Towards Explainable and Scalable Property Valuation via BIM and GA-Enhanced Ensemble Learning

Abstract

Purpose – This study proposes a novel BIM-integrated, explainable and scalable framework for property valuation. It aims to enhance predictive accuracy and transparency by combining IFC-based feature extraction with a Genetic Algorithm enhanced Ensemble Learning, addressing key limitations of traditional and opaque AI models in the built environment.

Design/methodology/approach – An IFC-based pipeline converts BIM geometry and semantics into machine-readable features. A Genetic Algorithm-enhanced Gradient Boosting Regressor (GA-GBR) is trained and tested on 152k transactions from China, and then further tested from building type perspective at smaller scales. Hyperparameter optimization is performed using a genetic algorithm, and model interpretability is enabled through SHAP.

Findings – The GA-GBR model outperforms 11 benchmark models, achieving a 3-4.6% gain on MAPE over recent state-of-the-art methods. SHAP analysis identifies key predictors - local price level, transaction timing and floor area - while submarket results highlight context-specific drivers such as elevator presence in high-rise buildings. GA-based optimization enhances both predictive performance and feature relevance.

Practical implications – The framework supports automated, explainable and scalable valuation using BIM-derived features, enabling end-to-end deployment and informed decision-making. It offers valuers, designers and policymakers a transparent tool for assessing property value at multiple scales.

Originality/value – This is the first study to integrate GA-optimized ensemble learning, IFC-derived features and xAI techniques into a unified BIM-valuation workflow validated against real world data. It contributes to methodological advancement while facilitating industry adoption of explainable AI applications in the built environment.

Keywords: Artificial Intelligence; Building Information Modeling; Property Valuation; SHAP (Shapley Additive Explanations); Ensemble Learning.

1. Introduction

The integration of Artificial Intelligence (AI) and Building Information Modeling (BIM) is reshaping the landscape of the building sector by enabling data-driven approaches to design optimization, risk analysis and decision-making. In particular, property valuation, critical to project feasibility, investment planning and cost-benefit analysis, is increasingly informed by AI models that can process large volumes of real estate and design-related data. However, a significant challenge persists - while AI models may offer high predictive accuracy, they often function as 'black boxes', providing little insight into how a decision is reached. This lack of explainability poses concrete problems in several practical contexts. For instance, in public-sector land valuation, a lack of transparency can lead to legal disputes or resistance to tax assessments. In private property valuation, opaque models make it difficult for investors or lenders to justify financial decisions or adapt designs to market signals. This undermines stakeholder trust, reduces model adoption in practice and limits the potential for AI-generated insights to support upstream design or project planning decisions within BIM workflows. Traditional approaches to property valuation, whether through human appraisers or even machine learning models like regression and neural networks (Burgess *et al.*, 2018; McCluskey *et al.*, 2013; Valier, 2020), have limitations as follows:

(1) They fail to integrate tightly with BIM-based design data, missing opportunities to create a feedback loop between valuation insights and design decisions.
(2) They often provide predictions without transparent justifications or interpretable reasoning.

This research addresses the urgent need to bridge that gap by exploring how Explainable AI (xAI) can be integrated into BIM workflow to provide automated and interpretable valuation insights. Specifically, it investigates how BIM-based workflow can enhance feedback loops within building design and inform decision making, and whether a Genetic Algorithm optimized Gradient Boosting Ensemble model (GA-GBR) outperforms traditional AI models in predictive performance and model interpretation. The main research questions guiding this study are: (1) how can xAI be effectively embedded within BIM workflows to improve interpretability in property valuation, (2) what actionable insights can xAI (SHAP) uncover at both global and individual property levels on feature importance and interaction effects to support design and investment decisions, and (3) how the proposed GA-GBR model performs compared to conventional AI models in delivering meaningful property valuation outputs.

The proposed xAI-BIM integration advances property valuation through three main innovations. First, explainable AI methods such as Shapley values and feature selection comparison are applied to improve interpretability and reveal relationships between property features and market value. Second, a Genetic Algorithm optimized Gradient Boosting Regressor (GA-GBR) is developed to enhance predictive performance and provide explainable outputs, evaluated at both large scale (national dataset) and finer submarket levels. Third, an Industry Foundation Classes (IFC)-based information extraction pipeline is implemented to standardize and integrate BIM-derived design parameters into valuation workflows. Collectively, these components establish a framework that strengthens feedback loops between valuation and design, supporting transparency and trust in AI-driven decision systems.

2. Related Work

Property valuation plays a pivotal role in construction, investment, taxation and lending. An accurate market price analysis should reflect property attributes, market culture fundamentals and geographical locations (Pagourtzi *et al.*, 2003). The dynamics of subjective factors and the real estate market's opacity make it hard for an accurate and objective property valuation. Typical information required for property valuation involves regional, city and neighborhood data, site data, building data, sales and cost data, income and expense data. The impact of macroeconomic variables on property value is beyond this research.

2.1 AI and BIM in Property Valuation

AI applications in property valuation have advanced from traditional regression models to neural networks and ensemble methods. While early research highlighted the superior performance of neural networks in capturing non-linear patterns (Lewis *et al.*, 1997; Liu *et al.*, 2011), limitations such as high computational demand and poor scalability remain, particularly for large-scale or real-time valuation (Rafiei and Adeli, 2016). To address these issues, genetic algorithms (GA) have been used for feature selection and model tuning, improving predictive accuracy in hybrid models (Ahn *et al.*, 2012; Sun, 2019). Rafiei and Adeli (2016) demonstrated the value of integrating GA with deep learning for early-stage valuation, though challenges with interpretability and generalization persist.

As a response, explainable AI (xAI) has emerged to improve model transparency and stakeholder trust. Shapley Additive Explanations (SHAP), a widely adopted xAI method, quantifies feature contributions and enhances the interpretability of tree-based models like XGBoost and Random Forests (Iban, 2022; Tchuente, 2024). These techniques help valuers understand how attributes such as location or building age influence predictions. However, xAI applications remain narrow. Most models rely on structured data, neglecting behavioral or contextual variables. Recent studies have integrated natural language processing (NLP) techniques and applied explainable AI (xAI) methodologies to address sustainability-related valuation challenges including regional generalizability constraints and limited model interpretability. This underscores the critical need for developing multi-scale explainable frameworks in contemporary valuation practice (Baur *et al.*, 2023; Doan *et al.*, 2024; Konhäuser and Werner, 2024; Tarasov and Dessoulavy-Śliwiński, 2024).

While several studies have examined BIM's technical capabilities, its integration into real-world valuation practice remains limited. Isikdag et al. (2015) and Wilkinson and Jupp (2016) proposed using 3D models and contextual data to support property professionals, yet their frameworks remain largely conceptual and detached from live valuation workflows. Practical efforts to link BIM and valuation include El Yamani et al. (2019), who applied hedonic pricing using IFC-derived variables. However, external market factors were excluded and limited case validation restricts generalizability. Arcuri et al. (2020) proposed a BIM-GIS framework based on the cost approach, while data-rich, lacks sensitivity to market dynamics. Other studies, such as RICS (2017) envisions broader BIM adoption, its practical use for market valuation requires an explainable AI layer to bridge BIM data and valuation outcomes.

Notably, Su et al., (2021) presented an early framework combining IFC data with machine learning-based valuation, but it lacked end-to-end automation, explainability and rigorous validation across submarket contexts. This study extends that prior work by embedding SHAP-based explanations, optimizing both features and hyperparameters via a GA mechanism, and conducting multi-scale empirical experiments.

Recent research from 2025 further pays attention to cloud-based BIM analytics and real-time valuation with automated valuation models (Hong and Guo, 2025; Muccio and Cannatella, 2026), underscoring the need for a scalable, explainable and automated framework for property valuation.

2.2 Research Gaps

Despite significant progress in integrating Artificial Intelligence (AI) and Building Information Modeling (BIM) into property valuation, several critical research gaps remain. Figure 1 provides a structured summary of these research gaps alongside the corresponding components of the proposed framework.

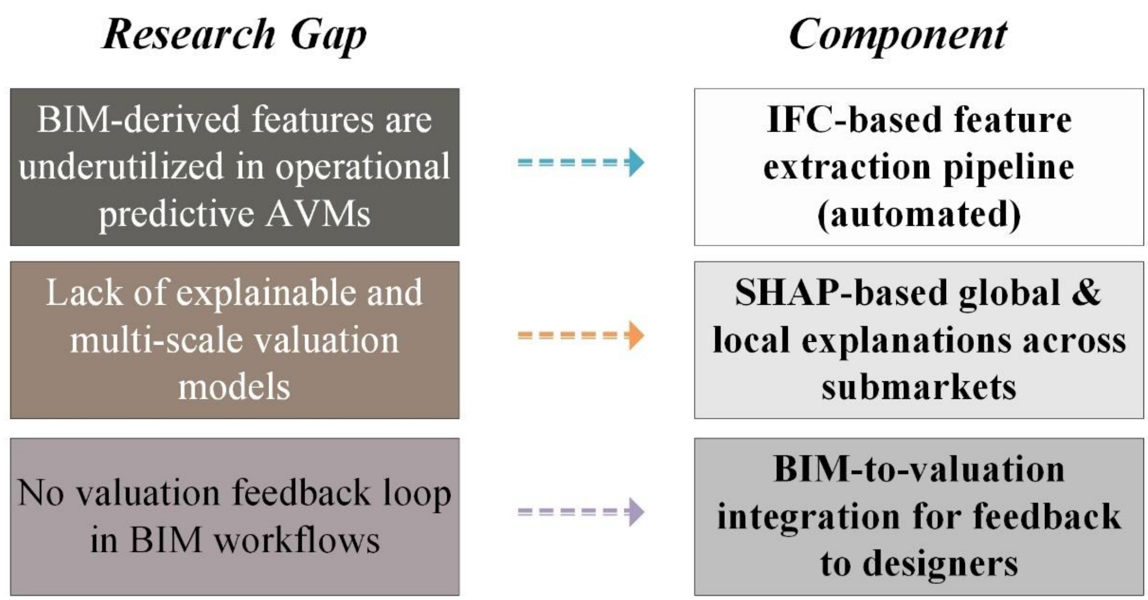

Figure 1: Linking research gaps to framework components

The first gap pertains to the limited use of semantically rich and BIM-derived variables within automated valuation models (AVMs). Although several studies have extracted IFC attributes, these features are rarely integrated into end-to-end predictive workflows. This study addresses the gap by implementing an IFC-based feature extraction pipeline that converts BIM elements into structured data for AI-driven valuation.

