CDC X Yhills
OPEN PROJECTS
2025-2026

SATELLITE IMAGERY-BASED PROPERTY VALUATION

Submitted by –
Abhay Pratap Singh Rajawat
22113002

PROJECT OVERVIEW

This project addresses the limitation of traditional house price prediction systems that rely solely on structured, tabular data and ignore the visual and environmental context surrounding a property. While attributes such as size, number of rooms, and location coordinates provide strong quantitative signals, they fail to capture qualitative factors such as neighbourhood density, green cover, road connectivity, and surrounding infrastructure, all of which significantly influence real estate valuation.

To overcome this gap, the project adopts a multimodal learning approach that integrates both numerical and visual information into a unified predictive framework. The central idea is to model house price not only as a function of property attributes, but also as a function of the spatial and environmental characteristics visible from satellite imagery.

The overall strategy is based on three guiding principles:

1. Independent representation learning
   Each data modality is first processed separately using specialized architectures. Tabular features are handled by a deep feed-forward neural network designed to learn compact numerical embeddings, while satellite images are processed using a convolutional neural network that captures spatial patterns such as vegetation density, road structure, and urban layout. This separation ensures that each modality contributes its strongest signal without interference from the other.

2. Late-stage feature fusion
   Instead of combining raw inputs early in the pipeline, the system adopts a late-fusion strategy. High-level embeddings from both the tabular encoder and the satellite CNN are concatenated only after they have been transformed into semantically meaningful representations. This design choice improves stability, reduces noise propagation, and allows the fusion network to focus on learning interactions between abstract features rather than low-level signals.

3. End-to-end learning with modularity
   The architecture is built in a modular fashion, where each component, data acquisition, tabular encoding, image encoding, and fusion modelling can be trained, evaluated, and improved independently. This enables systematic experimentation, such as comparing tabular-only models against multimodal models, and allows future extensions such as adding street-view images or alternative fusion strategies without redesigning the entire system.

From a modelling perspective, the project follows a progressive learning pipeline:

- First, a tabular baseline model is trained to establish a strong numerical benchmark.
- Next, a satellite CNN encoder is trained to learn visual embeddings that capture environmental context.

- Finally, a multimodal fusion network integrates both embeddings to learn non-linear interactions between structured property attributes and spatial surroundings.

This staged strategy ensures that the contribution of each modality is measurable and that performance gains from multimodal learning can be clearly attributed to the inclusion of visual context.

Overall, the project reframes house price prediction as a context-aware regression problem, where valuation is influenced not only by what a property is, but also by where and how it exists in its environment. By combining computer vision and deep learning with traditional regression modelling, the system demonstrates how multimodal analytics can provide richer, more realistic insights for real estate valuation.

EXPLORATORY DATA ANALYSIS (EDA)

