

حسن خليفة

Paper Clarivision System-2.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:404451967

Submission Date

Aug 2, 2026, 9:20 AM GMT+3

Download Date

Aug 2, 2026, 3:15 PM GMT+3

File Name

Paper Clarivision System-2.docx

File Size

3.1 MB

28 Pages

10,588 Words

60,536 Characters

49% detected as AI

The percentage indicates the combined amount of likely AI-generated text as well as likely AI-generated text that was also likely AI-paraphrased.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.



40 AI-generated only **49%**

Likely AI-generated text from a large-language model.



0 AI-generated text that was AI-paraphrased **0%**

Likely AI-generated text that was likely revised using an AI-paraphrase tool or word spinner.

Disclaimer

Our AI writing assessment is designed to help educators identify text that might be prepared by a generative AI tool. Our AI writing assessment may not always be accurate (i.e., our AI models may produce either false positive results or false negative results), so it should not be used as the sole basis for adverse actions against a student. It takes further scrutiny and human judgment in conjunction with an organization's application of its specific academic policies to determine whether any academic misconduct has occurred.

Clarivision AI: An Explainable Hybrid Framework for Residential Property Valuation and Similar Property Retrieval Using Structured Data and ResNet50-Based Deep Visual Features

Hassan Mohammed Ahmed Alamer Khalifa

ORCID ID :0009-0006-0611-1563

Collage of Informatics, Midocean University, Comoros

Email: hassanalamer2019@gmail.com

Supervisor Name: Dr / Amjad Qtaish

Abstract

Objectives: This study develops Clarivision AI, an explainable hybrid framework for residential property valuation that integrates structured property attributes with deep visual features extracted using ResNet50 and enhances model interpretability through Explainable Artificial Intelligence (XAI).

Study Location: The study was conducted using the Real Estate with Images dataset, comprising 15,474 residential property records with corresponding structured attributes and property images.

Methodology: The dataset was divided into 80% training data and 20% testing data using a fixed random seed of 42. Property prices were transformed using $\log(1 + \text{price})$ to reduce target skewness. Structured property attributes were preprocessed and combined with deep visual embeddings extracted using ResNet50 to construct hybrid features. An XGBoost regression model was then developed for property valuation, while SHAP was used to interpret the contributions of the structured features. A similar-property retrieval mechanism based on cosine similarity and an interactive decision-support interface developed using Streamlit were also incorporated.

Results: The proposed hybrid model achieved promising predictive performance, with an MAE of 0.20985, an RMSE of 0.29112, and an R^2 value of 0.67375. The MAE and RMSE values were calculated on the log-transformed property price scale. After back-transforming the predictions to the original price scale, the model achieved an MAE of **151,078.31** and an RMSE of **238,044.55**, expressed in the original price units of the dataset. The results indicate that the model explained approximately 67.4% of the variance in the log-transformed property prices. SHAP provided explanations of the contributions of the structured property features, while the developed platform supported similar-property retrieval and interactive visualization.

Recommendations: To include more structured and visual property attributes. To test framework on datasets from other countries. To apply more XAI methods. To extend the framework with property classification and cloud deployment of AI models.

Keywords: Real Estate Valuation, Hybrid Model, Explainable Artificial Intelligence, XGBoost, ResNet50, Clarivision AI

الكلمات المفتاحية: التقييم العقاري، النموذج الهجين، كلاريفيجن، الذكاء الاصطناعي القابل للتفسير

1. Introduction.

The real estate industry is critical to the growth of an economy and in making decisions. The accurate valuation of residential properties is critical for the buyers, sellers, financial institutions, and the government organizations. Nonetheless, the conventional methods of property valuation are majorly dependent on human knowledge and the structured data, hence making the process subjective, inconsistent and time-consuming because of the dependence on the quality of the information regarding the property.

However, recent developments in Artificial Intelligence (AI) and Machine Learning (ML) have greatly enhanced the accuracy of the Automated Valuation Models (AVMs). Some of the latest prediction algorithms like Random Forest, XGBoost, and Deep Learning have shown great performance in terms of their ability to identify non-linear relations between the attributes of the properties and market prices. However, most Automated Valuation Models (AVMs) still depend mainly on the structured attributes of the property while ignoring the visual information available from residential property pictures. Moreover, many of the AI valuation systems remain black boxes hence reducing the understanding of the valuation process.

In order to overcome the limitations described above, this paper puts forth Clarivision AI, a novel explainable hybrid approach for fusing structured property attributes with deep visual representations obtained via the ResNet50 convolutional neural network. Structured numerical property attributes and deep visual representations are fused in the hybrid learning architecture based on XGBoost, and SHapley Additive explanations (SHAP) are used for explaining predictions made by the model. In addition, Clarivision AI includes a Similar Property Retrieval component, where visually similar residential property listings are retrieved via cosine similarity measure over deep image embeddings.

Even though recent research showed significant improvements in residential property valuation by applying machine learning, multimodal learning, and XAI techniques, most of the available solutions consider these components separately. There are very few papers that introduce the framework combining all the discussed aspects into one decision-support solution. It is exactly this research gap that motivates the introduction of Clarivision AI hybrid framework.

The key contribution of the study is the creation of an integrated explainable decision support system which integrates hybrid property valuation, visual similarity search, model interpretability, and interactive reporting in one intelligent system. The experiment shows that the developed framework is not only better at valuating residential properties but also increases the transparency, usability, and trustworthiness of real estate decision making.

2. Research Problem

Although there have been tremendous advances made in the application of Artificial Intelligence (AI) in Automated Valuation Models (AVMs), some drawbacks continue to affect the accuracy, explainability, and usability of the available systems for residential properties. For instance, most of the valuation systems consider structural properties of the house such as size and number of rooms but ignore other visual cues about the property's appearance. Moreover, most of the existing deep learning-based valuation models are black box systems that make it hard for the user to comprehend how decisions are made by the valuation system. This limitation calls for the development of Explainable Artificial Intelligence (XAI) techniques to increase model explainability and user confidence. Although these models are capable of predicting prices accurately, they do not offer adequate explanations about their decisions, thus limiting the trust users have in such models. Additionally, previous research work focuses only on property price prediction without incorporating other decision support features like property matching, interactive visualization, and report generation.

Thus, it becomes necessary to have a hybrid explainable valuation system, which could incorporate structured property features, deep vision-based visual features, predictability and explainability, similar property searches, and decision support in an all-in-one intelligent system. Clarivision AI has been designed to overcome these drawbacks to provide precise and understandable valuation of residential properties.

Hence, the purpose of this study is to overcome the shortcomings of current residential property valuation systems and design an integrated Explainable Artificial Intelligence system, which would help with accurate predictions and decision making.

