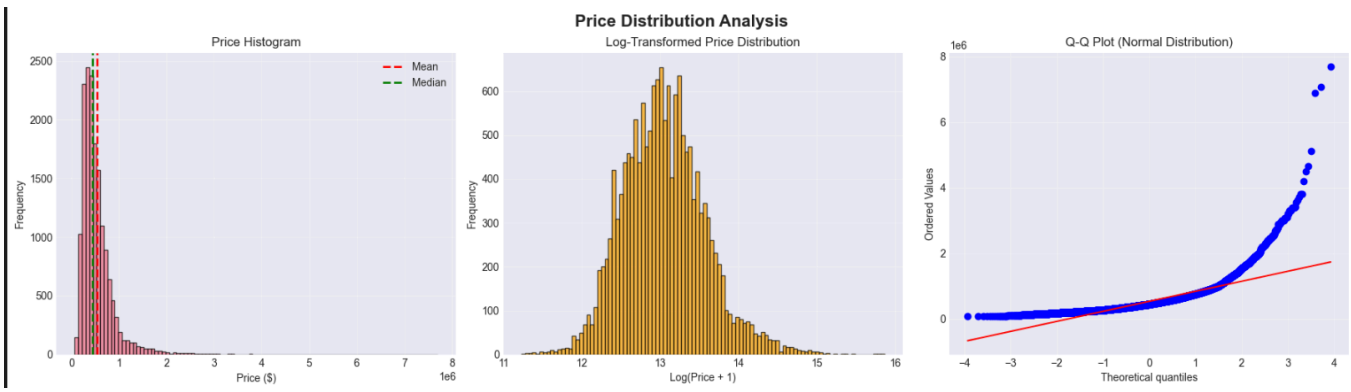


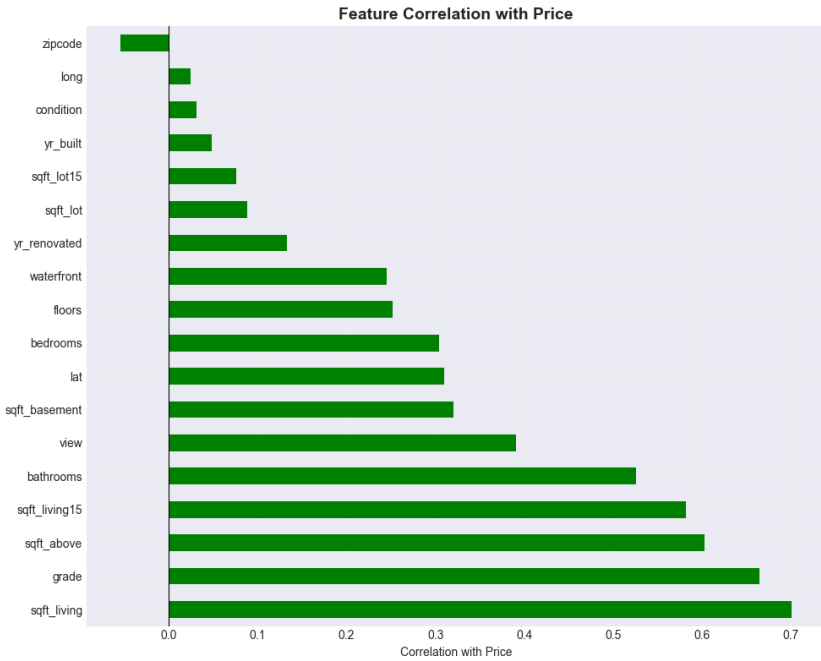
Table 1: Price distribution statistics showing effectiveness of log transformation

2.2 Key Feature Correlations with Price

Feature analysis revealed strong correlations between price and living space metrics, grade (construction quality), and premium features like waterfront access and view quality.(HIGHLY CORRELATED FEATURE PAIRS (|r| > 0.8) sqft_living <-> sqft_above : 0.876)