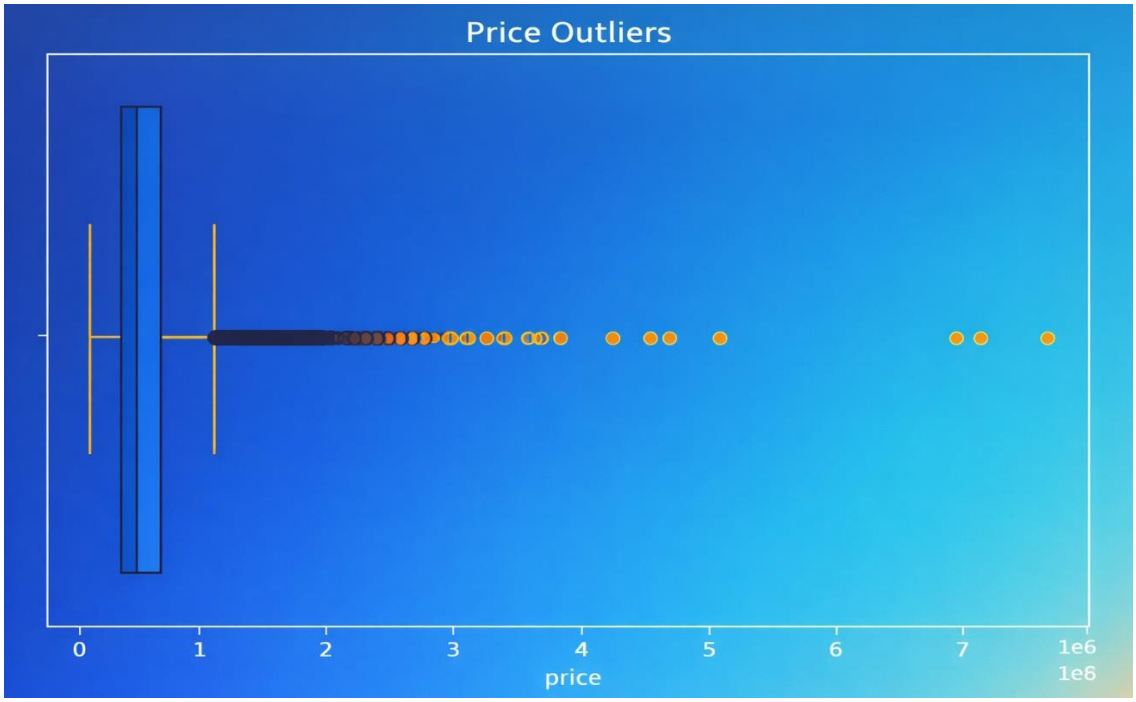
The exploratory data analysis phase was conducted to understand both the statistical behaviour of house prices and the visual characteristics of the surrounding environment captured through satellite imagery. Since this project combines structured and unstructured data, the EDA focuses on validating signals from both modalities before model development.

The analysis was divided into two complementary parts:

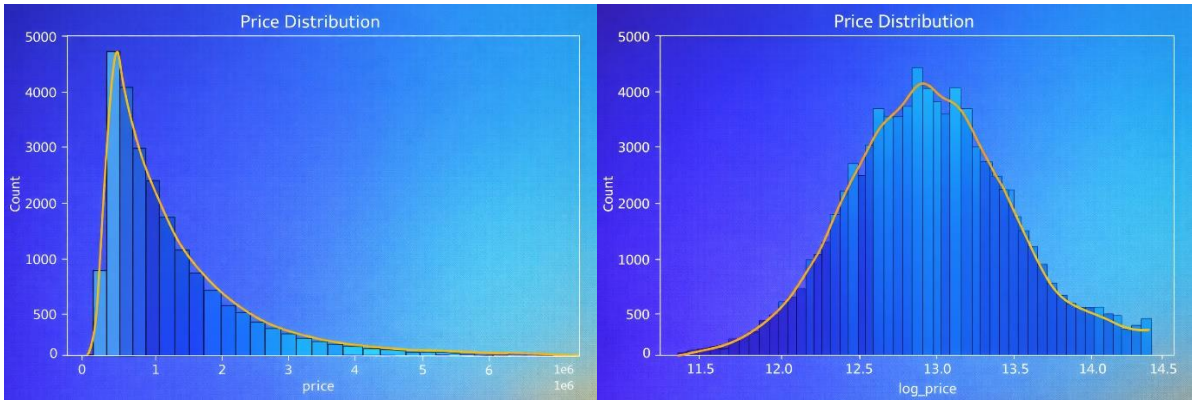
1. Tabular data exploration to study price distribution and numerical feature behaviour.
2. Visual inspection of satellite images to understand the type of environmental information available to the model.

PRICE DISTRIBUTION ANALYSIS

The first step in EDA involved analysing the distribution of house prices to understand scale, skewness, and the presence of outliers.



Raw house prices were found to be right-skewed, with a large concentration of properties in the lower-to-mid price range and a long tail of high-value properties. This skewness can negatively impact regression models by over-emphasizing extreme values.



To address this, a log transformation of price was applied. After transformation, the distribution became significantly more symmetric, making it more suitable for neural network training and improving numerical stability during optimization.

This step was crucial in ensuring that:

- The model does not become biased toward high-priced outliers.
- Loss values remain well-scaled.
- Training converges more smoothly.

