GeoSmart Price AI: A Multimodal Deep Learning and Geospatial Fusion Framework for Real Estate Valuation with Explainable AI

Mrs Roopa R[1], B Joshitha[2], K Sandhya[3], Nisha patil[4], Vidhya B[5]
[1] *Assistant Professor Department of CSE (Data Science) Vemana Institute of Technology Bengaluru, India*
[2,3,4,5] *Student Department of CSE (Data Science) Vemana Institute of Technology Bengaluru, India*

***Abstract*—GeoSmart Price AI is a multimodal real-estate valu-ation platform that combines structural property data, neighbor-hood amenities, listing text, and property photographs in a 544-dimensional representation. An ImageNet-pretrained ResNet-18 extracts visual features, XGBoost estimates market value, and TreeSHAP provides feature-level explanations. On a fixed 80/20 split of 600 Bengaluru properties, the full model reports a test MAE of INR 50.08 lakh, compared with INR 62.39 lakh for the structured baseline, and an R^2 of 0.6468. The web platform also supports property search, GIS visualization, financial tools, and a voice assistant. These results are limited to one hold-out split and require broader validation before high-stakes use.**

***Index Terms*—Multimodal learning, real estate valuation, geospatial intelligence, computer vision, ResNet-18, XGBoost, explainable AI, TreeSHAP, urban computing.**

I. INTRODUCTION

Automated valuation models are used by property market-places, financial institutions, and urban analysts to estimate residential property values. Real estate value is shaped by more than a home's recorded area and room counts: it also reflects accessibility to services, the quality of a listing's description, and the physical condition conveyed by photographs. These influences are heterogeneous and potentially nonlinear, making residential valuation a challenging prediction problem.

Many conventional AVMs use hedonic regression or tree-based learning on structured records. Such models can rep-resent broad locality differences through categorical labels, but may not resolve local variation in nearby amenities. They also cannot directly use the architectural and renovation details expressed in listing narratives or the appearance of a property in its photographs. In addition, complex ensembles can be difficult for users to interpret, which is a concern when estimates inform consequential financial decisions.

This work introduces GeoSmart Price AI, a property valua-tion framework that fuses structural, geospatial, textual, and vi-sual signals. The fused representation is passed to an XGBoost regression model. TreeSHAP is used to explain how input features contribute to individual predictions, with absolute attributions also aggregated into modality-level summaries. The valuation pipeline is part of a deployed web application that offers property discovery and GIS-based locality analysis.

The contributions of this work are as follows:

- A 544-dimensional early-fusion representation combining six physical and locality features, ten POI counts, sixteen TF–IDF features, and 512 ResNet-18 image embeddings.
- A permutation-invariant mean-pooling strategy that maps a variable-size gallery of property photographs to one fixed-size visual vector.
- A four-stage ablation using identical train/test partitions to report the marginal performance changes associated with geospatial, textual, and visual features.
- A TreeSHAP-based explanation method for examining feature-level and grouped modality contributions.
- Integration of the valuation approach with a GIS-enabled real-estate application, including listing search, authentication, financial calculations, and a

conversational voice interface.

II. RELATED WORK

[1] Artificial Intelligence and Real Estate Valuation: The Design and Implementation of a Multimodal Model—Karanikolas et al., 2025. This work presents a multi-modal AI model for real estate valuation by combining different types of property-related data to improve val-uation accuracy. Advantages: Combines multiple data sources for better property assessment. Disadvantages: Depends on data quality and computational resources.

[2] Machine Learning for Real Estate Investment Risk As-sessment: Developing Predictive Models for the US Market—Gerald and Jeremy, 2024. This work explores machine learning techniques for assessing real estate investment risks and supporting data-driven investment decisions in the US market. Advantages: Helps identify investment risks and predict market trends. Disadvan-tages: Mainly focuses on the US market and depends on historical data.