The second gap highlights the absence of explainable models capable of operating across multiple markets or building contexts. Many current models provide global insights but lack localized interpretation, which limits their utility for diverse stakeholders. The proposed framework incorporates SHAP-based global and local explanations, enabling transparency at both the aggregate and individual property levels, and is validated across different building types to ensure adaptability.

The last reflects the lack of bidirectional feedback between BIM design workflows and valuation insights. Existing models typically operate in isolation, providing limited guidance to designers or planners. This study contributes a BIM-integrated valuation pipeline that supports real-time feedback by linking predictive insights to BIM parameters.

By systematically aligning each identified gap with a specific component, this research contributes to both methodological rigor and practical applicability, thereby supporting the broader adoption of explainable and scalable AI for property valuation in BIM-enabled environments.

3. Research Methodology

Following Design Science Research (DSR) methodology, this research systematically develops and evaluates a novel BIM-integrated property valuation framework leveraging Genetic Algorithm (GA)-enhanced ensemble learning to achieve automation, transparency and scalability.

3.1 Problem statement

Prior research on AI-BIM property valuation integration remains either conceptual or employs opaque models, limiting practical adoption. The systematic integration of BIM-derived data into a valuation framework remains underexplored beyond isolated applications. This study addresses these gaps by proposing a transparent and automated valuation framework integrating BIM data via IFC into a GA-optimized Gradient Boosting Regressor (GA-GBR), enhanced with explainable AI techniques including Shapley values.

3.2 Requirement analysis

The framework development requires establishing functional, technical and data requirements. Functional requirements include automated IFC-based information extraction for BIM-derived features (floor area, room numbers), xAI methods (SHAP) for interpretability, multi-scale adaptability, and feedback-driven decision support for design optimization.

Technical requirements involve implementing the GA-GBR model for optimal predictive performance through automated hyperparameter tuning and feature selection. The framework integrates model-agnostic xAI techniques (SHAP) enabling global and local interpretability, which is vital for stakeholder trust. The IFC data extraction must ensure reliability and efficiency with prevalent BIM software.

Data requirements emphasize high-quality IFC-encoded BIM models containing standardized structural, spatial and semantic data. Transaction datasets should include historical prices, property characteristics, transaction timelines and geographic attributes to capture valuation drivers and enhance interpretability through diverse scenarios.

Together, these technical and data requirements provide the foundation for building an integrated, explainable and scalable BIM-xAI valuation framework that can support predictive accuracy and design feedback.

3.3 System design and development

The system architecture integrates BIM data, machine learning and xAI methods through modular components as shown in Figure 2:

- **IFC-Based Information Extraction**: BIM parser automatically extracts spatial and semantic features from IFC models, including geometric attributes (floor area, volume) and structural characteristics.

- **GA-GBR Modeling Pipeline**: Core module performing automated feature selection and hyperparameter tuning to optimize prediction accuracy and efficiency.
- **xAI Visualization Engine**: Incorporates SHAP and feature selection comparisons for global and local model interpretability, enabling stakeholders to trace feature influence on valuations.
- **Design Feedback Interface**: Generates valuation feedback supporting early-stage design optimization by mapping design element influence on predicted market value.

This architecture facilitates comprehensive property valuation by systematically integrating BIM and AI, fulfilling industry needs and research innovation requirements.

3.4 System validation

A multi-phase validation strategy ensures robustness, explainability and practical applicability: (1) comparative evaluation of GA-GBR against eleven machine learning models using transaction datasets across multiple scales, demonstrating improved predictive accuracy; (2) SHAP application to verify feature importance, supplemented by comparative analysis against conventional GBR; and (3) practical case study implementation using a Chinese BIM model, demonstrating IFC-based feature extraction feasibility and industry data integration.

This comprehensive validation ensures analytical robustness and practical applicability, advancing scalable, explainable and BIM-integrated property valuation.

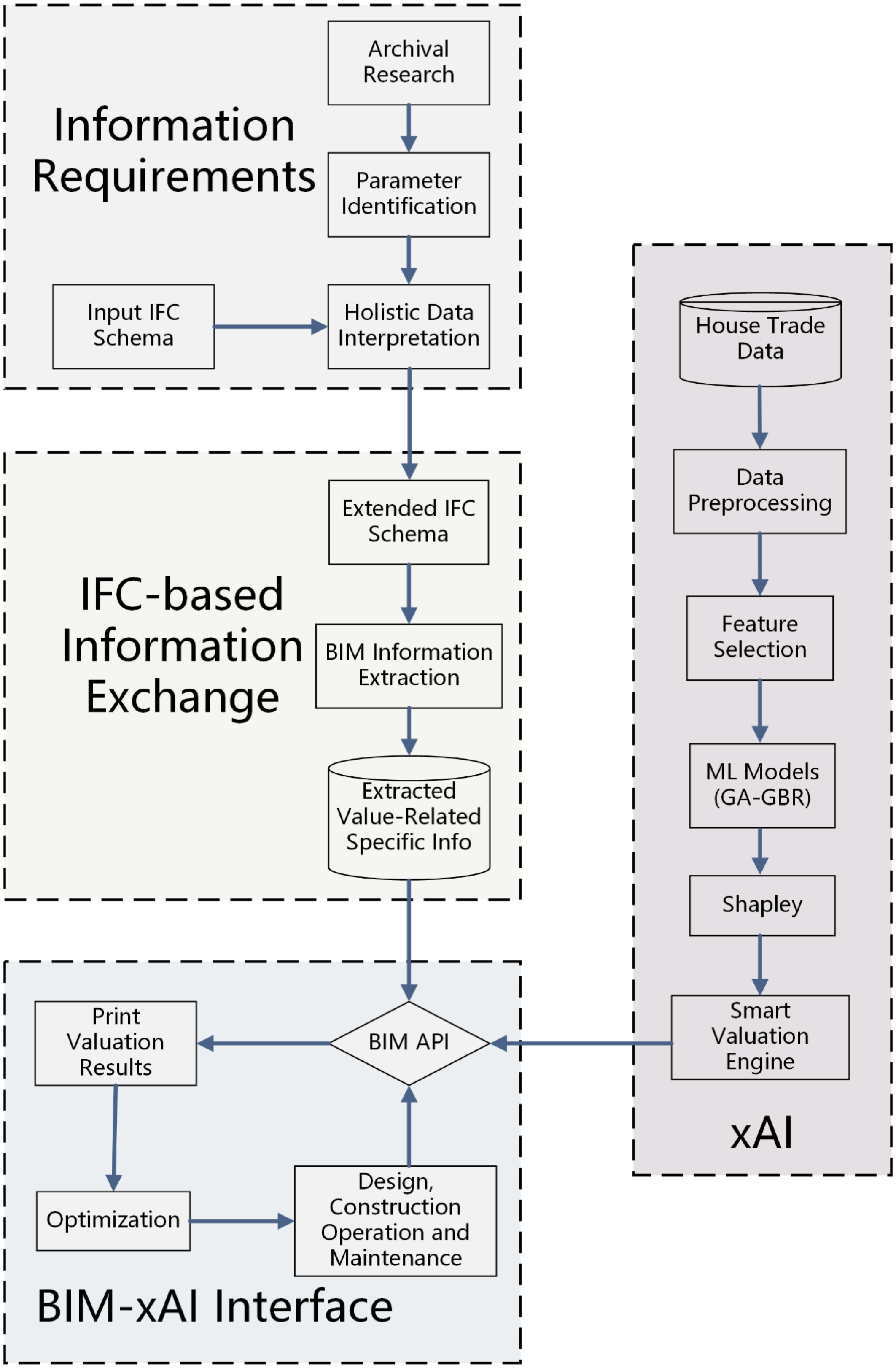

Figure 2: System design for the BIM-xAI integration

4. System development

With the conceptual architecture and validation roadmap established, section 4 moves into the implementation phase. It details the development of the two core modules: (1) the IFC-based information extraction pipeline, which transforms spatial and semantic data from BIM models into machine-readable features; and (2) the Genetic Algorithm optimized Gradient Boosting Regressor (GA-GBR), which performs predictive modeling enhanced by feature selection and interpretability. This section explains how these components are developed, integrated and prepared for empirical validation, forming the operational backbone of the proposed BIM-xAI valuation system.