Feature	Correlation	Insight
sqft_living	0.70	Strongest predictor - larger homes command premium
grade	0.67	Construction quality significantly impacts value
sqft_above	0.61	Above-ground space more valuable than basement
sqft_living15	0.58	Neighbor home size indicates neighborhood quality
bathrooms	0.53	More bathrooms = higher value
view	0.40	Properties with views have substantial premium

waterfront	0.26	Waterfront adds ~$1M premium on average



correlationwith price

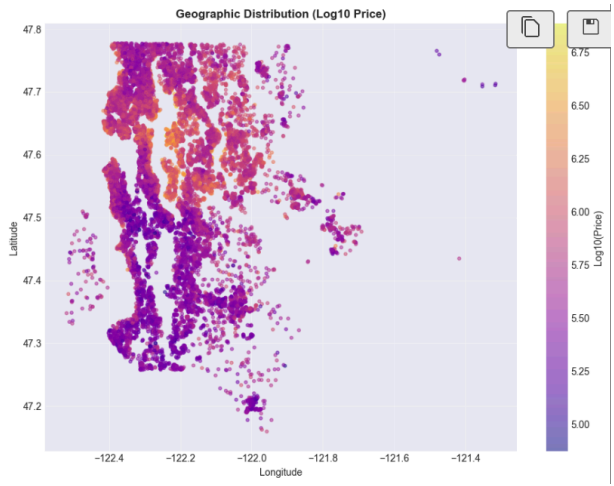
2.3 Categorical Feature Impact

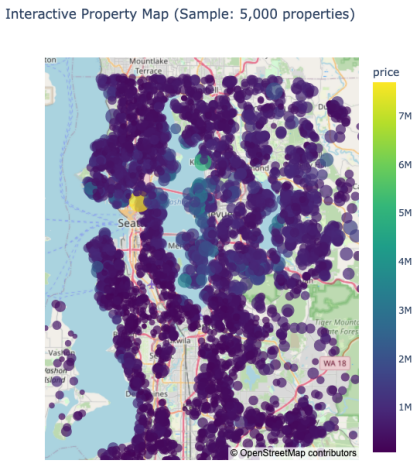
Feature	Category	Avg Price	Count
Waterfront	Yes (1)	$1,591,938	113 (0.7%)
Waterfront	No (0)	$530,068	16,096 (99.3%)

Grade	13 (Mansion)	$3,656,700	10
Grade	7 (Average)	$403,304	6,761
View	4 (Excellent)	$1,391,374	233
View	0 (None)	$494,968	14,604

Table 3: Impact of categorical features on property price

2.4 Geographical EDA

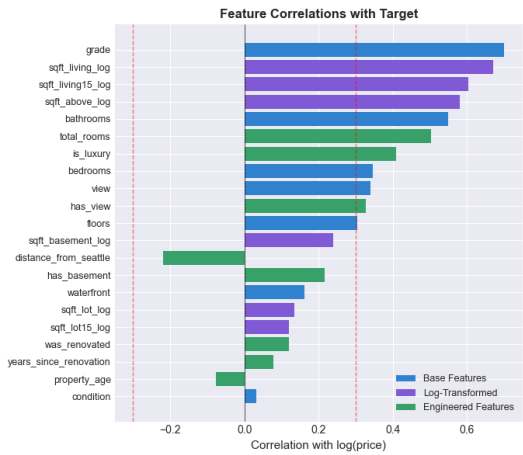


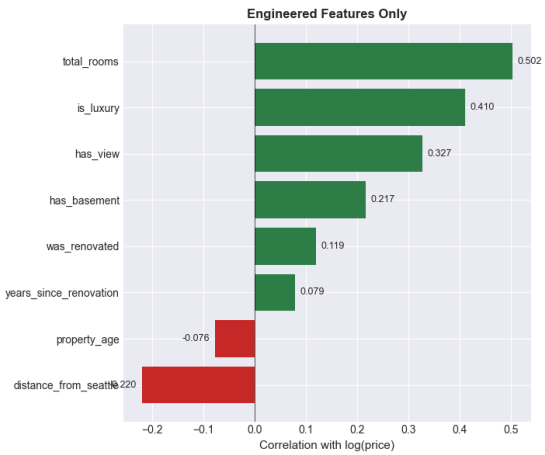


The waterfront corridor along Lake Washington (longitude ~-122.2) commands 4-5× price premiums ($2-2.5M average) despite low property density, while high-density urban Seattle areas show only moderate prices. This inverse density-price relationship demonstrates that location quality (waterfront, views) matters.