3. Importance of the Study.

2.1 Academic Significance

This study helps us better understand how to use Artificial Intelligence to value homes. It looks at a new way of combining information about a property, like its size and location, with pictures of the property to get a more accurate value. The study also uses a technique called SHAP to make it clearer how the computer is making its predictions, by showing which pieces of information are most important in deciding the value of a home.

2.2 Practical Significance

The above described Clarivision AI framework offers a comprehensive decision-making tool for assessing residential properties. It incorporates prediction, interpretability of models, similar property search, visualization and PDF reporting functionalities to help real estate experts, investors, banks, and other relevant parties make better valuation decisions.

4. Study objectives

Key Goals of the Study:

- Develop a hybrid residential property valuation model by integrating structural and visual property features using ResNet50.
- Enhance the accuracy of residential property valuation through an XGBoost regression model.
- Improve the transparency and interpretability of the models using Shapley Additive Explanations (SHAP).
- Create a Similar Property Retrieval module that fetches properties from a selected city and uses cosine similarity on ResNet50 visual embeddings to identify the most visually comparable properties for comparative analysis.
- Build an interactive decision support platform, Clarivision AI, incorporating residential property valuation capabilities.

5. Theoretical Literature of the Study.

Real estate valuation has changed remarkably during recent decades from conventional valuations methods towards intelligent Artificial Intelligence (AI) approaches. The conventional approaches to real estate valuation heavily rely on professional appraisers who calculate property values using their market experience and property characteristics. However, despite the widespread use of these valuation methods, they are typically time-consuming and inconsistent.

The rapid advancement in Machine Learning (ML) has resulted in remarkable improvements in Automated Valuation Model (AVM) performance. Such ML algorithms as Decision Trees, Random Forest, Gradient Boosting Machines, and XGBoost have shown remarkable ability to model complicated nonlinear relations between property characteristics and market prices. They have been constantly showing superiority in capturing complicated nonlinear relations between property characteristics and market prices which resulted in more accurate property valuations than conventional statistical approaches.

In more recent times, however, Deep Learning approaches have brought new possibilities to the domain of property valuation in that they use visual information derived from property pictures. ResNet50 CNNs, for example, have demonstrated outstanding capability in capturing architectural specifics, visual appearance of the house, its external state, which is impossible to encode in numeric form. Structured data combined with deep visual features can thus be viewed as an effective solution to improving the accuracy of residential property valuation models.

However, even with the development of new solutions, many AI-based valuation models lack interpretability. In order to solve this problem, a new approach called Explainable Artificial Intelligence (XAI) was proposed. Being an approach that gives transparent explanations of model outputs, it makes models easier to interpret. One of the most commonly used XAI techniques is SHapley Additive exPlanations (SHAP). This transparency increases user confidence and makes AI-driven models easier to adopt. In terms of residential property valuation, SHAP allows users to find out how much every property attribute contributes to the predicted price of the house.

Thus, recent research efforts have concentrated on designing hybrid explainable valuation models by combining structured information with deep visual features. Extending this research line, the proposed Clarivision AI model is designed to integrate structured information, deep visual features using ResNet50 architecture, prediction using XGBoost algorithm, SHAP explanations, Similar Property Retrieval, and decision support functionality in an intelligent platform.

6. Review of Previous Studies

Recent advances in Artificial Intelligence (AI) and Machine Learning (ML) have transformed residential property valuation, moving beyond traditional statistical methods. The reviewed literature can be categorized into four main research directions: (1) structured-data valuation models, (2) deep learning and multimodal valuation, (3) explainable artificial intelligence, and (4) intelligent decision-support systems. While these approaches have collectively enhanced valuation accuracy, transparency, and practical usability, they vary significantly in data modalities, model architectures, explainability techniques, and system functionalities.

6.1 Structured-Data Valuation Models

Machine learning has been used to value homes for a while now. At first, it mainly looked at things like how big a house is, where it's located, how many bedrooms and bathrooms it has, and information about when it was sold. But some researchers found that adding information about the area around the house can make these models better. For example, Wang and his team came up with a new way to value homes in Shenzhen, China, using a special kind of neural network that takes into account where the house is located. They showed that this can be a really good way to figure out how much a house is worth. Around the same time, Mora-Garcia and his team tried out a few different machine learning algorithms to see which one could best predict housing prices. They found that these new methods can work just as well as the old ways of doing things, and sometimes even better. This is a big deal, because it means that machine learning can be a powerful tool for figuring out how much a house is worth.

Researchers Balila and Shabri took a closer look at how different machine learning algorithms can predict property prices in Dubai. They used a lot of structured data about properties to see what works best. What they found out was that getting the data ready, picking the right features, and changing them into something useful is really important. They also compared different models to see which one is the best. One thing that stood out was that the size of the property is a big deal when it comes to figuring out its value. But there were some limitations to their study - they only used structured data and didn't look at pictures of properties or combine different types of features. This means there's still more to explore, like how pictures and other types of data can help predict property prices.

Generally, models that use structured data can give us useful and easy-to-understand estimates of a property's value. But they might not be able to fully capture the visual aspects of a property, like how it looks from the outside, the quality of the interior, the condition it's in, and the level of finishing. They also can't account for other things that you can see in pictures. This is why people started working on ways to value properties using images and combining different types of data.

6.2 Deep Learning and Multimodal Property Valuation

Recent studies have increasingly incorporated property images into automated valuation models. Deep learning techniques enable visual representations to be extracted directly from property images and combined with structured property information.

Yazdani and Raissi (2023) investigated the use of self-supervised vision transformers for real estate valuation, highlighting the potential of visual representations for improving property valuation. Hasan et al. (2024) proposed a multimodal deep learning approach for house price prediction by combining property images with structured information. Similarly, Deng et al. (2025) developed a multimodal machine learning pipeline that integrated multiple property information sources and explainable artificial intelligence techniques.

These studies show that using multiple ways of learning can give us more information about properties than just looking at numbers. But most of the studies we looked at were trying to make predictions better, and they didn't usually include things like finding similar properties based on pictures, interactive tools to help us understand the data, and automatic reports all in one place where we value properties.

Researchers like Huang and others found that using multiple types of data, like numbers and pictures, can help figure out how much a property is worth. They think it's really important to combine different kinds of information to get a better sense of a property's value. The problem is, the systems we have now to do this aren't very transparent or easy to use in real-life situations, so we need to make them better.

6.3 Explainable Artificial Intelligence in Property Valuation

As models for machine learning become more complicated, model transparency becomes crucial for assessing properties. The methods of Explainable Artificial Intelligence (XAI) could help detect the influential factors for property value estimation.