[3] Automated Land Valuation Models: A Comparative Study of Four Machine Learning and Deep Learning Methods Based on a Comprehensive Range of Influential Factors—Jafary et al., 2024. This work compares four machine learning and deep learning methods for auto-mated land valuation using various factors that influence land prices. Advantages: Compares different prediction methods and considers multiple valuation factors. Dis-advantages: Performance depends on data quality and regional market conditions.

[4] Multimodal Machine Learning for Real Estate Ap-praisal: A Comprehensive Survey—Huang et al., 2025. This work reviews multimodal machine learning tech-niques for real estate appraisal, including the integration of images, textual descriptions, and numerical property data. Advantages: Provides an overview of multimodal techniques and research challenges. Disadvantages: Fo-cuses on reviewing existing methods rather than devel-oping a new model.

[5] *Vision*-Based *Housing Price Estimation Using Interior, Exterior and Satellite Images*—Nouriani and Lemke, 2022. This work uses interior, exterior, and satellite images to estimate housing prices based on property appearance and surrounding environmental features. Ad-vantages: Uses visual information to improve property price estimation. Disadvantages: Requires suitable im-ages and may overlook economic factors.

[6] *NLP: Text Summarization and Keyword Extraction on Property Rental Listings—Part 1*—Kristiyanto, 2024. This work explores natural language processing tech-niques to summarize rental listings and extract impor-tant keywords from property descriptions. Advantages: Simplifies property descriptions and helps users find relevant information quickly. Disadvantages: May miss important details when processing complex text.

[7] *Precision in Real Estate: Hybrid Multi-Stage* Predictions *for House Prices*—Manasa et al., 2025. This work presents a hybrid, multi-stage prediction approach for estimating house prices using machine learning tech-niques to improve prediction performance. Advantages: Combines multiple prediction stages to improve estima-tion. Disadvantages: Increases model complexity and computational requirements.

[8] *Machine Learning Algorithms and Explainable Artifi*-cial *Intelligence for Property Valuation*—Maselli and Nesticò, 2025. This work examines machine learning algorithms and explainable AI techniques for property valuation, focusing on prediction accuracy and under-standing the factors influencing estimated prices. Ad-vantages: Improves prediction transparency and identi-fies influential property features. Disadvantages: Some models remain difficult to interpret fully.

[9] *A Novel Model for House Price Prediction with Machine Learning Techniques*—Belsare and Warkar, 2023. This work presents a machine learning-based approach for predicting house prices using relevant property features and historical data. Advantages: Automates house price prediction and supports data-driven valuation. Disad-vantages: Accuracy depends on the quality and rele-vance of the dataset.

[10] *Predicting Future Housing Prices: A Machine Learning* Approach—Verma et al., 2023. This work explores ma-chine learning techniques for

forecasting future housing prices using historical housing data and identified market patterns. Advantages: Helps forecast housing prices and supports investment decisions. Disadvantages: Sudden economic changes may reduce prediction accuracy.

III. PROPOSED FRAMEWORK

A. Overview

GeoSmart Price AI uses early feature fusion. Each modality is preprocessed into a fixed-dimensional vector; the vectors are concatenated and supplied to a gradient-boosted tree regressor. Fig. 1 summarizes the processing path.

The four vectors correspond to complementary evidence about the same property: its physical characteristics, nearby services, listing language, and photographed appearance. The structured and geospatial components represent measurable attributes and local context, while TF–IDF and image embed-dings encode unstructured listing content. Keeping these rep-resentations separate during preprocessing makes their dimen-sions and sources explicit; concatenation then gives XGBoost one aligned feature row per property. The staged ablation evaluates this design by introducing GIS, NLP, and visual features successively while retaining the same evaluation split.