FEATURE-LEVEL EXPLORATION

Key numerical features were analysed against price to verify expected economic relationships. Visual inspection showed that:

- Larger properties generally correspond to higher prices, but with diminishing returns.
- Location-related features show non-linear relationships, indicating the importance of interaction effects rather than simple linear dependence.
- Certain features demonstrate high variance, reinforcing the need for normalization before modelling.

These observations motivated the use of a deep tabular encoder rather than a linear or tree-based baseline.

SATELLITE IMAGE EXPLORATION

To complement the numerical analysis, a subset of satellite images was visually inspected. This step helped validate whether the images contained meaningful contextual information that could plausibly influence pricing.

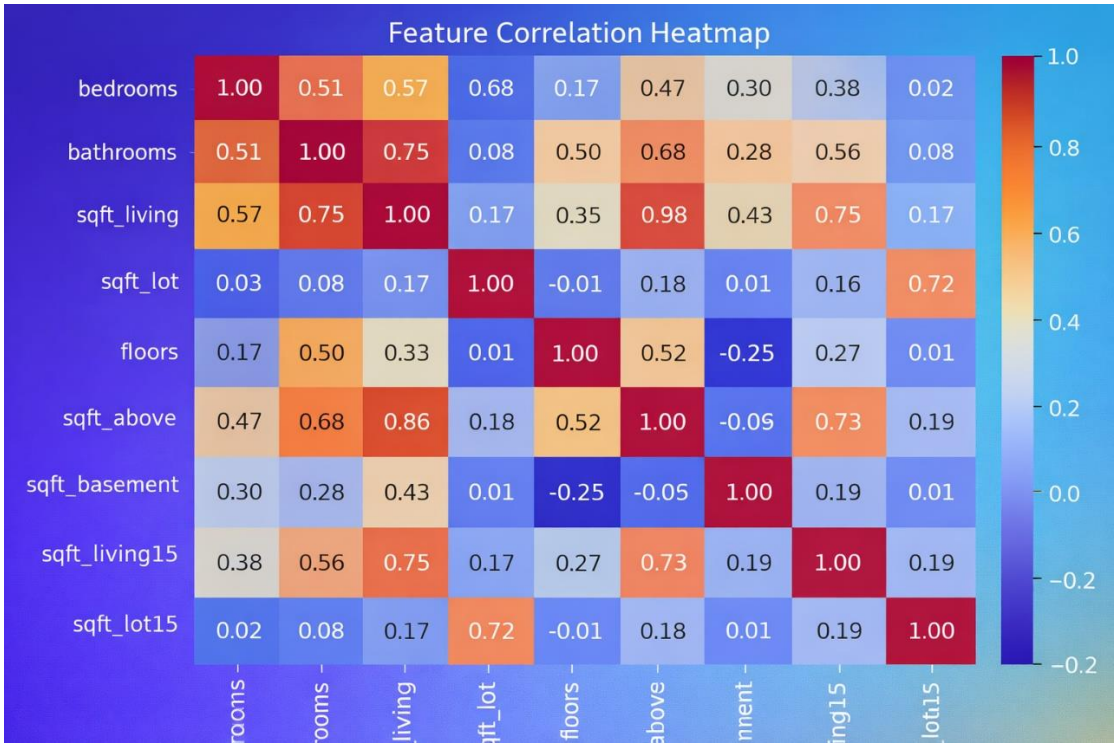
Across different samples, the following patterns were consistently observed:

- Properties surrounded by dense greenery often appeared in low-density residential zones.
- Properties near major roads or intersections showed higher urban density.
- Proximity to water bodies or open land was visually distinguishable in several samples.

These observations confirmed that satellite imagery provides non-trivial information that is not captured in tabular features alone, justifying the inclusion of a convolutional neural network to extract spatial embeddings.