In their study, Gunes (2024) analyzed model-agnostic interpretability methods of the residential property valuation and revealed the significance of explanation of model predictions. Moreover, Maselli and Nesticò (2025) stressed the importance of explaining AI models in terms of property valuation. Furthermore, the methods of XGBoost and Shapley values were used by Kee and Ho (2025) to explore the influence of property characteristics on predicted values.

This research proved the possibility of enhancing the transparency of the automated valuation models. Nevertheless, explainability was considered as an additional tool of analysis instead of a part of the whole decision support system, which includes price prediction, visual similarity retrieval, interactive visualization, and automated reporting.

6.4 Intelligent Decision-Support Systems

Recent research has explored intelligent systems that extend beyond numerical property price prediction. Geerts et al. (2025) examined the use of large language models in real estate appraisal, showcasing how language-based AI can support appraisal-related reasoning. However, their focus was primarily on language model capabilities rather than a comprehensive multimodal valuation system.

Rodriguez-Serrano (2025) studied prototype-based learning for real estate valuation, highlighting the benefits of example-based explanations. While prototype-based methods enhance interpretability, they do not offer image-based Similar Property Retrieval or a fully interactive decision-support environment.

Overall, the literature shows significant progress in structured-data valuation, multimodal learning, explainable AI, and intelligent decision-support systems. Yet, these capabilities are often investigated independently. Few studies integrate price prediction, image-based Similar Property Retrieval, model explainability, interactive visualization, and automated reporting within a single intelligent framework. This gap highlights the need for a more integrated and explainable approach to residential property valuation.

Research Direction	Dataset / Data Type	Modality	Model Type	Explainability	Similar Property Retrieval	Main Limitation
Structured-Data Valuation	Residential property attributes and geographic information	Structured data	GNNWR, ML, and ensemble models	Not emphasized	No	Limited consideration of property visual characteristics
Deep Learning and Multimodal Valuation	Property images combined with structured property data	Visual and multimodal data	Vision Transformers, CNN-based models, and hybrid ML models	Limited or study-dependent	No	Mainly focused on predictive performance rather than integrated decision support
Explainable AI in Property Valuation	Primarily structured property data	Structured data	Interpretable and model-agnostic ML approaches	SHAP and other XAI methods	No	Focused mainly on explaining predictions without visual retrieval or integrated services
Intelligent Decision-Support Systems	Structured and textual real-estate information	Structured and textual data	LLM-based and intelligent appraisal approaches	Model-dependent explanations	Limited	Limited integration of visual features and comprehensive valuation services
Clarivision AI — This Study	Real Estate with Images dataset containing 15,474 residential properties	Structured and visual data	ResNet50 visual embeddings + XGBoost regression	SHAP for structured feature contributions	Yes — city-filtered visual retrieval using cosine similarity	Explainability is currently focused primarily on structured features

Figure 1: Critical Comparison of Previous Research Directions (Source: Developed by the author based on the reviewed literature)

6.5 Critical Synthesis

According to the comparative analysis, the structured data-based valuation models have great predictive ability, although they tend to use only the numeric and geographic features of the properties without taking into account the visual properties. The deep learning and multimodal methods try to overcome this challenge by introducing the property images into the structure. Still, some of those studies emphasize predictive accuracy. The XAI approach tries to increase the level of transparency and explanation of the contribution of the features; still, those studies usually do not consider similar-property retrieval and decision-making assistance. The intelligent decision support studies go beyond the simple price prediction and create the valuation systems; however, they can operate mostly based on the structured or text-based data and ignore the visual features. Hence, it could be concluded that an integrated system is required for hybrid valuation, explanation, similar-property retrieval, visualization, and report generation.

6.6 Research Gap

From the reviewed literature, there has been a lot of success in the field of valuation of residential properties through the use of machine learning techniques for structured data, deep learning, multimodal methods, and Explainable Artificial Intelligence (XAI). Unfortunately, only very few of the studies in the literature seem to combine the use of structured property features, deep visual features, hybrid property valuation, SHAP-based explainability, similar property retrieval, visualization, and automatic reporting.

In order to fill this gap, this study proposes Clarivision AI; a hybrid explainable AI framework that combines the use of structured property features together with visual features generated using ResNet50, through an XGBoost regression algorithm. This framework also includes SHAP-based explanation, city-filtered Similar Property Retrieval using cosine similarity, visualization, and automatic PDF reporting.

7. Methodology.

This study adopted an experimental research methodology to develop, implement, and evaluate Clarivision AI, an explainable hybrid framework for residential property valuation. The proposed methodology integrates structured property attributes with deep visual representations extracted from property images to support residential property price prediction and improve the transparency of model outputs. The methodology consists of five sequential phases: (1) data collection and preprocessing, (2) deep visual feature extraction, (3) hybrid model development and evaluation, (4) model interpretability, and (5) similar-property retrieval and decision-support system development.

7.1 Dataset Source and Characteristics.

This study used the **House Prices and Images – SoCal** dataset, publicly available on Kaggle (ted8080, n.d.). The dataset includes **15,474 residential property records** from Southern California; each linked to one exterior property image via the **image_id** attribute. It consists of eight columns: **image_id**, **street**, **city**, **n_citi**, **bed**, **bath**, **sqft**, and **price**.

For the predictive model, four structured property features were selected: **city**, **number of bedrooms**, **number of bathrooms**, and **property area (sqft)**. The **property price** served as the target variable. The **n_citi** variable was used as the numerical representation of the categorical **city** attribute. The **image_id** linked each property record to its corresponding image. The **street** attribute was excluded from the model due to its high-cardinality and nature as a textual identifier.

The dataset was sourced from a publicly available Kaggle page with an unknown license status. Therefore, it was used solely for academic research and experimental purposes, with proper acknowledgment of the original source.

7.2 Structured Data Preprocessing.

We took a close look at the dataset to see if there were any missing or duplicate records that needed to be dealt with before we started building our model. Luckily, when we checked all eight attributes of the dataset, we found that none of them were missing any values. We also didn't find any duplicate records, which was a relief. Because of this, we didn't need to do anything about missing values or get rid of any records - everything was ready to go as it was.

When it comes to figuring out how much a house is worth, there are a few key things to look at. We picked four main features to help us come up with a good estimate: where the house is located in the city, how many bedrooms it has, how many bathrooms it has, and how big the house is in square feet. These things are important because they tell us about the house's location, size, and what it has to offer, which are all big factors in deciding how much a house is worth.

We used a different version of the city information, called **n_citi**, instead of the original one. This is because **n_citi** is a number, which the machine-learning model can understand. The **image_id** was kept so we can match each property with its picture. On the other hand, the street information was left out because it's a long piece of text with many different values, and we didn't use it to figure out the value of the properties.