4.1 IFC-based information extraction for property valuation

Partial data model retrieval in BIM can be categorized into schema-based and instance-based approaches. Schema-based methods rely on predefined structures like IFC to extract data, while instance-based approaches focus on retrieving specific object-level information directly from the model. Building on the no-schema algorithm by Won et al. (2013), this study adopts an instance-based method using the open-source *IfcOpenShell* library. This integration enables efficient extraction of common physical elements while offering flexibility to meet user-defined information needs. The extraction process involves three steps as follows:

(1) Target information identification in IFC instances

This step aims to identify the target information in an IFC instance model and define the representation of its data structure. The target information within an IFC instance model contains the current value-related design information in building objects (*IfcSpace*) and their value-specific properties (total area, built date and renovation condition). Therefore, the representation of the target information includes several key elements in an IFC data model: (1) the globally unique identifier number (GUID) of an IFC instance model, (2) the attributes of building objects including building object names, and (3) the attributes of required *IfcProperty* instances that contain the property set names, property names, property types, and their nominal values. Figure 3 gives an example of the representation of the target information based on IFC schema.

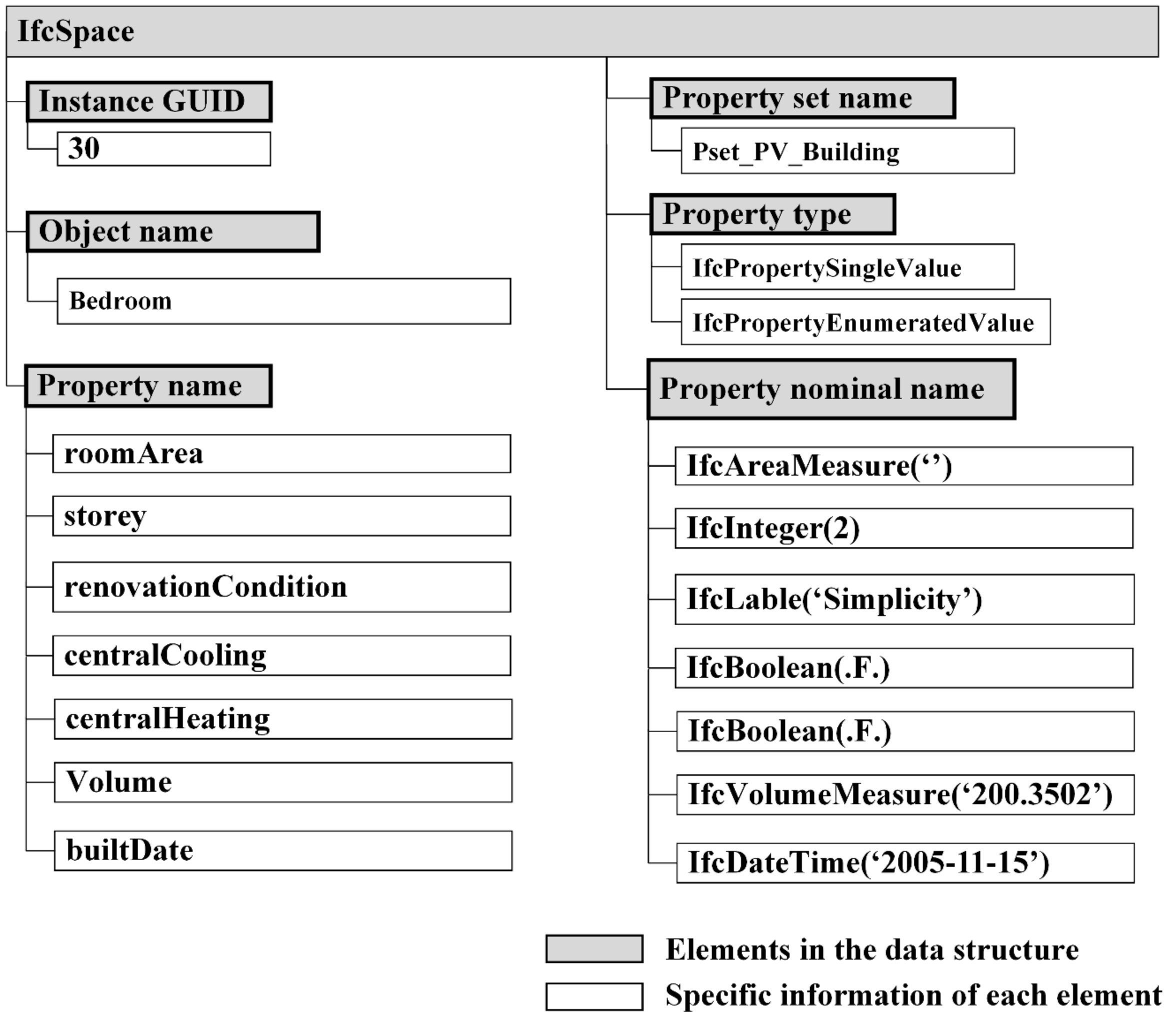

Figure 3: An example of an item in the data structure representation

(2) Information extraction algorithm development

To extract value-related information elements about building objects and their attributes for property valuation, information exchange between construction projects and property valuation is delivered through an IFC-based information extraction algorithm. The context of how to develop an IFC-based extraction algorithm is described in Figure A and Figure B in the Appendix. After analyzing the information elements and their relationships between *IfcObject* and *IfcProperty*, the IFC-based extraction algorithm was developed on Python using *IfcOpenshell-python* module on *Pycharm* software.

The algorithm initiates by identifying the relevant IFC entities associated with the specified building elements. It then recursively traverses all linked data instances until the complete set of associated properties is extracted. As illustrated in Figure C (Appendix), the flowchart outlines the logic of the developed IFC-based information extraction procedure. This approach retrieves building element attributes both directly via the *IfcRelDefinesByProperties* relationship and indirectly via the *IfcRelDefinesByType* relationship. The operational implementation of the algorithm, detailing both direct and indirect extraction paths, is presented in Table 1 with an example of the extracted data.

Table 1: An example of an extracted information item

Instance GUID	Object name	Property set	Property name	Property nominal value
30	Bedroom	Pset_PV_Building	renovationCondition	IFCLabel('simplicity')

(3) IFC Data cleaning and standardization

Next, IFC-derived data was preprocessed through four steps: (1) handling missing attributes by assigning default values or imputing with dataset-level medians; (2) standardizing heterogeneous property naming conventions by mapping synonyms (e.g. 'GrossFloorArea' and 'TotalArea') to unified categories; (3) resolving unit inconsistencies (e.g. converting square feet to square meters based on IFC unit assignments); and (4) detecting and removing outlier values such as negative areas or unrealistic floor counts. These steps ensure data consistency and improve the robustness of the GA-GBR valuation model.

4.2 Genetic algorithm optimized gradient boosting ensemble model (GA-GBR)

The rationale behind the model selection begins with a proof of concept, testing the 11 typical regression models including linear regressions, K-Nearest Neighbors (KNN), Supporter Vector Machine (SVM), decision tree-based models such as Classification and Regression Trees (CART), random forest and gradient boosting regression (GBR), and Artificial Neural Network (ANN) on the UCI Machine Learning repository - Boston housing dataset (Harrison and Rubinfeld, 1978). Typical performance metrics for regression analysis involve the mean absolute error (MAE), mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2) etc. To get a comprehensive understanding of eleven automated valuation models (AVMs) performances, all the above-mentioned metrics were selected. It was found that the GBR model had the highest prediction accuracy and lowest error among all the five metrics (results shown in Table B in the Appendix). In addition, from literature, it is concluded that compared to neural networks, ensemble learning has advantages in terms of model interpretability and flexibility, which is more suitable for knowledge mining and system development for property valuation.

The architecture of the proposed GA-GBR model consists of three key stages: base learner generation, problem encoding and genetic search, which has been provided in our previous study (Su *et al.*, 2021). While the GA-GBR model has demonstrated improved predictive accuracy, this subsection highlights how it enhances model interpretability in property valuation tasks.

To address current challenges within GBR including overfitting and the trade-off between independent and diverse of base learners, GA-GBR integrates a Genetic Algorithm (GA) with the traditional Gradient Boosting Regressor (GBR). The GA operates as an evolutionary feature selection engine, optimizing both the feature subset and model hyperparameters. This approach improves not only predictive accuracy but also interpretability by identifying and isolating the most relevant input variables.

The implementation of the proposed GA-GBR model focusing on enhanced model interpretability are provided as follows:

Base learner generation

The first step involves generating a pool of base learners using the input domain dataset. There are three common base-learner models: linear models, smooth models and decision trees. In this research, decision trees of uniform size are utilized as base learners, which are effective at handling mixed data types and modeling complex functions. The GBR ensemble combines multiple base learners $f_m(x_i)$ to generate a strong model $\hat{f}(x)$, which is displayed below.

$$\hat{f}(x) = \sum_{m=1}^{M} f_m(x_i) \tag{1}$$

The objective function for the base learner is to learn a mapping $f(x)$ between the input feature vector and the output (house price). Typical loss functions for regression models include Gaussian L_2 loss function, Laplace L_1 loss function, Huber loss function, and Quantile loss function (Natekin and Knoll, 2013). Initial testing on the GBR model indicated that it achieved the best prediction accuracy with Huber loss function.

Problem encoding

The problem encoding task is conducted in the next. This research employs binary encoding to represent solutions, focusing on exploring the relationship between input features and the target price. In the training of the GA-GBR model, each chromosome within the population corresponds to an individual that possesses N input features from the training data, as illustrated in Figure 4. For example, parameter A might represent house size, while parameter B could indicate the presence of central heating. A binary value (one or zero) denotes whether a feature is selected or not. The objective of using binary encoding for feature selection is to identify the near-optimal chromosome, where each bit corresponds to a feature, in order to discover the smallest subset of features that yields the highest predictive performance (Kanan *et al.*, 2007). The size of each chromosome is determined by the number of input features (N). For instance, in the experiments of the GA-GBR modeling, described in Section 5.1, N amounts to 56.