B. Structured Physical Attributes

The structural vector $\mathbf{v}_{\text{struct}} \in \mathrm{R}^6$ contains property area, bedroom count, bathroom count, balcony count, locality rate, and building age:

$$\mathbf{v}\text{struct} = [x\text{area}, x\text{bhk}, x\text{bath}, x\text{balc}, x\text{loc_rate}, x\text{age}]^T. \qquad (1)$$

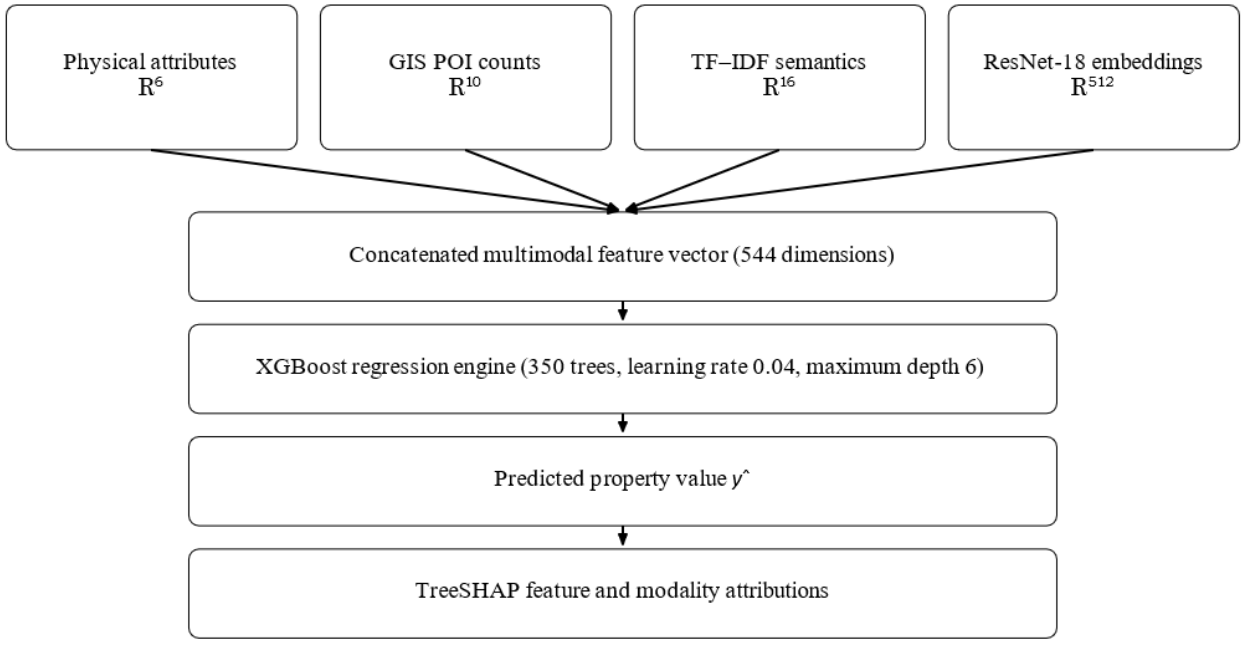

Fig. 1. GeoSmart Price AI multimodal feature-fusion and explanation pipeline.

Area is recorded in square feet. The locality rate is a median price-per-square-foot target encoding calculated from training data only:

$$\hat{\mu}_L = \text{median}\left\{ \frac{y_j}{A_j} : L_j = L,\ j \in \mathcal{D}_{\text{train}} \right\}, \qquad (2)$$

where y_j, A_j, and L_j denote the sale price, area, and locality of training observation j, respectively. Building age is mea-sured in calendar years.

C. Geospatial Infrastructure Features

For a property location $p = (\text{lat}_p, \text{lon}_p)$, the system obtains POI counts within radius $R = 1000$ m. The ten features represent schools, colleges, hospitals, metro stations, bus stops, restaurants, malls, banks, parks, and the total POI count:

$$\mathbf{v}_{\text{geo}} = [c_{\text{schools}}, c_{\text{colleges}}, c_{\text{hospitals}}, c_{\text{metro}}, c_{\text{bus}}, c_{\text{restaurants}}, c_{\text{malls}}, c_{\text{banks}}, c_{\text{parks}}, c_{\text{total}}]^T. \quad (3)$$

For category k, the count is

$$c_k = \sum_{j \in \mathcal{P}_k} \mathbb{I}\big[\text{dist}(p, j) \le R\big], \quad (4)$$

where P_k is the set of POIs in category k and I[·] is the indicator function.