We took a closer look at some properties that seemed a bit off. We checked to see if the number of bedrooms, bathrooms, property area, and price made sense together. Two properties really stood out as being unlikely - one had 36 bathrooms and the other had 13, but they both only had four bedrooms and weren't very big. We decided to remove those two from our dataset. However, we kept other properties with high values if they seemed to match up with what you'd expect from a large residential property.

We took a closer look at how the original property prices were spread out, using something called the skewness statistic. What we found was that the prices were skewed to the right, with a skewness value of 1.349 - this means that there were some really high prices that were pulling the average up. To make the numbers more stable and easier to work with, we used a logarithmic transformation, which is like a math formula that helps reduce the impact of those really high prices. We added 1 to the price, and then took the log of that - this helped to even out the distribution and make it more normal. We used this transformed price variable when we were training and testing our model, because it made the whole process more reliable.

We were left with 15,472 property records after checking the data for quality and removing two records that didn't seem right. These records will be used to develop our model further.

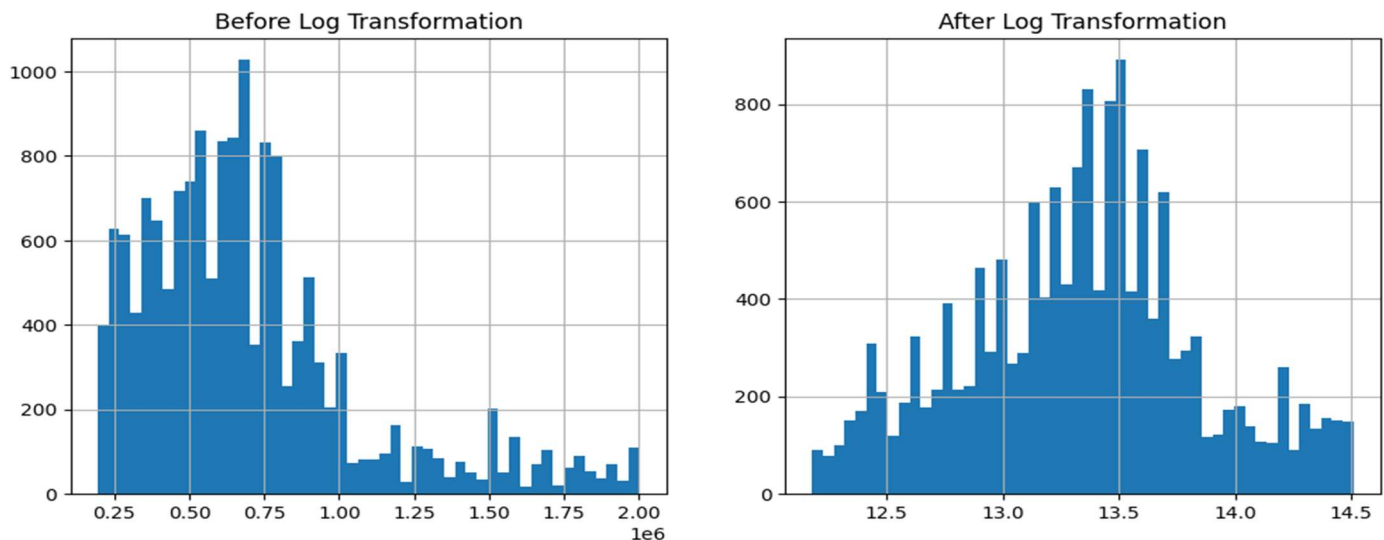


Figure 2 Distribution of Property Prices Before and After Log Transformation.

7.3 Image Preprocessing and Deep Visual Feature Extraction

Images were loaded and resized to a fixed size of 224 x 224 pixels, with three RGB color channels. Images were converted to a numerical array and a batch dimension was added to generate an input tensor with the dimensions (1, 224, 224, 3).

Preprocessing was carried out using the `preprocess_input` function from the TensorFlow/Keras ResNet50 library. The preprocessing process carries out the transformation of the input according to the requirements of the pre-trained ResNet50 model, providing uniformity between property images and the model based on the ImageNet dataset.

A pre-trained ResNet50 convolutional neural network served as a fixed deep visual feature extractor. The model was pre-initialized using the weights of the ImageNet dataset, and the layers responsible for classifying images were excluded using the parameter `include_top=False`. Global average pooling was used by setting `pooling='avg'`, generating a 2048-dimensional fixed visual embedding of each property image. No further training of the layers was done for this pre-trained ResNet50 model.

The extracted visual embeddings represented high-level visual characteristics of the residential properties and were subsequently combined with the selected structured property features to construct the hybrid feature representation used in the valuation model.

7.4 Hybrid Feature Construction and XGBoost Model Development

To get a sense of how stable our model is and to have something to compare it to, we did something called five-fold cross-validation on our training data. We had already set aside 20% of our data as a test set, which we didn't use for this part. When we did the cross-validation, we made sure to shuffle the data and used a fixed random seed, which was 42. We looked at two kinds of models: XGBoost and Random Forest regression. Both of these models used the same features, like the number of cities, bedrooms, bathrooms, and square feet. We didn't use the hold-out test set for the cross-validation, we saved it for when we finally evaluated our hybrid model. This way, we could get a good idea of how well our model would work on data it had never seen before.

When we look at how well the XGBoost model did, we see it got a mean cross-validation score of 0.6620, with a standard deviation of 0.0118. On the other hand, the Random Forest model, which we used as a baseline, did a bit better with a mean score of 0.6765 and a standard deviation of 0.0101. What's interesting here is that both models were pretty consistent across the different parts of the data they were trained on, which is shown by their relatively low standard deviations. In the end, the Random Forest model came out slightly on top in this experiment, at least in terms of its mean score. This suggests that, for this particular set of data, the Random Forest approach might be a better choice for making predictions.

7.5 Model Interpretability Using SHAP

To make the proposed way of figuring out property values clearer and easier to understand, a method called Shapley Additive Explanations (SHAP) was used to see how different parts of a property's information affect the predicted price. The visual features found using ResNet50 are complex and can't be easily understood by people, so the analysis focused on the parts of the property's information that are more straightforward, like its attributes.

We trained a separate XGBoost regression model using some key property features. These features included a numerical representation of the city, the number of bedrooms, the number of bathrooms, and the property area in square feet. The goal was to predict the log-transformed property price. To do this, we split our dataset into two parts: training data and testing data. The training data made up 80% of the dataset, while the testing data made up the remaining 20%. We used a fixed random seed of 42 to ensure the split was consistent. This resulted in 12,379 property observations for training and 3,095 for testing.