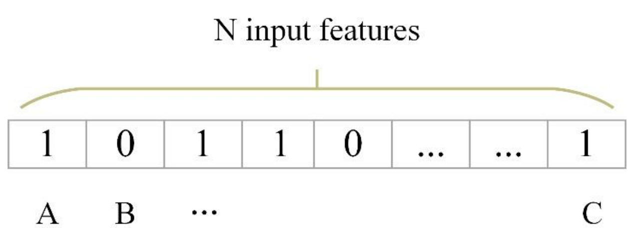

Figure 4: Binary chromosome encoding for feature selection in the GA-GBR

Genetic search

Through the operations of selection, crossover and mutation, the genetic algorithm (GA) iteratively evolves a population of candidate feature subsets toward near-optimal solutions. The fitness function that guides this evolutionary search is primarily defined by the coefficient of determination (R^2), with an added emphasis on model sparsity and generalization. This dual-objective design serves to mitigate overfitting while enhancing the interpretability of the resulting model.

5. Empirical studies

The preliminary construction of the proposed GA-GBR model involved three main steps: data collection, verification and data preprocessing. The dataset is collected from Kaggle, uploaded by Qiu, which comprises 152186 traded house data for properties registered in China from 2009 to 2018. Variables are classified as continuous (e.g., house price, total area), binary (e.g., elevator presence, property rights) and categorical (e.g., building structure).

Before model training, categorical variables are systematically encoded using appropriate techniques based on their characteristics:

- Binary variables (elevator, fiveYearsProperty, subway): retained as 0/1 binary indicators
- Ordinal variables (renovationCondition: 1-4 scale): preserved ordinal encoding reflecting quality hierarchy
- Nominal variables (buildingType, buildingStructure, district): one-hot encoded creating dummy variables.

To validate encoding robustness, we conducted systematic sensitivity testing:

- Compared one-hot and target encoding for nominal variables: R^2 difference < 0.5%
- Tested ordinal and one-hot for renovationCondition: ordinal preserved 2.1% better performance.

The GBR and GA-GBR models are developed and tested in Python using scikit-learn. Both models are trained and tuned using genetic search to optimize key hyperparameters, including the number of estimators, learning rate, tree depth, and loss function. Each dataset is split into 70% training and 30% testing sets, and further segmented by building categories to analyze feature-price relationships.

The GA parameters are systematically determined through preliminary experiments and sensitivity analysis to balance exploration and exploitation trade-offs:

(1) Population parameters

- Population size: 600 individuals (optimized through convergence analysis)
- Generations: 32 (with early stopping criteria)
- Chromosome length: 56 bits (corresponding to feature count)

(2) Evolutionary operators

- Selection: Tournament selection (k = 3)
- Crossover: Single-point crossover, probability = 0.8
- Mutation: Bit-flip mutation, probability = 0.1/chromosome length

(3) Validation and overfitting control

- Train-Test split: 70% training, 30% testing with stratified sampling preserving price distribution
- Cross validation: 5-fold CV for hyperparameter tuning and stability assessment
- Regularization: Huber loss, learning rate of 0.2, maximum tree depth of 7, and at least 50 samples per leaf
- Overfitting check: Training and testing R^2 gap < 1%, confirming genuine generalization across building types and time periods.

After trial-and-error testing, for full datasets, GA-GBR performs best with 200 estimators, learning rate (0.2), and Huber loss; for smaller 1,000-sample groups, optimal settings included 100 estimators and learning rate (0.1). The GA-GBR model performs best with a genetic algorithm with 600 individuals, 32 generations, and standard crossover and mutation rates. Model fitness is evaluated using R^2, and chromosomes outperforming the baseline GBR are selected, demonstrating improved accuracy and generalizability across scales.

5.1 Model performance

The performance of the trained GA-GBR model is initially evaluated on the big dataset, comparing the model performances with the other 11 different machine learning models. Table 2 provides a comparative analysis of five typical metrics for regression models. The metrics include Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) for error measurement, and R-squared (R^2) for prediction accuracy.

Table 2: Predictive accuracy of the 12 different automated valuation models

Accuracy metrics	**MAE**	**MAPE**	**MSE**	**RMSE**	**R^2**
Linear regression	76.79	26.0%	13042	114	80.3%
Ridge regression	76.75	26.0%	13033	113.9	80.3%
Lasso regression	77.11	26.1%	13265	114.9	80.0%
Elastic Net regression	77.89	26.3%	13602	116.4	79.5%
KNN	129.19	35.5%	40468	200.7	39.0%
SVM	171.28	51.4%	69810	263.7	5.3%
ANN	156.73	52.9%	51625	227.2	22.2%
CART	54.28	15.7%	8610	90.4	86.9%
AdaBoost	52.89	18.5%	6119	77.9	90.8%
Random forest	38.65	12.0%	4545	66.7	93.2%
GBR	36.87	11.8%	3832	61.9	93.9%
GA-GBR (proposed)	**32.76**	**10.55%**	**3026**	**55.01**	**95.2%**

Table 2 shows that linear and regularized regression models performed poorly, with high error rates and low model fit. KNN and SVM showed even weaker results, with R^2 below 0.40. In contrast, tree-based models delivered significantly better performance. CART reduced the MAE to 54.3 and increased R^2 to 86.9%, while ensemble methods like Random Forest and GBR further improved R^2 to 93.2%. and 93.9% The proposed GA-GBR model outperformed all others, achieving the lowest MAE (32.76), lowest MAPE (10.55%) and highest R^2 (95.2%). Notably, all 32 GA-generated models exceeded the predictive accuracy of the baseline GBR, demonstrating the effectiveness and generalization capability of the GA-optimized approach in large-scale valuation tasks.

To ensure the robustness of GA-GBR's superior performance, we conducted bootstrap analysis to establish 95% confidence intervals for GA-GBR and GBR. It shows that R^2 for GA-GBR at 0.952 [0.949-0.955] and R^2 for GBR at 0.939 [0.936-0.942]. This confirms that GA-GBR's 1.3% performance improvement is statistically significant ($p < 0.001$), demonstrating the effectiveness of genetic algorithm optimization.

5.2 Model explainability analysis

This section presented the explainability analysis of the proposed GA-GBR model using SHAP in the big dataset. First, feature importance was assessed using SHAP, enabling both global and local interpretability of feature contributions in a non-linear context. These insights were used to explain model behavior and identify influential variables beyond what correlation alone could reveal. Second, the performance and interpretability of the GA-GBR model were compared with a baseline GBR to highlight the advantages of GA-based optimization in enhancing both prediction accuracy and feature relevance. Together, these explainability techniques enhanced the transparency, interpretability and decision-support capabilities of the proposed BIM-xAI valuation framework.

1) SHAP feature importance analysis

Improving model performance and interpretability requires understanding the relationships between input features and property prices. SHAP values quantify each feature's contribution to a prediction by computing its marginal effects across all possible feature combinations. This provides both global interpretability (across the entire model) and local interpretability (specific to individual predictions), enabling more transparent and accountable decision-making in property valuation. Two types of SHAP summary plots are generated below to illustrate both global and local model behavior.

Figure 5 shows that the three most impactful features are: (1) *communityAverage*, which dominates as the strongest driver by capturing neighborhood price levels; (2) *tradeTime*, reflecting temporal market conditions and seasonality; and (3) *square* (floor area), which remains a major physical determinant of price. Among the next tier, *DOM* (active days on market) has the largest additional impact, followed by *bathRoom* and *livingRoom*. Geographic coordinates (Lat, Lng) contribute to a lesser extent. A long tail of variables - *constructionTime, ladderRatio, district, renovationCondition, drawingRoom, followers, buildingStructure, elevator, subway, fiveYearsProperty, floor,* and *buildingType* - exhibits minimal average contributions, indicating weaker or context-specific effects.