Final Feature Set (21 Features)- Feature Engineering

Base Features (7): bedrooms, bathrooms, floors, waterfront, view, condition, grade
Log-Transformed (6): sqft_living_log, sqft_lot_log, sqft_above_log, sqft_basement_log, sqft_living15_log, sqft_lot15_log (they were highly skewed features)
Engineered (8): property_age, was_renovated, years_since_renovation, has_basement, total_rooms, has_view, is_luxury, distance_from_seattle





3. Visual Feature Analysis

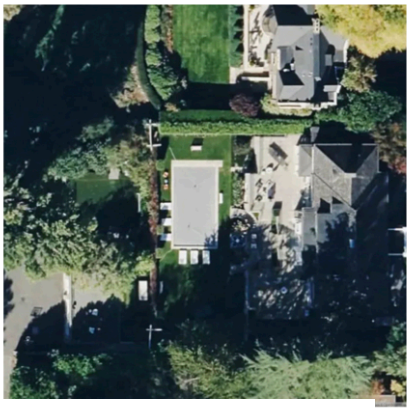
Satellite Imagery Configuration

Satellite images were acquired at zoom level 20 (approximately 0.3m/pixel resolution), providing detailed views of individual properties including roof characteristics, landscaping, pools, driveways, and surrounding neighbourhood features. Zoom -20 shows house level specific details while also maintaining visual information about the neighbourhood, while zoom-19 level captures a lot of neighbourhood specific details while fails to capture detailed house- specific view

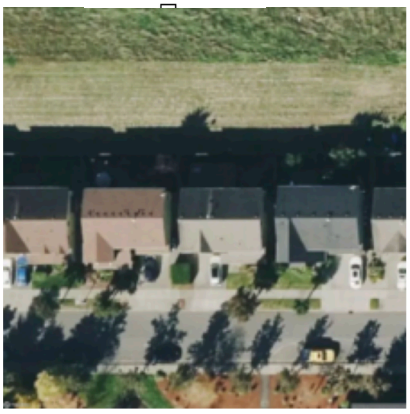
```
HIGH PRICE typically shows:
• Larger roof footprint
• Swimming pools
• Manicured landscaping
• Waterfront/lake access
• Larger lot size
• Complex roof structure
• Multiple structures

LOW PRICE typically shows:
• Smaller footprint
• Dense neighborhood
• Less greenery
• Standard lot size
• Simple roof structure
• Closer to neighbors
```