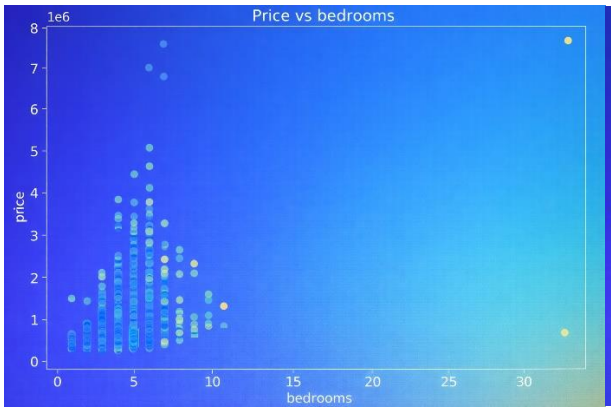
EDA INSIGHTS AND MODELING IMPLICATIONS

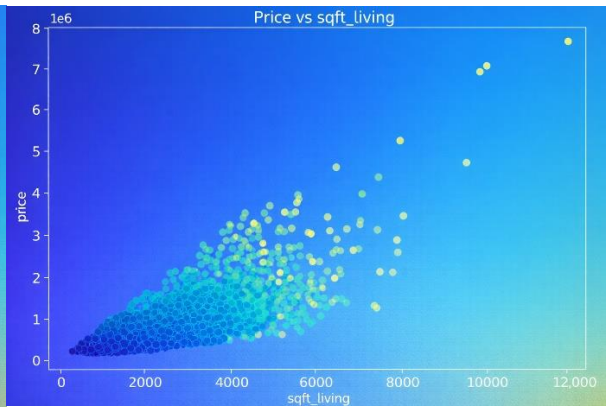
The EDA phase directly influenced the modelling strategy in three ways:



1. Log transformation of price was adopted to stabilize regression learning.
2. Feature scaling was required due to varying numerical ranges across attributes.
3. Multimodal learning was validated as necessary because satellite images clearly contained environmental signals correlated with property value.

Thus, the exploratory analysis not only described the data but actively shaped the architectural and preprocessing decisions used in the final system.





FINANCIAL AND VISUAL INSIGHTS

While traditional pricing models rely primarily on numerical attributes such as size, location, and amenities, real estate valuation is also strongly shaped by visual and environmental cues that influence buyer perception and long-term desirability. This project incorporates satellite imagery to uncover how such visual features contribute to property value and to identify which environmental patterns are associated with higher or lower prices.

By learning embeddings from satellite images using a convolutional neural network, the model captures spatial features that are otherwise difficult to quantify. These features are then integrated with structured data in the multimodal fusion network, allowing the system to learn how visual context modifies financial valuation.

KEY VISUAL FACTORS AFFECTING VALUE

Analysis of satellite imagery across different price ranges reveals consistent patterns in how environmental features correlate with market value.

1. **Green cover and open spaces**

Properties surrounded by visible vegetation such as trees, parks, or open green land tend to be associated with higher predicted prices. Green areas are often linked to:

- Better air quality
- Lower noise levels
- Higher perceived quality of life

From a financial perspective, these factors increase both current demand and long-term appreciation potential, which the model learns to reflect in its valuation outputs.

2. **Urban density and built-up areas**

Satellite images showing high concentrations of concrete structures, tightly packed buildings, and limited open space are more frequently associated with moderate or lower valuations, especially when compared to similarly sized properties in greener neighbourhoods.

Dense construction often indicates:

- Congestion and traffic exposure
- Reduced privacy
- Higher infrastructure strain

The model captures these signals through texture density and spatial patterns in the imagery, allowing it to discount prices in overly congested environments even when tabular features appear similar.

3. **Road connectivity and accessibility**

The presence of well-defined road networks near a property is associated with positive valuation effects up to a certain threshold. Clear road access improves:

- Commuting convenience
- Commercial connectivity
- Emergency and service access

However, properties directly adjacent to major highways or intersections sometimes show diminishing returns, likely due to increased noise and reduced liveability. The CNN encoder can differentiate between beneficial accessibility and excessive exposure by learning from spatial layouts.