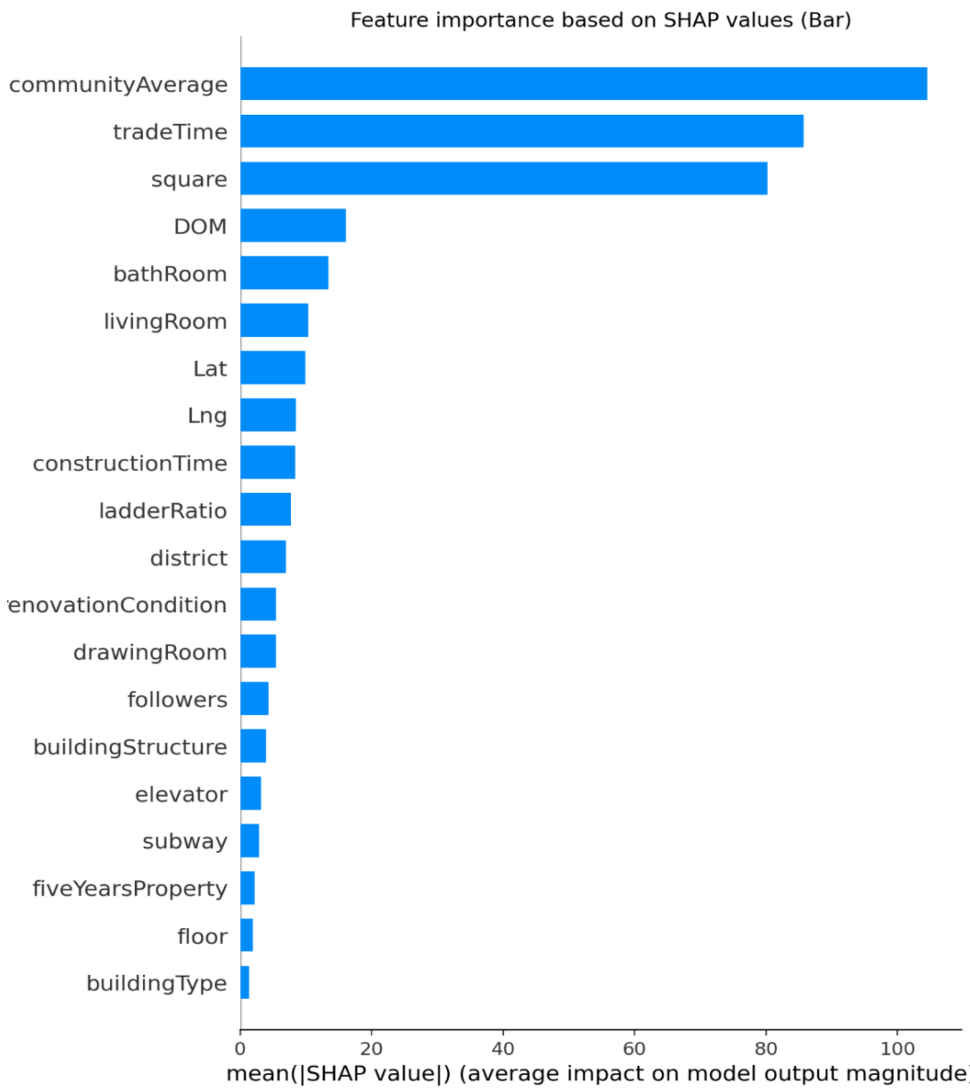

Figure 5: Global feature importance based on mean SHAP values

To complement the global feature rankings, SHAP summary plot in Figure 6 is generated to illustrate how individual feature values influence model predictions at the local level. Each point represents a single prediction, with the SHAP value reflecting the marginal contribution of a feature to the predicted sale price. Color denotes feature magnitude - red indicates higher values and blue indicates lower values.

Several key interpretations from Figure 6: (1) *communityAverage*: higher neighborhood averages (red) strongly push SHAP values to the right, raising predicted prices; lower values (blue) pull them down. (2) *tradeTime*: later transactions (red) increase predictions, while earlier dates (blue) depress them - consistent with market appreciation over time. (3) *square*: larger floor area (red) consistently lifts prices; smaller area (blue) reduces them. (4) *DOM* (active days on market): *DOM* (red) are associated with negative SHAP values, indicating a discount for stale listings. (5) *bathRoom* and *livingRoom*: higher counts (red) generally contribute positively, though with smaller magnitude than the top three drivers. (6) *Lat/Lng* and remaining attributes (*constructionTime, ladderRatio, district, renovationCondition, drawingRoom, followers, buildingStructure, elevator, subway, fiveYearsProperty, floor, buildingType*) show mixed, mostly modest effects - suggesting context-specific or second-order influences.

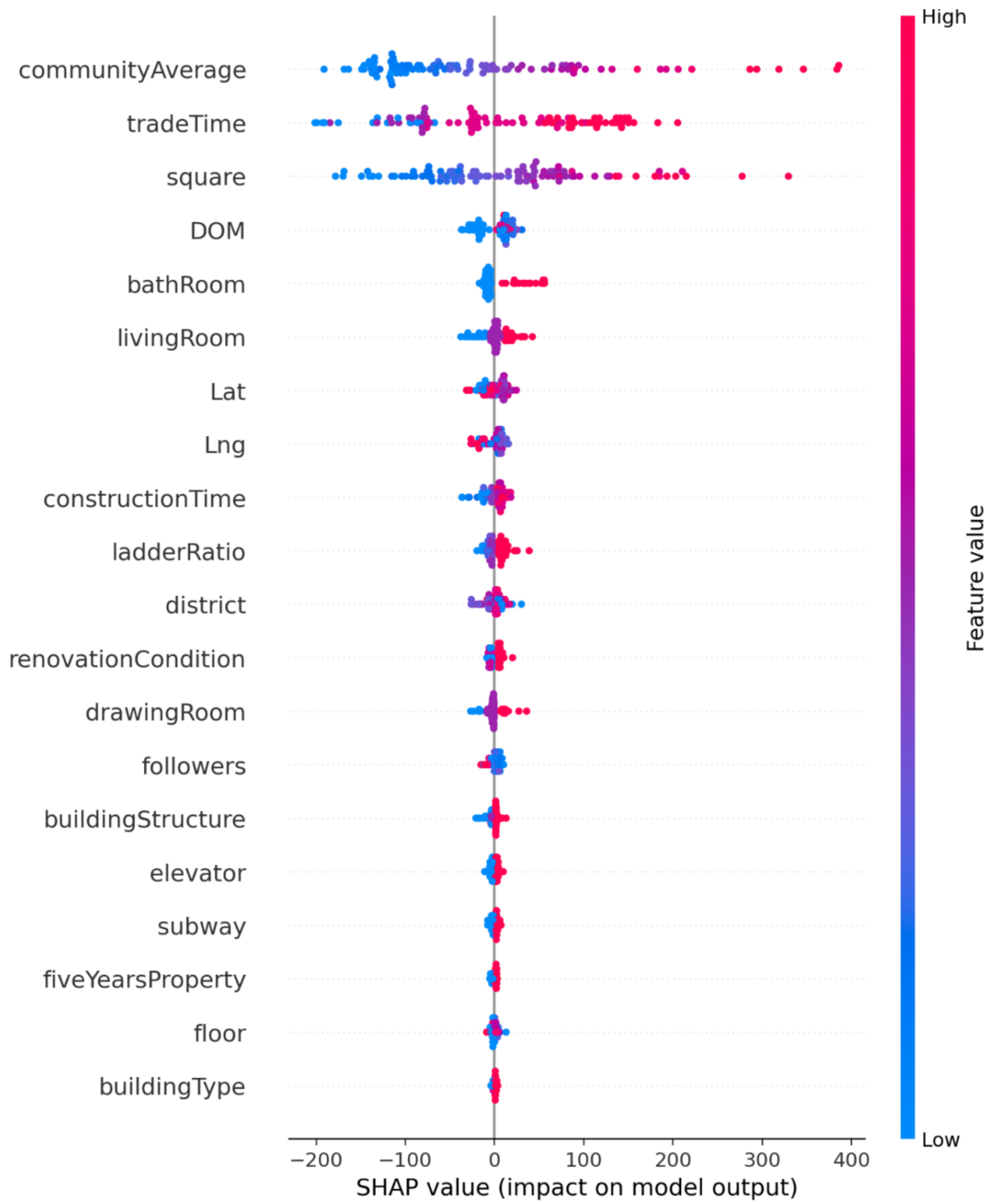

Figure 6: Local feature importance using SHAP values for individual predictions

To ensure robustness of SHAP-based insights, we performed bootstrap analysis (1,000 iterations) to establish 95% confidence intervals for mean absolute SHAP values of the top six features:

- Square: 96.01 [95% CI: 86.03–106.91]
- CommunityAverage: 90.21 [95% CI: 80.05–99.96]
- TradeTime: 84.95 [95% CI: 78.74–91.16]
- DOM: 15.60 [95% CI: 14.42–16.85]
- BathRoom: 14.48 [95% CI: 12.03–17.45]
- Lat: 11.86 [95% CI: 10.83–12.94]

Non-overlapping confidence intervals confirm statistically significant differences in feature importance rankings. This statistical validation confirms that the identified feature hierarchy is robust and not due to sampling variation.

Furthermore, a representative property transaction with its complete SHAP value breakdown is provided below, to demonstrate the GA-GBR model's interpretability at the individual instance level. A high-rise tower property case:

- Property profile: Tower building, 199m², 27th floor, 3Bedrooms/2Baths, 1 Elevator, District 7
- Transaction date: 2016-09-11
- Actual price: ¥14,300,000 (¥71,860/m²)
- Predicted price: ¥14,118,822 (¥70,951/m²)
- Prediction error: 1.27% (MAPE)

Key SHAP Contributions to property features:

- CommunityAverage: SHAP (+0.18 to +0.22) - Neighborhood pricing levels represent the dominant value driver, contributing 18-22% price premium, reflecting the strong influence of location and local market dynamics
- Square: SHAP (+0.14 to +0.17) - The 199m² floor area, approximately 2.2 times the market median, adds 14-17% to valuation as the primary physical characteristic
- TradeTime: SHAP (-0.03 to -0.04) - The 2016 transaction date applies a 3-4% discount relative to more recent market conditions
- Floor: SHAP (+0.002 to +0.003) - The 27th floor location adds minimal value (0.2-0.3%), indicating floor level has negligible impact in this specific case despite being a high floor
- Elevator: SHAP (+0.001 to +0.002) - Elevator access contributes marginally (0.1-0.2%), as it's a standard expectation for high-rise properties.

This case illustrates how structural features (elevator, floor) dominate valuation in high-rise properties, aligning with market expectations for vertical developments.

2) Further interpretation from the perspective of three building categories

To test the model performance at small scales, the dataset was subdivided into groups of 1000 property transactions, stratified by building categories. This approach enabled consistent comparative analysis across different property types while ensuring statistical robustness at the subgroup level. The predicted prices by the GBR and GA-GBR models were first compared with the actual prices using three building category-related datasets: (1) the tower group, (2) the combination of plate and tower group, and (3) the plate group. As shown in Figure 7, the GA-GBR model's predicted prices were closer to the actual prices than those predicted by the GBR model, with smaller MAPE values of 0.6%, 1.6%, and 1.02% for the three groups respectively. This demonstrated the higher prediction accuracy of the proposed GA-GBR model compared to the traditional GBR.