HIGH PRICE PROPERTY
$7,700,000



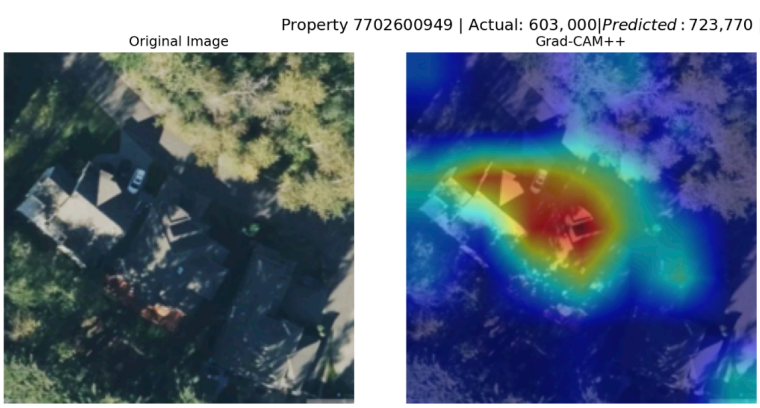
LOW PRICE PROPERTY
$275,000

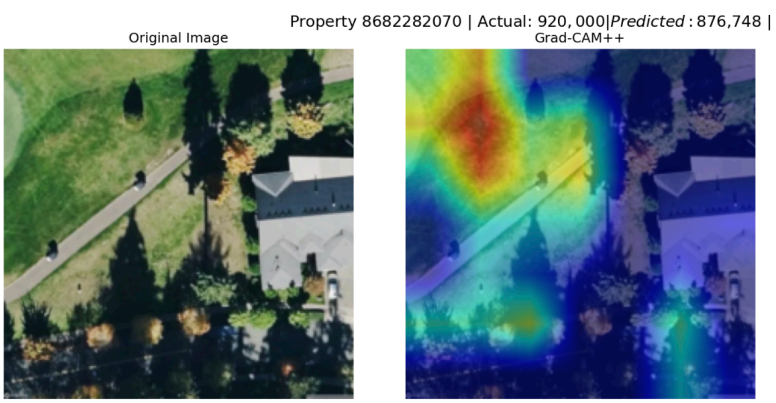


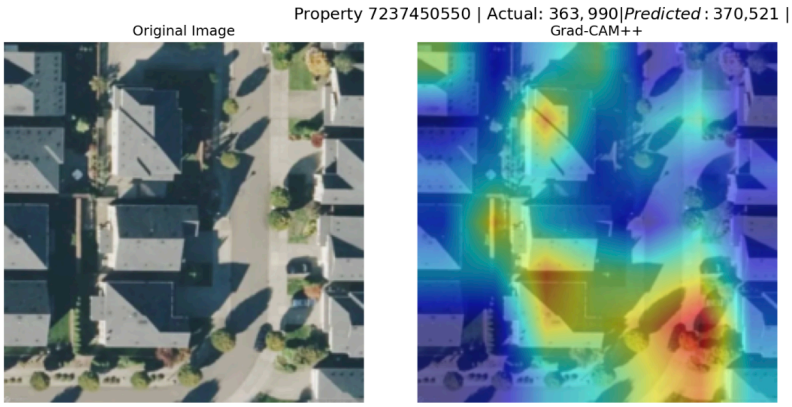
Driving Property Value (Grad-CAM++ Analysis)

Using Grad-CAM++ (Gradient-weighted Class Activation Mapping) on the FPN output layer, we analyzed which image regions most influenced the model's price predictions. The attention patterns reveal interpretable focus on value-relevant features:

- **Waterfront & Water Bodies:** Strong attention on lakes, rivers, and oceanfront areas. Waterfront properties average $1.59M vs $530K for non-waterfront, validating the model's focus.
- **Building Density & Neighbourhood:** The model attends to surrounding housing density and neighbourhood character, small houses in densely populated neighbourhoods tend to have a lower price , while stand-alone houses tend have a higher price.
- **Greenery & Landscaping:** Well-maintained spread out lawns, trees, and landscaping features receive attention as seen in the Grad-cam overlays, correlating with higher-grade properties and better-maintained homes.
- **Roof Size & Structure:** Larger roof footprints (indicating larger homes) and complex roof structures draw model attention.
- **Driveways & Access:** Paved driveways, multiple-car garages, and property access points are highlighted, indicating infrastructure quality.
- **Pools & Luxury Amenities:** Swimming pools and outdoor amenities are strongly attended to, consistent with their presence in luxury properties.







4. Model Architecture

The architecture follows a late fusion strategy where image and tabular branches are processed independently before being combined in the fusion head. This design allows each modality to develop specialized representations before integration.The multimodal model is designed so that visual features contribute complementary context rather than overwhelming the tabular signal.