D. Listing Text Features

Listing narratives can contain terms that describe finishes, amenities, and architectural features. The text pipeline normalizes tokens, removes stopwords, and calculates unigram and bigram TF–IDF values. Let df(t, D) = $|\{d \in D : t \in d\}|$ denote the number of documents containing term t. The smoothed inverse document frequency and TF–IDF are then defined as:

$$\text{idf}(t, D) = \log\left(\frac{1 + |D|}{1 + \text{df}(t, D)}\right) + 1. \quad (5)$$

$$\text{TFIDF}(t, d, D) = \text{TF}(t, d)\,\text{idf}(t, D). \quad (6)$$

E. Visual Embeddings

An ImageNet-pretrained ResNet-18 processes each prop-erty photograph. Images are resized to 256 × 256 pix-els, center-cropped to 224 × 224, then normalized using channel means [0.485, 0.456, 0.406] and standard deviations [0.229, 0.224, 0.225]. Global average pooling yields a 512-dimensional vector $\mathbf{h}_k$ for image I_k.

For a gallery G = $\{I_1, \ldots, I_N\}$, where $1 \le N \le 15$, the property representation is the element-wise mean:

$$\mathbf{v}_{\text{vision}} = \frac{1}{N}\sum_{k=1}^{N} \mathbf{h}_k \in \mathbb{R}^{512}. \quad (7)$$

Mean pooling is invariant to image ordering and maps galleries of different sizes to a fixed-length vector.

F. Feature Fusion and Regression

The four modality vectors are concatenated as

$$\mathbf{X}_{544} = [\mathbf{v}_{\text{struct}} \,\|\, \mathbf{v}_{\text{geo}} \,\|\, \mathbf{v}_{\text{nlp}} \,\|\, \mathbf{v}_{\text{vision}}] \in \mathbb{R}^{544}. \quad (8)$$

The target is continuous property value. XGBoost is config-ured with 350 estimators, learning rate 0.04, maximum depth 6, subsample ratio 0.85, and column-sampling ratio 0.8. The regularized squared-error objective is

$$\mathcal{L} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{m=1}^{M}\left(\gamma T_m + \frac{\lambda}{2}\sum_{j=1}^{T_m} w_{mj}^2\right), \quad (9)$$

where T_m is the number of leaves in tree m, w_{mj} is a leaf weight, and γ and λ are regularization parameters.

IV. EXPLAINABLE VALUATION

TreeSHAP decomposes a prediction into a baseline value and additive feature contributions:

$$f(x) = \phi_0 + \sum_{i=1}^{544}\phi_i(x), \qquad \phi_0 = \mathbb{E}[f(X)]. \quad (10)$$

The Shapley value for feature i is defined over subsets S of the remaining feature set F:

$$\phi_i(x) = \sum_{S \subseteq F\setminus\{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!}\left[f_{S\cup\{i\}}(x) - f_S(x)\right]. \quad (11)$$

To summarize the contribution of modality group G, feature attributions are aggregated as

$$\Phi_G(x) = \frac{\sum_{i\in G}|\phi_i(x)|}{\sum_{j=1}^{544}|\phi_j(x)|} \times 100\%. \quad (12)$$

These values describe how the fitted model allocates its prediction across features; they do not establish causal effects of amenities, text, or appearance on market price.

V. EXPERIMENTAL EVALUATION

A. Protocol

The evaluation sample contains 600 residential properties from 30 Bengaluru micro-markets. A random 80/20 split with seed 42 assigns 480 properties to training and 120 to testing; the project description reports no property overlap. All four

ablation conditions use this split and identical XGBoost hyperparameters. Locality-rate target encoding uses training data only to avoid incorporating test targets.