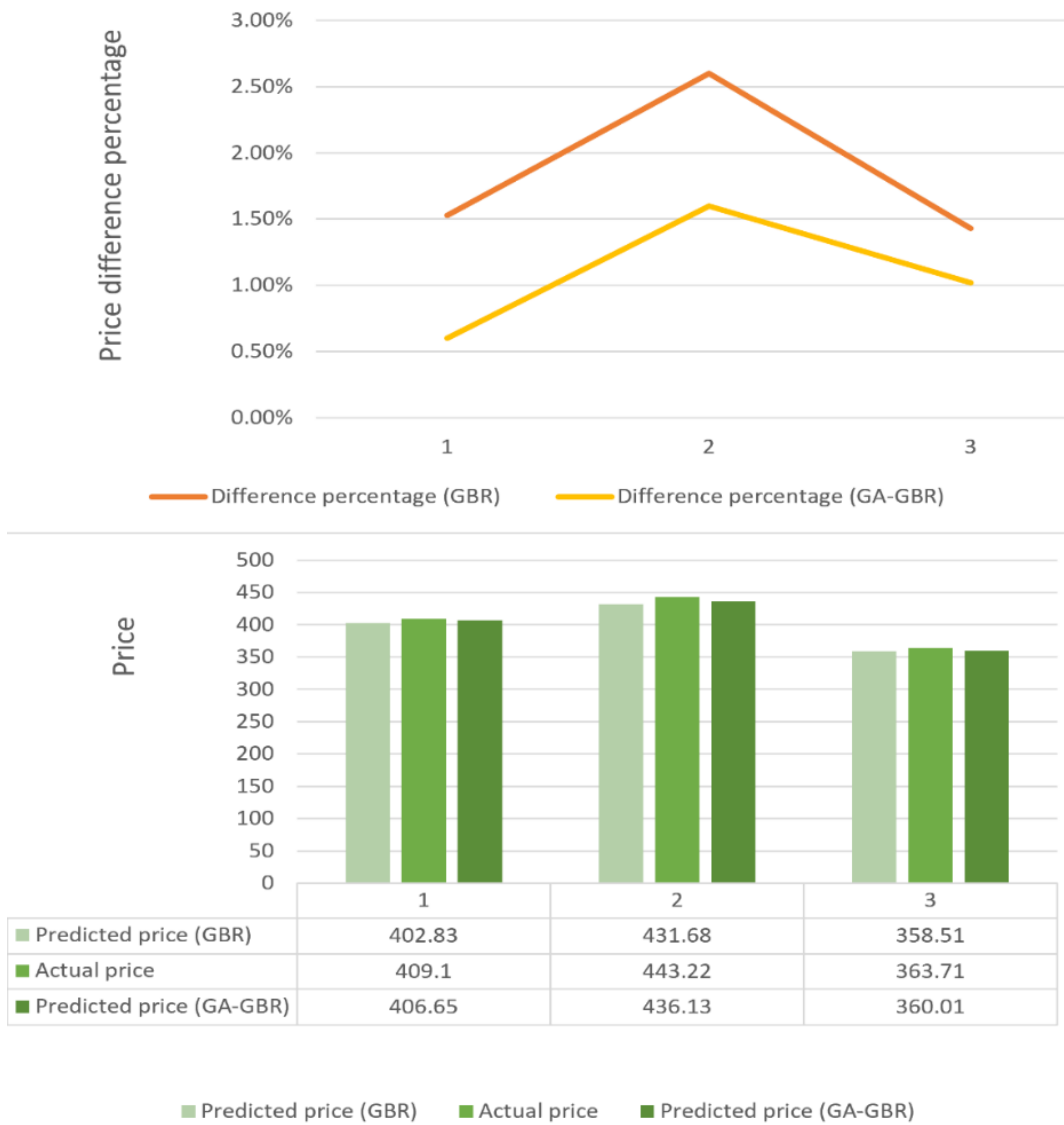

Figure 7: Comparison of the actual price and predicted price by the two models in the three divided building-category-related datasets

To demonstrate the model interpretability at small scales, Table 3 compared the top three features identified by the GA-GBR and GBR models across three building types: tower, mixed (tower and plate) and plate. Feature importance was measured using mean absolute SHAP values for consistency.

Table 3: Feature importance ranking (SHAP) in the three building categories

Tower		Plate and tower		Plate	
GBR feature	**Rank**	**GBR feature**	**Rank**	**GBR feature**	**Rank**
Square	15.08%	*CommunityAverage*	13.79%	*CommunityAverage*	13.89%
TradeTime	12.64%	*Square*	13.72%	*TradeTime*	12.50%
CommunityAvera	12.44%	*TradeTime*	12.70%	*Square*	11.31%
GA-GBR feature	**Rank**	**GA-GBR feature**	**Rank**	**GA-GBR feature**	**Rank**
Elevator	14.58%	*Kitchen*	17.06%	*Square*	24.34%
TradeTime	13.94%	*Floor*	16.04%	*DOM*	14.22%
Floor	13.47%	*DOM*	14.39%	*ConstructionTime*	14.08%

In the tower group, GA-GBR ranked *Elevator* highest (14.58%), replacing *CommunityAverage* (13.79%) in GBR. The importance of *Square* slightly declined (from 15.08% to 14.58%), while *TradeTime* gained relevance (13.94% vs. 12.64%). For mixed buildings, GA-GBR emphasized *Kitchen* (17.06%) and *Floor* (16.04%), features absent in GBR's top rankings, highlighting GA-GBR's ability to capture more detailed spatial and interior characteristics. In the plate group, GA-GBR prioritized active days on market - *DOM* (14.22%) and *ConstructionTime* (14.08%), replacing GBR's focus on *CommunityAverage* and *TradeTime*. This shift suggests GA-GBR better identifies property-specific temporal and lifecycle features.

In addition, *TradeTime* remained a consistently important feature across both models and all subgroups, confirming its general impact on price. Overall, GA-GBR demonstrated stronger adaptability to context-specific factors, thanks to its genetic optimization of feature selection and model tuning.

5.3 Practical implementation of the BIM-AI framework from a building perspective

In this section, the practical implementation of the BIM-AI framework from a building perspective is demonstrated using an IFC-based BIM model from Revit and the GA-GBR model integrated on PyCharm platform. The implementation follows a two-stage process: (1) extracting the value-related information from the IFC property valuation extension, and (2) calculating the property values using the GA-GBR model. The case study utilized a real residential building project from Beijing, with the BIM model originally created in Revit for construction documentation. We selected a representative property transaction from our dataset and manually integrated its 22 distinct features into the BIM model's IFC schema, due to limited time and resources. Subsequently, the necessary value-related information for property valuation was incorporated into the spaces and zones defined in the Revit models which were illustrated in Figure 8, in accordance with the proposed property sets and properties in the extended IFC schema and the input features from the testing dataset.

The syntactic and semantic validation of the IFC models was conducted using the *Solibri* Model Checker, in reference to ISO 10303-11 (ISO, 2014), confirming that there were no missing mandatory entities or incorrect data structures. Lastly, the required value-relevant information was automatically extracted using the developed IFC-based information extraction algorithm.

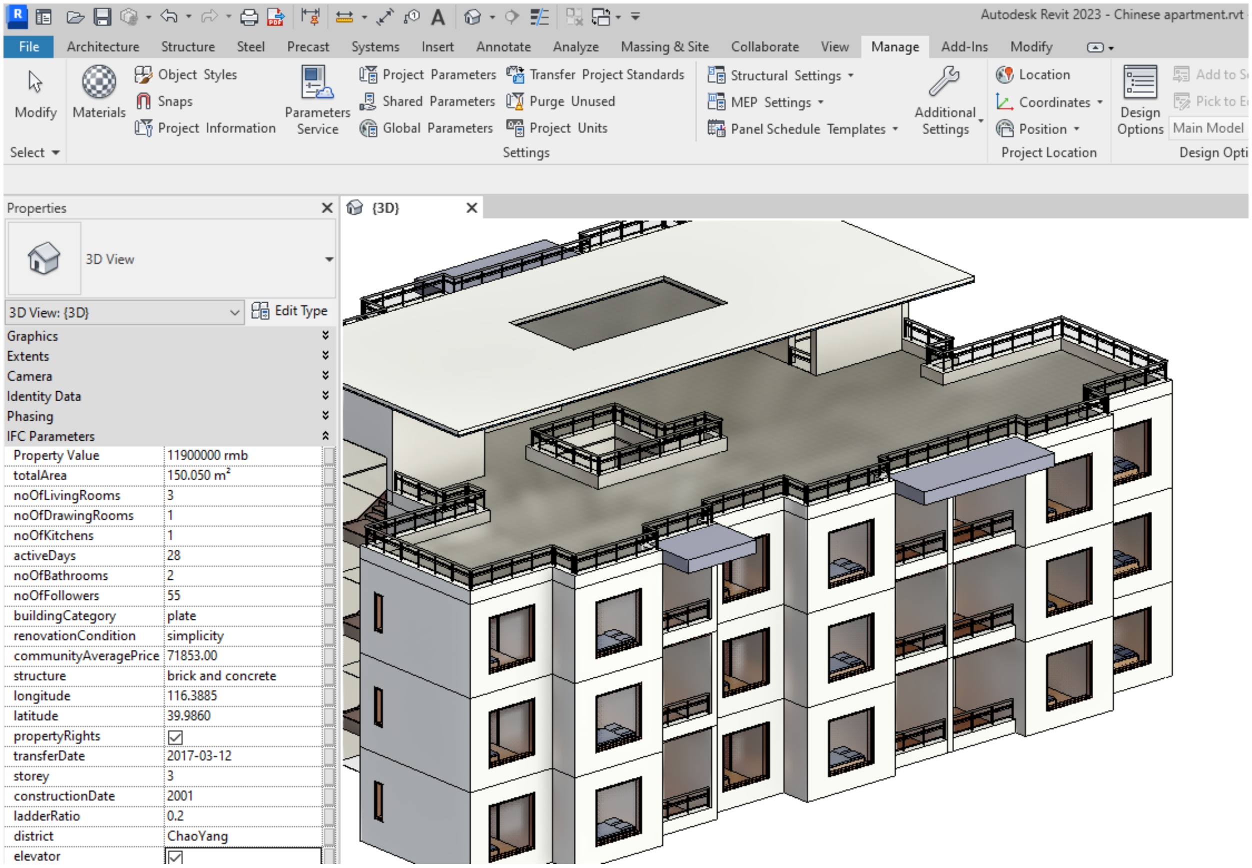

Figure 8: An IFC-based BIM model of the duplex house with required value-related information added according to the 22 input features in the experiment dataset