The structured XGBoost model was configured with 300 estimators, a maximum tree depth of 6, a learning rate of 0.05, and a fixed random seed of 42. After training the model using the structured training data, SHAP values were calculated for the hold-out test set using the TreeExplainer method. TreeExplainer was selected because it is specifically designed to efficiently explain predictions generated by tree-based machine-learning models, including XGBoost.

The SHAP results came back with a matrix that had 3095 rows and 4 columns, which meant we had 3095 test cases and 4 features that were structured. We used the SHAP analysis to get a big picture of how important and influential these structured features were across all the test data. At the same time, it also helped us understand what was going on at a local level, like how each of the structured features affected the price predictions for individual properties.

The way the current system works is that it can explain how it made decisions based on things like property size and location, but it can't really tell you why it liked or didn't like a particular picture of a house. This is because the system uses a special kind of math to look at pictures, called ResNet50, which breaks down images into tiny pieces that don't really mean anything on their own. Even though the system is good at looking at pictures and numbers together to

figure out how much a house is worth, it can't point to a specific part of the picture and say "oh, I liked this part" or "I didn't like that part". It can only tell you things like "I thought the house was worth more because it has a big backyard" or "I thought the house was worth less because it's close to a busy road". The system is good at explaining the things it can measure, like numbers and facts, but it's not so good at explaining the things it sees in pictures.

Future studies could take the explainability part a step further by using image-based methods like Gradient-weighted Class Activation Mapping, or Grad-CAM for short, to figure out which parts of the images are most important for predicting property values. This could give us more visual explanations and help us understand the whole valuation process better. It might also be useful to look at how these visual explanations work together with the other parts of the framework to give us a more complete picture of how property values are determined.

7.6 Similar Property Retrieval

To extend the functionality of Clarivision AI beyond residential property price prediction, a Similar Property Retrieval module was developed. The module identifies residential properties that are visually similar to a property image uploaded by the user and provides relevant comparison properties to support real-estate valuation and decision-making.

The system uses a database of property features that's stored in a file called `property_features.pkl`. This database has lots of information about each property, like the image identifier, city, street, price, number of bedrooms and bathrooms, and the size of the property. It also includes a special set of features that describe what the property looks like, which is called the deep visual feature representation. When someone uploads a new property image, the system uses the same process it used to build the database to extract the visual features from the image. It then changes the features into a simple format that can be used to compare the image to others in the database.

When you're looking for a place to live, the location is a big deal. It affects how much a property is worth, and different cities have different markets, features, and prices. So, to make the search more useful, the system only looks at properties in the city you've chosen. This way, you can compare properties that are similar and in the same local market. The system filters out properties from other cities because they might not be a good match. By doing this, you get a more accurate idea of what's available in the area you're interested in. This helps you make a better decision when you're looking for a place to call home.

The visual similarity between the uploaded property image and each candidate property image is calculated using cosine similarity. For an uploaded-property feature vector ($\mathbf{u}$) and a candidate-property feature vector ($\mathbf{v}$), cosine similarity is calculated as:

$$\text{CosineSimilarity}(\mathbf{u}, \mathbf{v}) = \frac{\mathbf{u} \cdot \mathbf{v}}{\|\mathbf{u}\| \|\mathbf{v}\|}$$

The cosine similarity score measures the similarity between the directions of the two deep visual feature vectors. A higher similarity score indicates that the uploaded property image and the candidate property image have more similar visual feature representations, whereas a lower score indicates weaker visual similarity. The calculated similarity scores are stored with the corresponding property information.

The property candidates are sorted in decreasing order based on their cosine similarity scores. The duplicate property entries are filtered out by using the criteria of city, street name, price, property area, number of bedrooms, and number of bathrooms. The top five properties (Top-K with K=5) are returned as output by the system.

The returned properties are shown using the Streamlit app interface. Each property is displayed along with its image, price, visual similarity score, city, number of bedrooms, number of bathrooms, and property area (in square feet). The similarity score is transformed into a percentage and is shown as a percentage and a progress bar that intuitively represents the degree of similarity.

In the current version of the Similar Property Retrieval module, there is an interactive qualitative assessment of similarity between property images in the form of returning the top 5 visually similar properties from the chosen city. A quantitative evaluation based on metrics like Precision@5, Recall@5, or NDCG@5 was not performed due to the absence of labeled ground-truth data indicating visually similar properties.