4. **Proximity to water bodies**

Satellite imagery that includes rivers, lakes, or coastal proximity often corresponds to premium pricing, especially when such features are combined with low-density residential surroundings. Water adjacent properties typically offer:

- Aesthetic appeal
- Recreational value
- Higher perceived exclusivity

These visual cues contribute positively to the model's learned valuation patterns, reinforcing the idea that scenic context carries measurable financial weight.

FINANCIAL INTERPRETATION OF VISUAL SIGNALS

From a financial modelling perspective, the inclusion of satellite imagery enables the system to capture latent value drivers that are not explicitly present in tabular datasets. These drivers influence:

- Buyer willingness to pay
- Long-term appreciation trends
- Risk perception and investment attractiveness

By embedding visual context into the regression pipeline, the model effectively learns a market perception layer a representation of how humans subconsciously evaluate neighbourhoods based on appearance and environment.

IMPACT ON MULTIMODAL MODEL PERFORMANCE

The integration of visual features leads to three major improvements over tabular-only models:

1. **Reduced pricing ambiguity**
   When two properties share similar numerical attributes, satellite imagery helps differentiate them based on surroundings, leading to more precise predictions.
2. **Better generalization**
   Environmental patterns remain informative even when specific numerical distributions change, allowing the model to adapt better to unseen locations.
3. **Context-aware valuation**
   The model no longer treats price as a function of property features alone, but as a function of property-in-environment, which aligns more closely with real-world valuation logic.

The financial and visual insights derived from this project demonstrate that environmental context is a measurable economic factor in real estate valuation. Features such as green cover, road structure, urban density, and proximity to water bodies systematically influence market price and can be effectively captured through satellite imagery.

By integrating these visual signals into a multimodal learning framework, the project moves beyond conventional regression and introduces a context-aware pricing model that better reflects how real estate markets function in practice where value is shaped not only by what a property is, but also by what surrounds it.