The integration of value-relevant information from the *Lianjia* Company into the Revit model involved embedding data from 22 distinct input features based on the Chinese dataset. This process utilized shared parameters in the Manage tab's settings panel, guided by the expanded IFC schema. For example, the model incorporated 'brick and concrete' as the *IfcLabel* in the Structure property of the *Pset_PV_Building* set, '50' as the *IfcInteger* in the *activeDays* attribute of the *Pset_PV_Transaction* set, and '11,900,000' as the *IfcReal* in the Property Value facet of the *Pset_PV_Valuation* set. These details extracted using the information extraction algorithm were listed in Table 4 below.

Table 4: The extracted value-related information from the BIM model

Property set name	**Property name**	**Data type**	**Adapted nominal value**
Pset_PV_Building	totalArea	IfcAreaMeasure (150.05)	150.05 ㎡
Pset_PV_Building	noOflivingRooms	IfcInteger (3)	3
Pset_PV_Building	noOfDrawingRooms	IfcInteger (1)	1
Pset_PV_Building	noOfKitchens	IfcInteger (1)	1
Pset_PV_Transaction	transferDate	IfcDateTime (2017-03-12)	2017-03-12
Pset_PV_Transaction	activeDays	IfcInteger (28)	28
Pset_PV_Building	noOfBathrooms	IfcInteger (2)	2
Pset_PV_Transaction	noOfFollowers	IfcInteger (55)	55
Pset_PV_Building	buildingCategory	IfcLabel ('plate')	plate
Pset_PV_Building	renovationCondition	IfcLabel ('simplicity')	simplicity

Pset_PV_Transaction	communityAveragePrice	IfcReal (71853)	71853 RMB
Pset_PV_Building	structure	IfcLabel ('brick and concrete')	brick and concrete
Pset_PV_Building	elevator	IfcBoolean(.F.)	False
Pset_PV_Parcel	longitude	IfcLabel ('116.3885')	116.3885
Pset_PV_Parcel	latitude	IfcLabel ('39.9860')	39.986
Pset_PV_Transaction	propertyRights	IfcBoolean(.T.)	True
Pset_PV_Building	storey	IfcInteger (3)	3
Pset_PV_Building	constructionDate	IfcDateTime (2001)	2001
Pset_PV_Building	ladderRatio	IfcReal (0.2)	0.2
Pset_PV_Parcel	district	IfcLabel ('ChaoYang')	ChaoYang
Pset_PV_Valuation	Property Value	IfcReal (11900000)	11900000 RMB

Next, the BIM-AI system was further implemented on PyCharm. Based on the extracted information from the Chinese BIM model, the house values predicted by the GBR and GA-GBR models are 10,819,100 RMB and 11,022,905 RMB respectively. With the actual value at 11,900,000 RMB, the prediction by the GA-GBR model is more accurate than that by the GBR.

5.4 Experiment findings

The main findings from the empirical experiments of the proposed BIM-xAI system are presented as follows:

1) Predictive performance: The GA-GBR outperformed 11 benchmark AVMs, improving R^2 by 1.3% over the baseline GBR and reducing MAPE to 10.6%. The MAPE improvement represents a 3-4.6% gain over recent state-of-the-art methods (Baur *et al.*, 2023; Konhäuser and Werner, 2024; Tchuente, 2024). The model also preserved its advantage at finer market levels, achieving lower MAPEs across three building-specific categories.

2) Explainability results: Global SHAP analysis revealed *CommunityAverage* as the dominant predictor, followed by *TradeTime* and *Square*. Secondary factors include *DOM* (showing discounts for stale listings), *BathRoom*, and *LivingRoom*. Geographic coordinates provide moderate influence while structural attributes show minimal impact. Local interpretability confirmed expected relationships: prestigious neighborhoods, recent transactions, and larger properties command premiums, while extended market exposure reduces values. When analyzed by different building types (Tower, Mixed, Plate), the model adaptively prioritized context-specific features like *elevator*, *kitchen*, and *stories* over generic predictors, demonstrating its ability to capture market-specific valuation effects. The enhanced accuracy with improved interpretability validates that genetic optimization not only boosts predictive performance but also produces more transparent and market-aligned feature selection across multiple scales.

6. Discussion

This study developed and validated a Genetic Algorithm optimized Gradient Boosting Regressor (GA-GBR) integrated with BIM workflows for property valuation, achieving superior predictive performance compared to 11 benchmark models. By synthesizing evolutionary optimization, ensemble learning and explainable AI, the framework transcends traditional "black box" limitations while adapting to diverse market contexts. The novel framework's impact extends well beyond technical innovation, offering transformative potential across multiple domains.

Research Implications: This work establishes a reproducible methodological framework that advances several research frontiers. It demonstrates how evolutionary algorithms can optimize feature selection in high-dimensional property datasets while maintaining interpretability - a balance rarely achieved in complex AI systems. The framework opens new research avenues including: (1) transfer learning applications to adapt models across geographic markets with minimal retraining; (2) temporal market evolution studies tracking how feature importance shifts over economic cycles; (3) integration with emerging data sources such as IoT sensors for real-time occupancy data, satellite imagery for neighborhood analysis, and social media for sentiment analysis; and (4) extension to related domains like commercial real estate, infrastructure valuation and urban resilience assessment. The open-source nature of the approach encourages reproducibility and collaborative improvement within the research community.

Practice and Industry Applications: The framework transforms professional practice across the real estate value chain. Architects and designers are able to receive immediate feedback on how design choices affect property values, enabling value for investment adjustment during conceptual design rather than costly post-construction modifications. Property appraisers can leverage the tool to augment traditional valuation methods, reducing assessment time significantly while providing defensible and transparent valuations. Banks and mortgage lenders can deploy the system for automated underwriting, reducing loan processing time and improving risk pricing accuracy. Insurance companies benefit from more precise property replacement cost estimates. The standardized IFC-based approach ensures compatibility with existing BIM workflows, requiring minimal disruption to current practices.

Societal and Economic Impact: The framework delivers both substantial economic value and critical societal benefits. Economically, the 1.3% R^2 improvement represents substantial value in high-volume markets. Given that even minor valuation improvements can affect lending decisions on properties worth hundreds of thousands of dollars, this accuracy gain has meaningful implications for risk management and capital allocation. Beyond these commercial benefits, the framework addresses critical societal challenges by democratizing property valuation knowledge and reducing information asymmetry that historically disadvantages first-time homebuyers and minority communities. The transparent SHAP explanations promote fair lending practices and help urban planners understand how public investments (transit, schools, parks) affect property values, enabling more equitable resource allocation. The system's ability to identify gentrification patterns early allows proactive affordable housing interventions, while real-time market monitoring helps policymakers detect speculative bubbles or distressed areas requiring targeted support. This dual impact - improving both market efficiency and social equity - demonstrates how advanced AI can serve both commercial interests and public good, creating value that extends from individual transactions to community-wide benefits.

7. Limitation and future work

Although the GA-GBR framework shows high accuracy and interpretability, several limitations remain. First, the scarcity of BIM-price datasets forced reliance on proxy inputs, potentially reducing real-world fidelity. Second, the generalizability of results remains untested beyond the studied urban datasets. Markets with different regulatory systems, cultural contexts, or rural transaction norms may yield varying performance, highlighting the risk of geographic bias. Third, the framework is inherently dependent on the quality and completeness of BIM data. Inconsistent IFC modelling practices, missing attributes, or heterogeneous data standards could reduce reliability and hinder cross-project comparability. Last, integrating BIM with transaction data raises privacy concerns, as detailed spatial layouts linked to prices could identify individual owners. Future deployments must implement anonymization protocols and bias audits to ensure equitable valuations across diverse communities.

From a methodological perspective, the use of genetic algorithms introduces additional computational cost, which may restrict real-time or large-scale deployment without optimization. Moreover, while SHAP analysis enhances interpretability, it cannot fully capture higher-order feature interactions or eliminate potential model bias arising from skewed training data.

Future research should prioritize developing comprehensive IFC-linked transaction datasets with robust privacy protections. The GA-GBR framework requires enhancement through multi-objective or Bayesian optimization to reduce computational costs. Cloud-based and lightweight deployments could enable real-time design feedback integration with existing BIM platforms. Transfer learning methodologies could facilitate cross-regional model adaptation without extensive retraining. Additionally, user-centered explainable AI interfaces must be developed to effectively communicate valuation insights to diverse stakeholders. These tools should provide transparent, interpretable outputs for architects, developers and policymakers, thereby supporting data-driven decision making in the built environment.