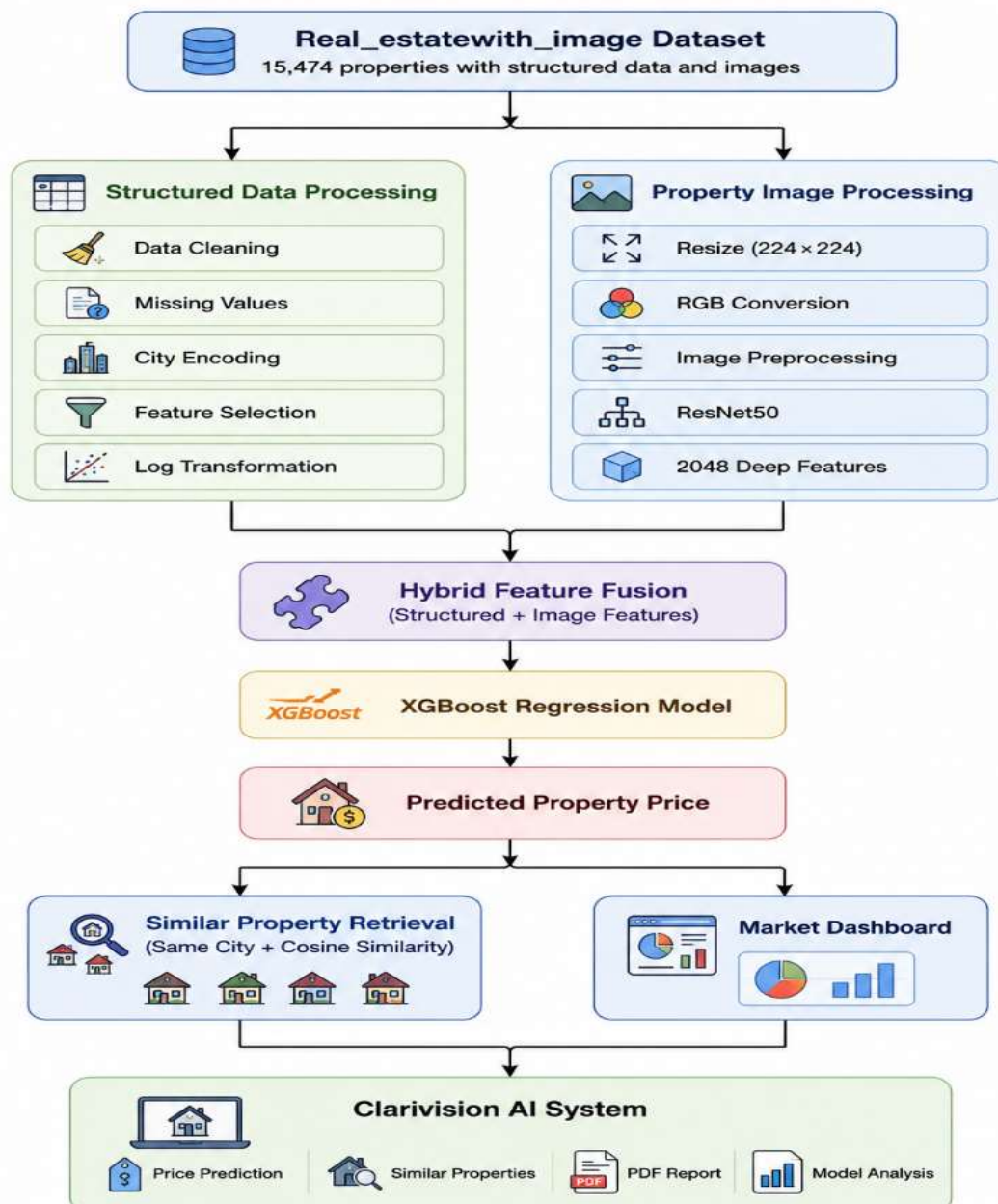


Figure 3: Overall Architecture of the Proposed Clarivision AI Framework

Figure 3 presents the architecture of the overall design of the proposed Clarivision AI framework. The framework starts from the real estate dataset in multimodal form that has structured property features and residential property images. The structured data goes through the process of data cleaning, data preprocessing, and feature engineering, and the images go through the processing by the ResNet50 convolutional neural network to extract the visual deep features. These two types of feature sets are combined in a unified hybrid form to be used as training data for the XGBoost regression model of residential property valuation. The resulting valuation is further used by the modules of Similar Property Retrieval and interactive visualization dashboard in the process of decision-making. Lastly, all the modules are combined in one Clarivision AI platform.

8. Results

The Clarivision AI framework was put to the test using a dataset that had 15,474 residential properties, each with its own set of details and pictures. This test looked at how well the new hybrid valuation model could predict things, and compared it to a different model that only used structured data. It also checked how stable the models were by trying them out multiple times, and used a technique called SHAP to see which property details were the most important. Additionally, it evaluated how well the Similar Property Retrieval module worked, which is a key part of the framework. The goal was to see if the new framework could provide more accurate valuations by combining different types of data.

We took our dataset and split it into two parts: 80% for training and 20% for testing. This was done using a random seed, which was set to 42, so we could get the same results every time. The testing data was put aside and only used at the very end to see how well our new hybrid model worked. Now, let's go through the results one by one: first, how we split the data and set up the model, then how well it predicted property prices, followed by a comparison with a baseline model that uses structured data, and so on. We also looked at how the model performed when we tested it multiple times, and used a technique called SHAP to understand what was going on inside the model. Additionally, we checked how well the model could find similar properties and how it worked with the Clarivision AI system.

8.1 Dataset Split and Model Setup

The experimental dataset consisted of 15,474 residential property records after preprocessing and data-quality checks. The dataset was divided into 80% training data and 20% testing data using a fixed random seed of 42. The property price was transformed using the natural logarithmic transformation ($\log(1 + \text{price})$) before model training to reduce the positive skewness of the target variable.

The proposed hybrid model integrated structured property attributes with deep visual features extracted from property images using ResNet50. XGBoost was used as the main regression model. The XGBoost model was configured with 300 estimators, a maximum tree depth of 6, a learning rate of 0.05, and a random seed of 42.

The structured-data-only XGBoost model was used as a baseline for evaluating the contribution of the deep visual features. The structured baseline used four property attributes: numerical city representation (`n_citi`), number of bedrooms (`bed`), number of bathrooms (`bath`), and property area in square feet (`sqft`).

8.2 Performance Evaluation of the Proposed Hybrid Model

The predictive ability of the developed Clarivision AI Hybrid model has been tested through the use of MAE, MSE, RMSE, and the coefficient of determination (R^2). This set of evaluation criteria gives complementary information about the mean prediction error, the size of prediction error, and the percentage of variance explained by the model.

Table 1. Performance Evaluation of the Proposed Clarivision AI Hybrid Model

Metric	Value
Mean Absolute Error (MAE)	0.20985
Mean Squared Error (MSE)	0.08475
Root Mean Squared Error (RMSE)	0.29112
Coefficient of Determination (R^2)	0.67375

The proposed hybrid model achieved an MAE of 0.20985, an MSE of 0.08475, an RMSE of 0.29112, and an (R^2) value of 0.67375. The MAE and RMSE values were calculated on the log-transformed property-price scale. The (R^2) value indicates that the model explained approximately 67.4% of the variance in the log-transformed property prices.

The results look good, but let's not get too excited - they're not perfect. With an R^2 value of 0.67375, we're talking about a moderate to promising performance, not an excellent one. This means that about 32.6% of the variance is still unexplained, which is a significant amount. So, while the results are promising, there's still room for improvement to get a better understanding of what's going on.

When we looked at the results in a more practical way, we changed the predictions back to the original scale of property prices. This gave us a better idea of how accurate they were. We found that the average difference between the predicted and actual prices was around 151,078.31, and the overall difference was around 238,044.55. To really understand what these numbers mean, we need to think about the currency and price units used in the original dataset.

8.3 Comparison with the Structured-Data Baseline Model

To evaluate the contribution of the deep visual features extracted using ResNet50, the proposed Clarivision AI hybrid model was compared with an XGBoost model trained using only the structured property attributes.

Table 2. Performance Comparison Between the Structured XGBoost Model and the Proposed Clarivision AI Hybrid Model

Model	MAE	RMSE	(R^2)
XGBoost (Structured Only)	0.22924	0.31369	0.62749
Clarivision AI (Hybrid)	0.20985	0.29112	0.67375

From this comparison, it is clear that the hybrid model performs better than the structured data only XGBoost model in all the metrics used. For instance, the model reduces the MAE from 0.22924 to 0.20985 and reduces the RMSE from 0.31369 to 0.29112. Moreover, the (R^2) metric rises from 0.62749 to 0.67375.