The project overview also describes platform coverage across 62 residential micro-markets. This is broader platform coverage, distinct from the 30-market, 600-property ablation sample reported here. The backend's reported 19,778 real-estate records refer to the application data store, not to the ablation sample.

B. Ablation Results

Table I reports the four-stage results supplied for this study. Stage-to-stage changes are measured against the immediately preceding model.

The full model's test MAE is INR 50.08 lakh, compared with INR 62.39 lakh for the structured baseline, a reported cumulative reduction of 19.73%. Test RMSE falls from INR

96.90 lakh to INR 80.90 lakh (16.51%). The R^2 score in-creases by 0.1535, from 0.4933 to 0.6468, equivalent to a 31.1% relative increase over the baseline R^2. Adding visual embeddings produces the largest individual stage improvement in the reported ablation: a 14.87% MAE reduction from Model C to Model D.

These results describe a single held-out split. No confidence intervals, repeated random splits, or external-market validation results were supplied; accordingly, the reported metric differ-ences should not be interpreted as statistical significance or proof of performance in other markets.

VI. PLATFORM IMPLEMENTATION

A. Service and Data Layer

The application exposes asynchronous REST services using FastAPI. NumPy and PyTorch support feature and tensor oper-ations in the inference pipeline; the project reports inference latency below 150 ms. PostgreSQL persistence is managed with SQLAlchemy and Alembic migrations. The application data store is reported to contain 19,778 real-estate records,together with lead-pipeline and transaction-log data.

This service layer connects valuation and property data functions to the web client while keeping persistent application records in PostgreSQL. SQLAlchemy provides the persistence interface, and Alembic is used for schema migrations. The reported inference time describes the project implementation; the available evaluation does not specify hardware, request volume, or a latency measurement protocol, so it should not be treated as a general service-level guarantee. The 19,778 stored records are application data and are distinct from the 600-property sample used to calculate the model metrics.

B. Web and Geospatial Interface

The client is built with Vite and React 18. Leaflet provides interactive map views, including locality-level POI markers, amenity circles, and price-oriented GIS visualization. The interface supports locality search, property catalog browsing, filtering, and property listing. The project description identi-fies Geoapify for geocoding and categorized-place data and OpenStreetMap layers for map display. Figs. 2–6 show the platform interface.

These components support a user flow from finding a local-ity to inspecting available properties and their surroundings. Geoapify geocoding and categorized-place data provide loca-tion context, while Leaflet presents that context through map markers and amenity circles over OpenStreetMap layers. The locality report complements the catalog by presenting price-oriented GIS information alongside nearby points of interest; the maps therefore serve as a browsing and neighborhood-inspection interface rather than as a separate valuation model.

C. Additional Application Features

The platform also includes an interactive loan-amortization calculation, role-based access control, and payment verifica-tion for valuation reports. A Twilio and Gemini Live inte-gration supports a bidirectional voice callback agent intended to answer prospective buyers' questions using property infor-mation. These application features complement the valuation pipeline; they are not evaluated by the ablation metrics in Section V.

The loan calculator, account controls, payment verification, and voice callback support property exploration; their usabil-ity, security, and response quality are outside the scope of the predictive evaluation.

TABLE I FOUR-STAGE ABLATION RESULTS ON THE FIXED HELD-OUT TEST SPLIT.

Stage	Model	Modalities	Dim.	MAE (lakh)	RMSE (lakh)	R^2	Δ MAE
I	A (baseline)	Structured	6	62.39	96.90	0.4933	–
II	B	Structured + GIS	16	61.64	95.55	0.5073	−1.21%
III	C	Structured + GIS + NLP	32	58.83	93.16	0.5316	−4.56%
IV	D (full)	Structured + GIS + NLP + vision	544	50.08	80.90	0.6468	−14.87%