8. Conclusion

This study proposed and validated a transparent, explainable and scalable property valuation framework by integrating BIM with a Genetic Algorithm optimized Gradient Boosting Regressor (GA-GBR). The framework addresses critical challenges in traditional automated valuation models, including overfitting, feature redundancy and limited interpretability. By incorporating GA for feature selection and hyperparameter tuning, the GA-GBR model demonstrated superior performance compared to 11 benchmark machine learning models across datasets at multiple scales. In particular, the model achieved consistent improvements in regression accuracy metrics, including MAE, MAPE, MSE and R^2, while enhancing interpretability through SHAP based xAI techniques.

Empirical results revealed that GA-GBR effectively promotes feature independence, capturing localized market dynamics and structural characteristics overlooked by traditional models. Through systematic experimentation on large-scale datasets and stratified building-type subgroups, the proposed model exhibited robust adaptability to varying analytical scales. Furthermore, the integration of xAI methods provided both global and local explanations of predictions, supporting its use in decision-making processes for developers, planners and policymakers.

The proposed BIM-xAI system lays a foundation for scalable, transparent and design-responsive valuation tools in the building sector. As the digital transformation of the built environment continues, the framework holds promises to improve valuation accuracy, design feedback and stakeholder trust.

The research makes contributions as follows:

- **Novel Optimization Approach:**

  A genetic algorithm approach that simultaneously optimizes feature selection and hyperparameters in gradient boosting for property valuation, demonstrating measurable improvements over traditional approaches.

- **Hierarchical Explainability Framework for Property Valuation:**

  Combines global (dataset-level) and local (building-type-specific) SHAP analysis to reveal both universal and context-specific valuation drivers. The systematic application across different building typologies, not just SHAP usage.

- **IFC-to-Valuation Pipeline Integration:**

  A specific technical framework for extracting valuation-relevant features from IFC-based BIM data and feeding them directly into ML models.

Overall, this research contributes a replicable, scalable and explainable valuation framework that can inform both academic inquiry and industry application, supporting more transparent and intelligent property market assessments.

Reference

Ahn, J.J., Byun, H.W., Oh, K.J. and Kim, T.Y. (2012), “Using ridge regression with genetic algorithm to enhance real estate appraisal forecasting”, *Expert Systems with Applications*, Vol. 39 No. 9, pp. 8369–8379, doi: https://doi.org/10.1016/j.eswa.2012.01.183.

Arcuri, N., De Ruggiero, M., Salvo, F. and Zinno, R. (2020), “Automated Valuation Methods through the Cost Approach in a BIM and GIS Integration Framework for Smart City Appraisals”, *Sustainability*, Vol. 12 No. 18, p. 7546, doi: https://doi.org/10.3390/su12187546.

Baur, K., Rosenfelder, M. and Lutz, B. (2023), “Automated real estate valuation with machine learning models using property descriptions”, *Expert Systems with Applications*, Vol. 213, p. 119147, doi: https://doi.org/10.1016/j.eswa.2022.119147.

Burgess, G., Jones, M. and Muir, K. (2018), “BIM in the UK house building industry: opportunities and barriers to adoption”, *University of Cambridge*, Cambridge, available at: https://www.cdbb.cam.ac.uk/.

Doan, Q.C., Chen, C., He, S. and Zhang, X. (2024), “How urban air quality affects land values: Exploring non-linear and threshold mechanism using explainable artificial intelligence”, *Journal of Cleaner*

Production, Vol. 434, p. 140340, doi: https://doi.org/10.1016/j.jclepro.2023.140340.

Harrison, D. and Rubinfeld, D. (1978), "The Boston Housing Dataset", available at: https://www.kaggle.com/datasets/schirmerchad/bostonhoustingmlnd (accessed 13 September 2025).

Hong, Y. and Guo, F. (2025), "A framework of BIM-IoT application in construction projects through multiple case study approach", *Journal of Building Design and Environment*, Vol. 3 No. 1, p. 202439, doi: https://doi.org/10.70401/jbde.2025.0004.

Iban, M.C. (2022), "An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants", *Habitat International*, Vol. 128, p. 102660, doi: https://doi.org/10.1016/j.habitatint.2022.102660.

Isikdag, U., Horhammer, M., Zlatanova, S., Kathmann, R. and Van Oosterom, P. (2015), "Utilizing 3D building and 3D cadastre geometries for better valuation of existing real estate", in FIG (Ed.), *Proceedings FIG Working Week 2015 "From the Wisdom of the Ages to the Challenges of Modern World"*, International Federation of Surveyors (FIG), Sofia, pp. 1–18.

ISO. (2014), "Industrial automation systems and integration — Product data representation and exchange — Part 11: Description methods: The EXPRESS language reference manual", available at: https://www.iso.org/standard/38047.html (accessed 19 September 2025).

Kanan, H.R., Faez, K. and Taheri, S.M. (2007), "Feature Selection Using Ant Colony Optimization (ACO): A New Method and Comparative Study in the Application of Face Recognition System", in Petra Perner (Ed.), *Industrial Conference on Data Mining*, Springer Berlin Heidelberg, pp. 63–76, doi: https://doi.org/10.1007/978-3-540-73435-2_6.

Konhäuser, K. and Werner, T. (2024), "Uncovering the financial impact of energy-efficient building characteristics with eXplainable artificial intelligence", *Applied Energy*, Vol. 374, p. 123960, doi: https://doi.org/10.1016/j.apenergy.2024.123960.

Lewis, O.M., Ware, J.A. and Jenkins, D. (1997), "A novel neural network technique for the valuation of residential property", *Neural Computing & Applications*, Vol. 5 No. 4, pp. 224–229, doi: https://doi.org/10.1007/BF01424227.

Liu, X.S., Deng, Z. and Wang, T.L. (2011), "Real estate appraisal system based on GIS and BP neural network", *Transactions of Nonferrous Metals Society of China*, Vol. 21 No. Supplement 3, pp. s626–s630, doi: https://doi.org/10.1016/S1003-6326(12)61652-5.

McCluskey, W.J., McCord, M., Davis, P.T., Haran, M. and McIlhatton, D. (2013), "Prediction accuracy in mass appraisal: A comparison of modern approaches", *Journal of Property Research*, Vol. 30 No. 4, pp. 239–265, doi: https://doi.org/10.1080/09599916.2013.781204.

Muccio, E. and Cannatella, D. (2026), *Mapping Real Estate Values: A Semi-Systematic Literature Review of Spatial Evaluation Methods and Approaches*, *Lecture Notes in Computer Science*, Vol. 15893 LNCS, Springer Nature Switzerland, doi: https://doi.org/10.1007/978-3-031-97645-2_16.

Natekin, A. and Knoll, A. (2013), "Gradient boosting machines, a tutorial", *Frontiers in Neurorobotics*, Vol. 7, doi: https://doi.org/10.3389/fnbot.2013.00021.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. and French, N. (2003), "Real estate appraisal: a review of valuation methods", *Journal of Property Investment & Finance*, Vol. 21 No. 4, pp. 383–401, doi: https://doi.org/10.1108/14635780310483656.

Rafiei, M.H. and Adeli, H. (2016), "A Novel Machine Learning Model for Estimation of Sale Prices of Real Estate Units", *Journal of Construction Engineering and Management*, Vol. 142 No. 2, doi: https://doi.org/10.1061/(asce)co.1943-7862.0001047.

RICS. (2017), "The Future of Valuations", available at: https://www.rics.org/content/dam/ricsglobal/documents/to-be-sorted/future_of_valuations_insights_paper_rics.pdf (accessed 19 September 2025).

Su, T., Li, H. and An, Y. (2021), "A BIM and machine learning integration framework for automated property valuation", *Journal of Building Engineering*, Vol. 44, p. 102636, doi: https://doi.org/10.1016/j.jobe.2021.102636.

Sun, Y. (2019), "Real estate evaluation model based on genetic algorithm optimized neural network", *Data Science Journal*, Vol. 18 No. 1, pp. 1–9, doi: https://doi.org/10.5334/dsj-2019-036.

Tarasov, S. and Dessoulavy-Śliwiński, B. (2024), "Algorithm-Driven Hedonic Real Estate Pricing – an Explainable Ai Approach", *Real Estate Management and Valuation*, Vol. 33 No. 1, pp. 22–34, doi: https://doi.org/10.2478/remav-2025-0003.

Tchuente, D. (2024), "Real Estate Automated Valuation Model with Explainable Artificial Intelligence Based on Shapley Values", *Journal of Real Estate Finance and Economics*, doi: https://doi.org/10.1007/s11146-024-09998-9.

Valier, A. (2020), "Who performs better? AVMs vs hedonic models", *Journal of Property Investment and Finance*, Vol. 38 No. 3, pp. 213–225, doi: https://doi.org/10.1108/JPIF-12-2019-0157.

Wilkinson, S.J. and Jupp, J.R. (2016), "Exploring the value of BIM for corporate real estate", *Journal of Corporate Real Estate*, Vol. 18 No. 4, pp. 254–269, doi: https://doi.org/10.1108/JCRE-11-2015-0040.

Won, J., Lee, G. and Cho, C. (2013), "No-Schema Algorithm for Extracting a Partial Model from an IFC Instance Model", *Journal of Computing in Civil Engineering*, Vol. 27 No. 6, pp. 585–592, doi: https://doi.org/10.1061/(asce)cp.1943-5487.0000320.

El Yamani, S., Ettarid, M. and Hajji, R. (2019), "BIM potential for an enhanced real estate valuation approach based on the hedonic method", *3rd International Conference on Building Information Modelling (BIM) in Design, Construction and Operations*, Séville, Spain, available at: https://orbi.uliege.be/handle/2268/250013 (accessed 19 September 2025).