These findings suggest that the inclusion of the deep features learned by ResNet50 together with the structured property data helps in enhancing the property valuation performance by providing more information. This is evident from the fact that the (R^2) metric increases by 0.04626.

8.4 Cross-Validation and Structured-Model Comparison

To check how well the models for structured data work, a test called five-fold cross-validation was done. This was only done on the part of the data used for training, after 20% of it was set aside for testing. When doing the cross-validation, the parts of the data were mixed up and a fixed random seed, 42, was used to make sure the results were the same each time.

We compared how well two models worked: one that used structured data with XGBoost and another that used Random Forest. To do this, we looked at the average and spread of the scores from a test called R², which we ran five times with different sets of data.

Table 3. Five-Fold Cross-Validation Results for Structured-Data Models

Model	Mean (R ²)	Standard Deviation
XGBoost	0.66200	0.01175
Random Forest	0.67651	0.01011

The structured XGBoost model achieved a mean (R²) value of 0.66200 with a standard deviation of 0.01175. The Random Forest model achieved a slightly higher mean (R²) value of 0.67651 with a standard deviation of 0.01011.

The relatively low standard deviations indicate that both structured-data models demonstrated stable performance across the five validation folds. Random Forest achieved slightly higher average performance in this structured-data cross-validation experiment. However, these results should be interpreted separately from the final evaluation of the proposed hybrid model because the cross-validation experiment used only the four structured property attributes and did not include the ResNet50 visual features.

8.5 SHAP-Based Feature Importance Analysis

SHapley Additive exPlanations (SHAP) were used to improve the transparency of the structured valuation component and to examine the contributions of the structured property attributes to the model predictions. A separate XGBoost model was trained using the four structured features: numerical city representation (n_citi), number of bedrooms (bed), number of bathrooms (bath), and property area in square feet (sqft).

The resulting SHAP value matrix had a shape of ((3095, 4)), corresponding to 3,095 test observations and four structured property attributes. The SHAP analysis indicated that property area (sqft) was the most influential structured feature, followed by the numerical city representation (n_citi), number of bathrooms (bath), and number of bedrooms (bed).

The relative feature importance values were approximately 50.16% for property area, 28.56% for city, 14.28% for the number of bathrooms, and 7.00% for the number of bedrooms.

FEATURE IMPORTANCE OF STRUCTURED PROPERTY ATTRIBUTES

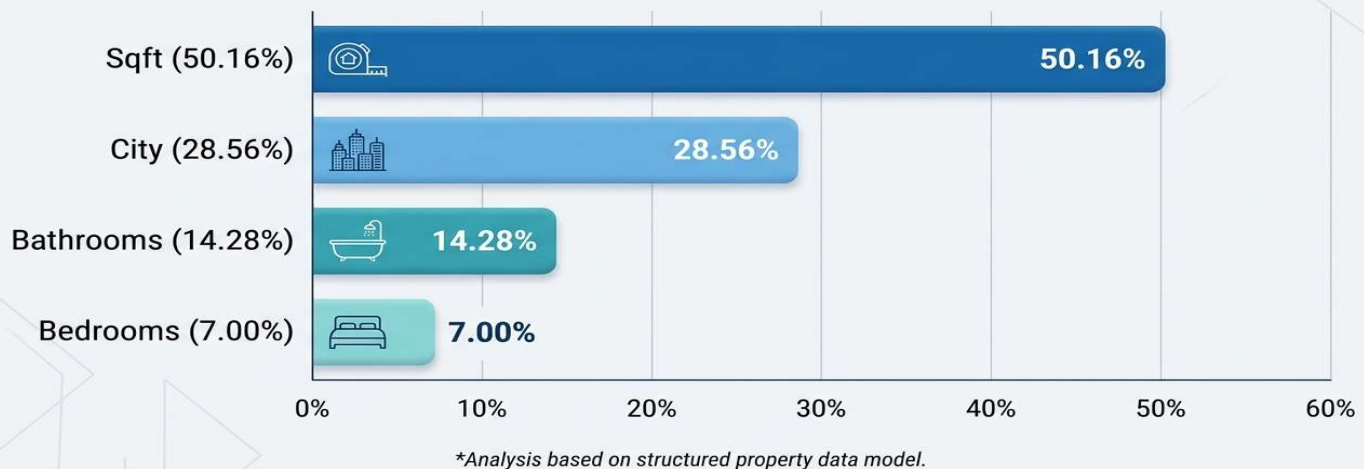


Figure 4. Relative Importance of Structured Property Features

The results demonstrate that for the structured XGBoost model, the mean score is equal to 0.66200 with the standard deviation of 0.01175. At the same time, the Random Forest model demonstrated a slightly higher result – the mean score was equal to 0.67651, while the standard deviation equaled 0.01011. Therefore, we can conclude that the Random Forest model is not only more accurate but also more consistent.

The results prove that both models, based on the structured data, performed well during all experiments. The Random Forest model performed a bit better than the other one in this particular experiment. However, it is crucial to distinguish the results of this experiment from the final results concerning the evaluation of our hybrid model since this experiment included just four attributes of the structured data but no visual attributes from ResNet50 model.

8.6 Similar Property Retrieval

The Similar Property Retrieval tool was created to help with making decisions about residential property prices. It works by finding properties that look similar to each other. To do this, it uses a special kind of math called cosine similarity to compare the visual features of different property images. These features are extracted using a pre-trained computer program called ResNet50, which is really good at looking at pictures and figuring out what's important about them. Each property image is represented by a long list of numbers - 2,048 of them, to be exact - that describe what the image looks like. The tool then uses these lists of numbers to find properties that are similar to each other, which can be really helpful when trying to figure out how much a property is worth.

To make the search results more relevant to the user's location, we limited the search to properties in the same city that the user chose. This was necessary because properties that look similar but are in different cities can have very different values due to being in different real estate markets. By focusing on properties in the same city, we can provide more accurate and helpful information for people trying to figure out the value of a residential property and make informed decisions. This way, the results are more tailored to the user's specific needs and location, making it easier for them to find what they're looking for.

The system ranked the properties that were most similar to what we were looking for, and it showed us the top 5 results. For each of these properties, we could see a picture of it, how similar it was to what we were looking for, the price, the city it was in, the number of bedrooms and bathrooms, and how big it was. We tested the system and it worked really

well - it was able to find and rank properties that looked similar to the one we were searching for. If the property we were searching for was already in the system, it would show up with a perfect match score. Other properties would have lower scores, depending on how similar they looked. The system used something called "deep visual representations" to figure out how similar the properties were, and it did a pretty good job.

The similarity between two property pictures can be measured using a score. This score shows how alike the features of the two pictures are. A higher score means the features are more similar. But this score doesn't show the probability of the properties being the same or having the same market value. It just compares the features that a computer program called ResNet50 can see.