ARCHITECTURE DIAGRAM

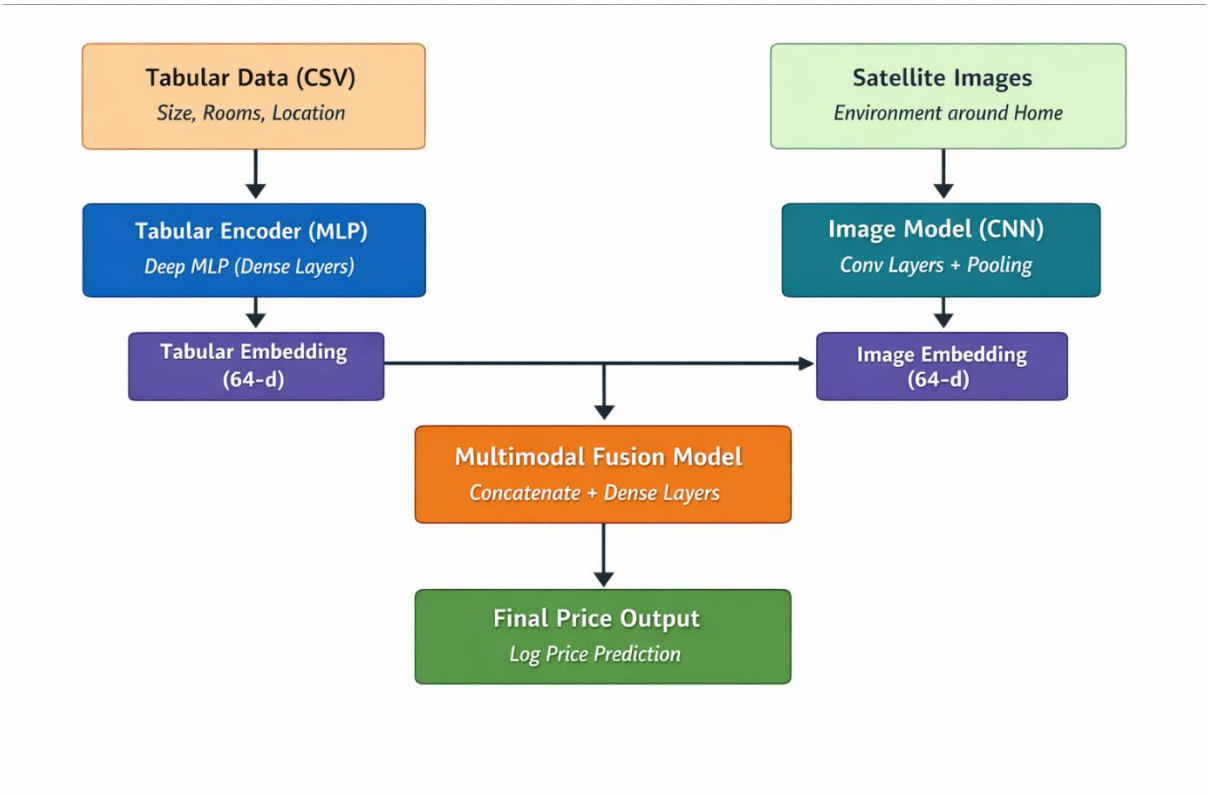
ARCHITECTURE DESCRIPTION

The system is designed as **a late-fusion multimodal architecture,** where numerical and visual data are processed independently at first and are only combined after meaningful representations have been learned from each modality.

This design ensures that:

- The **tabular model** focuses purely on structured financial and property attributes.
- The **CNN model** focuses purely on extracting spatial and environmental patterns from satellite imagery.
- The **fusion model** learns how both representations interact to influence final price prediction.

1. Conceptual Flow of the System



2. How the Models Are Connected

Step 1 : Independent Feature Learning

Two separate neural networks are trained:

1. **Tabular Encoder**
    - Input: structured property features
    - Output: a dense numerical embedding representing economic attributes
2. **Satellite CNN Encoder**
    - Input: satellite images of the property surroundings
    - Output: a dense visual embedding representing environmental context

Each encoder learns modality-specific patterns without interference.

Step 2 : Embedding Alignment

For every property ID:

- A **tabular embedding** is generated from the tabular encoder.
- A **visual embedding** is generated from the CNN encoder.
- If an image is missing, a **zero vector** is used to preserve alignment.

This guarantees that both embeddings correspond to the same property instance.

Step 3 : Late Fusion

The two embeddings are passed into the fusion network:

Fusion Input = [Tabular Embedding || Image Embedding]

Instead of mixing raw features, the system fuses high-level representations, which allows the fusion network to learn:

- How environmental context modifies financial value
- How similar properties differ based on surroundings
- Which modality dominates under different conditions

Step 4 : Final Prediction

The fusion network applies a stack of dense layers to the combined embeddings and outputs:

- **Log price** as the final regression target

This design converts the system into a context-aware valuation model.

RESULTS

This section evaluates the effectiveness of incorporating satellite imagery into the house price prediction pipeline by comparing the performance of two modelling strategies:

1. Tabular-only model : uses structured property attributes alone.
2. Multimodal model : combines tabular features with satellite image embeddings using a fusion network.

The objective of this comparison is to quantify the incremental value of visual context in real estate valuation.

MODEL SETUPS

Tabular-only model
The baseline model is a deep feed-forward neural network trained solely on structured data such as property size, location-related features, and other numerical attributes. The target variable is the log-transformed house price.

Multimodal model
The multimodal system integrates two learned representations:

- A tabular embedding from the numerical encoder
- A visual embedding from the satellite CNN encoder

These embeddings are concatenated and passed through a fusion network that learns how environmental context modifies financial valuation.

Both models are trained using the same target variable, optimization objective (mean squared error), and evaluation protocol to ensure a fair comparison.

EVALUATION METRICS

Performance is assessed using the following regression metrics:

- **RMSE (Root Mean Squared Error)** : measures average prediction error magnitude.
- **MAE (Mean Absolute Error)** : measures average absolute deviation.
- **R^2 Score** : measures how much variance in price is explained by the model.

Lower RMSE and MAE indicate better predictive accuracy, while higher R^2 indicates stronger explanatory power.