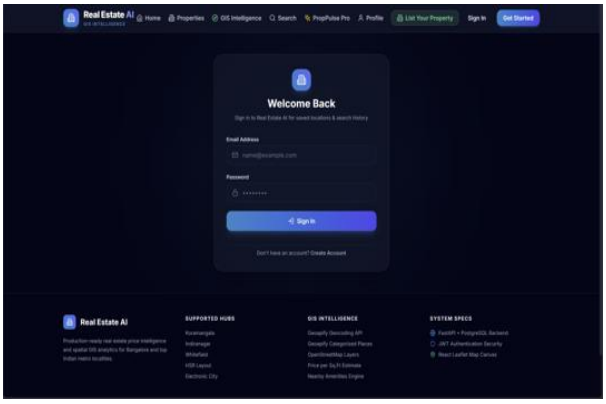

Fig. 2. Sign-in screen.

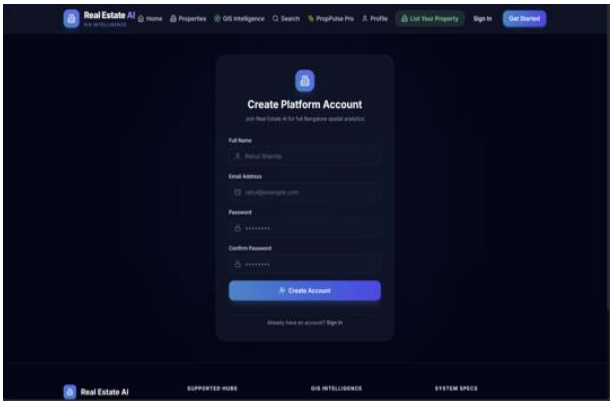

Fig. 3. Account creation screen.

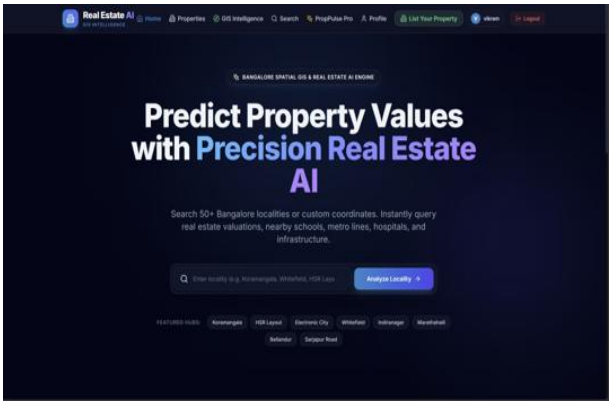

Fig. 4. Locality-search landing page.

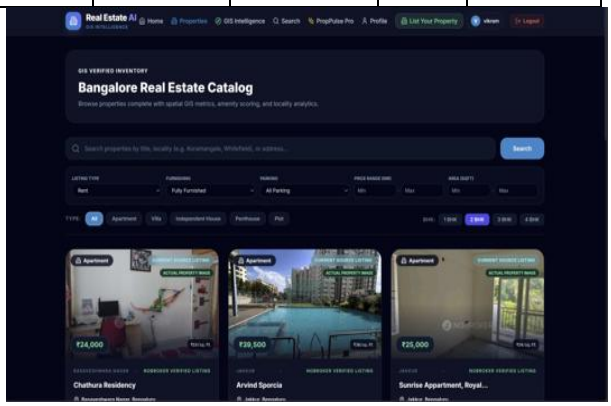

Fig. 5. Searchable and filterable property catalog.

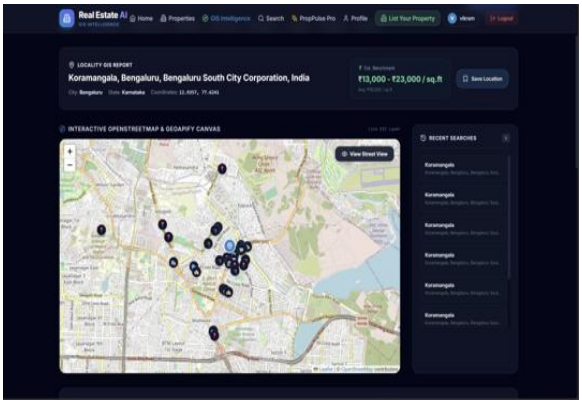

Fig. 6. GIS locality report with price benchmark and POI markers.