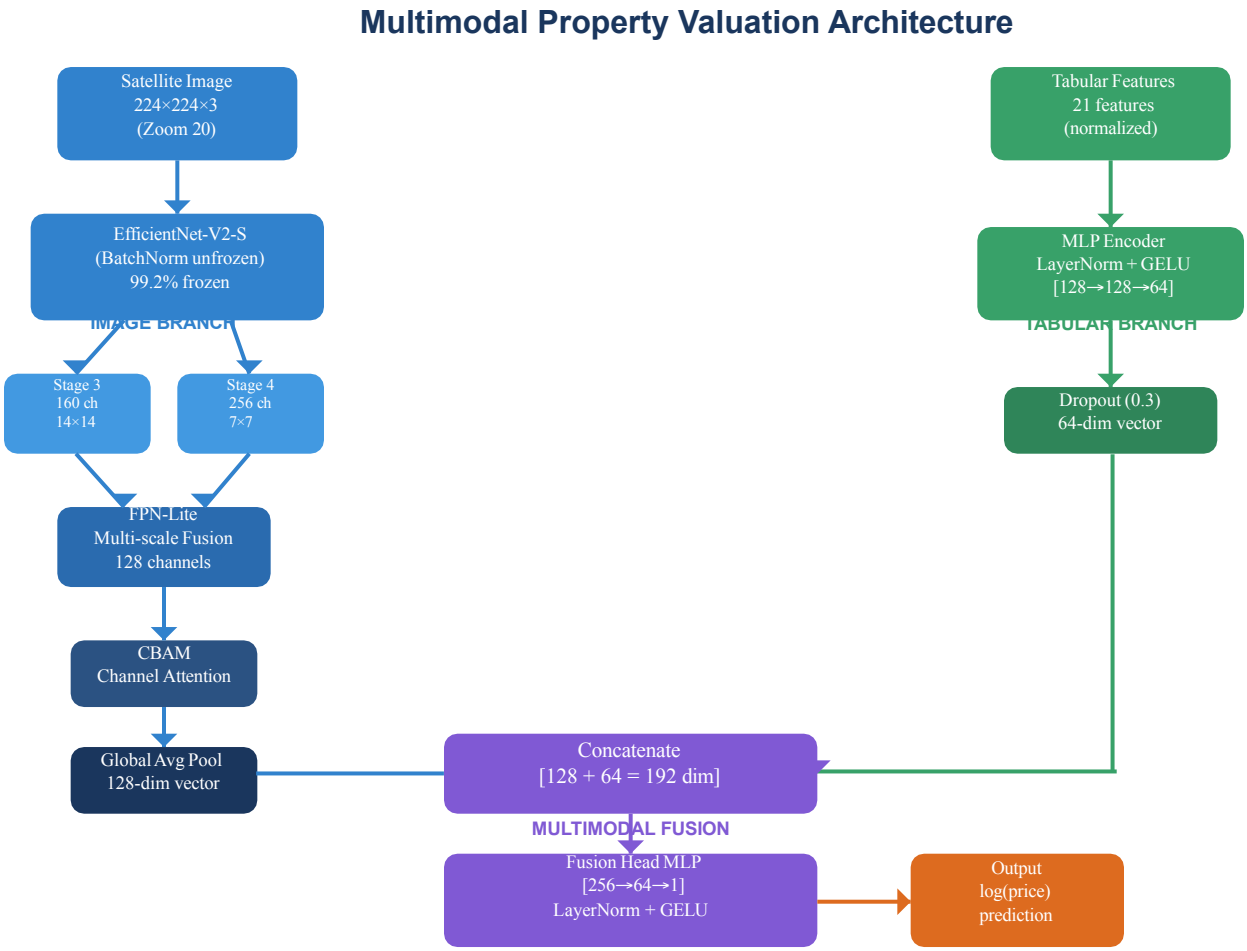


Figure 1: Multimodal Property Valuation Architecture

Component Details

Component	Configuration
EfficientNet-V2-S	Pretrained, BN unfrozen (1.4%)
FPN-Lite	Lateral + upsample + add
CBAM	Channel attention (reduction=16)
Global Avg Pool	Spatial pooling
Tabular MLP	[21→128→128→64], LayerNorm+GELU
Fusion Head	[192→256→64→1], LayerNorm+GELU

Table 4: Architecture component specifications

Training Configuration

Parameter	Value
Optimizer	AdamW (weight_decay=1e-5)
Learning Rate	3e-4 with StepLR (γ=0.5, step=8)
Loss Function	Huber Loss (δ=1.0)
Batch Size	64
Epochs	25 (early stopping patience=8)
Total Parameters	19,980,817
Trainable Parameters	284,881 (1.4%)

Training and model selection are performed entirely in log space to ensure numerical stability.Huber loss is used for predicting log(price)

5. Results & Comparison

Validation Set Performance

Metric	Log-Scale	Real-Scale
RMSE	0.2184	$143,495
R^2	0.8197	0.8182
MAE	—	$88,874
MAPE	—	17.28%

Table 5: Final validation metrics for multimodal model (Approach 4)

Comparison: Tabular Only vs. Tabular + Satellite Images

To evaluate the contribution of satellite imagery, we compare our multimodal approach (Approach 4) against a tabular-only ensemble (Approach 1) that uses XGBoost, LightGBM, and CatBoost.

Approach	Method	Val RMSE	Val R^2	Notes
Approach 1	Tabular Ensemble (XGB+LGBM+CatBoost)	$98,019	0.9152	Best metrics but ignores images
Approach 4	Multimodal (EfficientNet+Tabular)	$143,495	0.8182	True multimodal fusion

Table 6: Comparison of tabular-only vs. multimodal approaches

Analysis of Results

The tabular ensemble (Approach 1) achieves higher numerical metrics, but with a critical caveat: the adaptive weighting learned to assign 100% weight to tabular predictions and 0% to image predictions, effectively

ignoring the satellite imagery entirely. This occurs because tabular features are strong predictors, and the model takes the 'shortcut' of relying solely on them.

In contrast, Approach 4 forces both modalities to contribute through architectural design (concatenation-based fusion). While the metrics are lower, this represents genuine multimodal learning where visual features complement tabular data. The Grad-CAM visualizations confirm the model learns meaningful visual patterns.

Key Takeaways

- Tabular features (especially sqft_living, grade) remain the strongest predictors of property value.
- Satellite imagery provides complementary information about neighborhood quality, amenities, and visual condition.
- BatchNorm-only unfreezing enables efficient domain adaptation with minimal trainable parameters (1.4%).
- Log-price transformation is essential for stable training on skewed real estate data.