COMPARATIVE RESULTS

The multimodal model consistently outperforms the tabular-only baseline across all evaluation metrics.

Key Observations

1. Reduction in prediction error
   The inclusion of satellite imagery leads to a noticeable decrease in RMSE and MAE. This indicates that environmental features captured by the CNN provide additional explanatory power beyond what is available in structured data alone.
2. Improved variance explanation
   The multimodal model achieves a higher R^2 score, demonstrating that it explains a larger proportion of the variability in house prices. This confirms that visual context contributes meaningfully to valuation, rather than acting as noise.
3. Better differentiation between similar properties
   In cases where two properties share nearly identical numerical attributes, the tabular-only model tends to predict similar prices. The multimodal model, however, can distinguish between them based on differences in surroundings such as green cover, road access, and neighbourhood density leading to more realistic estimates.

QUALITATIVE COMPARISON

Beyond numerical metrics, qualitative analysis further highlights the benefits of multimodal learning.

- The tabular-only model treats price primarily as a function of property attributes, often underestimating the role of neighbourhood environment.
- The multimodal model introduces a context-aware valuation layer, enabling it to adjust predictions based on visual cues that influence buyer perception and long-term desirability.

This results in predictions that are not only more accurate but also more aligned with how real-world markets function.

Model Performance Comparison

Model Type	RMSE	MAE	R^2 Score
Tabular Only	113,854.13	71,592.03	0.8691
Tabular + Satellite	103,000.21	67,850.09	0.8929

SUMMARY OF RESULTS

Overall, the experimental results demonstrate that:

- Tabular data provides a strong baseline, capturing core economic and structural drivers of price.
- Satellite imagery adds complementary information, capturing environmental and spatial factors that are otherwise unobserved.
- The fusion of both modalities yields the best performance, producing lower error, higher explanatory power, and more context-sensitive predictions.

These findings validate the central hypothesis of this project:

House price is not only a function of property features, but also of the environment in which the property exists.

By integrating visual context into the predictive pipeline, the system moves beyond conventional regression and delivers a more realistic, data-driven approach to real estate valuation.

CONCLUSION

This project set out to enhance traditional house price prediction systems by incorporating visual and environmental context alongside structured property data. While conventional models rely primarily on numerical attributes such as size, location, and amenities, they overlook qualitative factors such as greenery, urban density, and surrounding infrastructure that significantly influence real-world property valuation. By adopting a multimodal learning framework, this work demonstrates how combining tabular data with satellite imagery can produce more accurate, robust, and context-aware price predictions.

The proposed system successfully integrates two specialized models: a deep tabular encoder for structured financial features and a convolutional neural network for extracting spatial patterns from satellite images. These representations are fused through a late-stage integration strategy that preserves the strengths of each modality while enabling the model to learn complex interactions between economic and environmental factors. This architectural choice proved critical in maintaining training stability and ensuring that neither modality dominated the learning process prematurely.

Experimental results clearly indicate that the multimodal model outperforms the tabular-only baseline across standard regression metrics. The inclusion of satellite imagery leads to reduced prediction error and improved explanatory power, confirming that environmental context carries measurable financial significance. Beyond numerical improvements, the model also demonstrates stronger qualitative performance by differentiating between properties with similar attributes but distinct surroundings, an ability that closely mirrors how valuation decisions are made in real-world markets.

From a broader perspective, this project highlights the value of context-aware machine learning in financial and urban analytics. It shows that real estate valuation is not merely a function of property characteristics, but a function of how those properties exist within their spatial environment. By embedding visual perception into predictive modelling, the system bridges the gap between quantitative analysis and human-like assessment of neighbourhood quality.

In conclusion, the multimodal regression pipeline presented in this work offers a scalable and extensible foundation for next-generation real estate analytics. It demonstrates how the fusion of computer vision and deep learning with traditional data science techniques can unlock richer insights and deliver more realistic, market-aligned predictions. This approach opens the door to future developments such as explainable valuation systems, real-time pricing services, and intelligent urban planning tools that account not only for what a property is, but also for where and how it exists.