VII. CONCLUSION AND FUTURE WORK

This paper presented GeoSmart Price AI, a multimodal property valuation framework combining physical attributes, local POI counts, listing-text semantics, and ResNet-18 image embeddings. XGBoost performs regression on the resulting 544-dimensional representation, and TreeSHAP provides ad-ditive model explanations. On the supplied fixed Bengaluru test split, the full model reports lower MAE and RMSE and higher R^2 than the structured-only baseline. The screenshots illustrate how valuation and locality intelligence are situated within

a property discovery platform.

The current evaluation is limited to 600 records, one train/test split, and Bengaluru markets. Future work should evaluate additional cities and time periods, repeat the ablation across multiple splits, report uncertainty intervals, assess cali-bration and subgroup errors, and audit the provenance and rep-resentativeness of listing text and photographs. Explanations should be presented as model attributions rather than causal claims, and estimates should support, not replace, qualified appraisal and due diligence.

REFERENCES

[1] N. Karanikolas, E. Kyriakidou, and E. Athanasouli, "Artificial Intelli-gence and Real Estate Valuation: The Design and Implementation of a Multimodal Model," *Information*, vol. 16, no. 12, Art. no. 1049, 2025, doi: 10.3390/info16121049.

[2] D. Gerald and E. Jeremy, "Machine Learning for Real Estate Investment Risk Assessment: Developing Predictive Models for the US Market," ResearchGate, Feb. 2024. [Online]. Available: ResearchGate record.

[3] P. Jafary, D. Shojaei, A. Rajabifard, and T. Ngo, "Automated land valuation models: A comparative study of four machine learning and deep learning methods based on a comprehensive range of influential factors," Cities, vol. 151, Art. no. 105115, 2024, doi: 10.1016/j.cities.2024.105115.

[4] C. Huang, Z. Li, F. Chen, and B. Liang, "Multimodal Machine Learning for Real Estate Appraisal: A Comprehensive Survey," arXiv:2503.22119, 2025, doi: 10.48550/arXiv.2503.22119.

[5] A. Nouriani and L. Lemke, "Vision-based housing price estimation using interior, exterior & satellite images," Intelligent Systems with Applica-tions, vol. 14, Art. no. 200081, 2022, doi: 10.1016/j.iswa.2022.200081.

[6] D. Kristiyanto, "NLP: Text Summarization and Keyword Extraction on Property Rental Listings—Part 1," Towards Data Science, Jul. 8, 2024. [Online]. Available: article page.

[7] A. Manasa, P. Chandana, M. S. Ahmed, and K. Baradur, "Precision in Real Estate: Hybrid Multi-Stage Predictions for House Prices," International Research Journal on Advanced Engineering Hub, vol. 3, no. 5, pp. 2559–2563, May 2025, doi: 10.47392/IRJAEH.2025.0381.

[8] G. Maselli and A. Nesticò, "Machine Learning Algorithms and Explain-able Artificial Intelligence for Property Valuation," Real Estate, vol. 2, no. 3, Art. no. 12, 2025, doi: 10.3390/realestate2030012.

[9] H. S. Belsare and K. V. Warkar, "A Novel Model for House Price Prediction with Machine Learning Techniques," International Journal of Scientific Research in Science and Technology, vol. 10, no. 3, pp. 743–754, May–Jun. 2023, doi: 10.32628/IJSRST523103134.

[10] P. K. Verma, S. Arya, and C. Asbe, "Predicting Future Housing Prices: A Machine Learning Approach," Multidisciplinary Science Journal, vol. 5, Art. no. 2023ss0206, 2023, doi: 10.31893/multiscience.2023ss0206.