Clarivision AI is different from traditional Automated Valuation Models, which just give you a number for how much a property is worth. Instead, Clarivision AI combines the valuation of a property with pictures of similar properties. This lets users compare the property they're looking at with other properties that look similar, and get more information to help them make decisions about the property's value.

The current study was mainly about how things work and the quality of the results, because we didn't have the actual answers to know which properties looked similar. So, we couldn't use the usual ways to measure how good the results were, like Precision, Recall, and NDCG. In the future, we should get experts or users to say what they think is similar and use that to test how well our system works. We need to get some labeled data that shows what's similar and what's not, so we can really see how good our system is at finding similar properties. This will help us make it better and more accurate.

Here's how the process works for finding similar properties using Clarivision AI. First, an image of a property is uploaded and then processed using a special tool called ResNet50. This tool helps create a unique set of features for the image, kind of like a fingerprint, but for pictures. This fingerprint is made up of 2,048 different points that describe what the property looks like. Next, the system compares this fingerprint to a big database of other property images to find the ones that look the most similar. It does this by calculating something called cosine similarity, which is a way of measuring how alike two things are. To make sure the results are relevant, the system only looks at properties in the same city as the one in the uploaded image. After comparing all the images, the system ranks them based on how similar they are to the original property. It then shows the top 5 properties that look the most similar. This helps users find properties that are similar in style or design to the one they're interested in, which can be really useful for things like real estate or design inspiration.

The Similar Property Retrieval module was created in order to aid in the visual comparison of properties whose visual attributes were similar to that of the query property. Cosine similarity was used in calculating the similarity using the feature vector extracted from the pre-trained ResNet50 convolutional neural network. In order to enhance the context of the retrieved data, the search was narrowed down to only properties that are in the same city. Up to five visually similar properties were supposed to be retrieved by the module. However, in some instances, fewer than five similar properties could be retrieved due to the small number of properties that qualified for search within the selected city after filtering out duplicate properties.

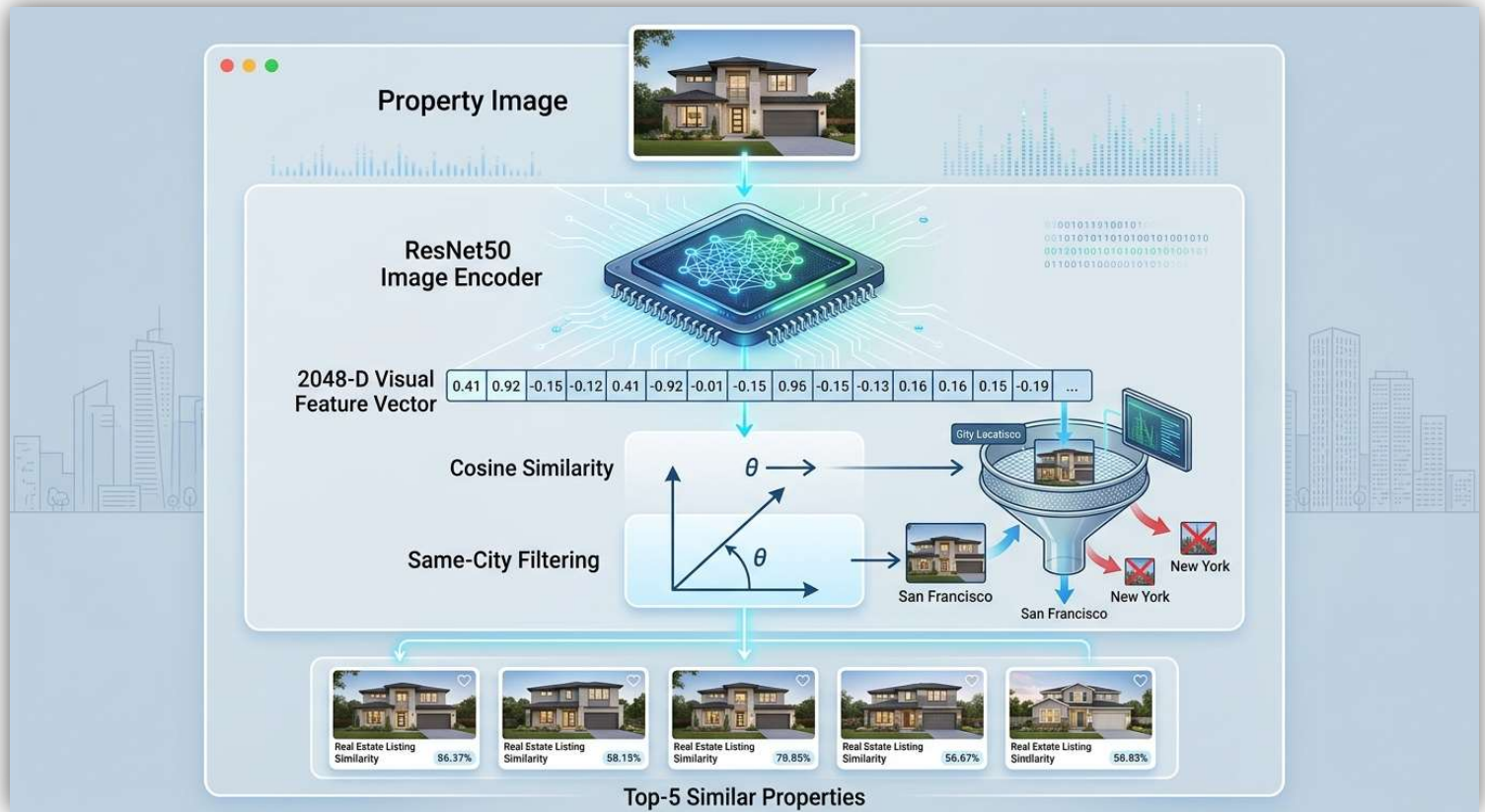


Figure 5. Workflow of the Similar Property Retrieval Module in Clarivision AI. Source: Developed by the author.

Figure 6 presents an example of the Similar Property Retrieval module. The retrieved properties are ranked according to their cosine similarity scores, with the highest-ranked properties displayed first. For readability, the retrieval results are presented in two parts: Figure 5(a) shows the first three retrieved properties, while Figure 5(b) shows the fourth and fifth retrieved properties.

Property Image



■ Top Similar Properties

Similar Property #1



Price: \$555,900

Similarity: 81.37%

City: Anaheim, CA

Bedrooms: 3

Bathrooms: 2.0

Square Feet: 1428

Similar Property #2



Price: \$739,000

Similarity: 80.39%

City: Anaheim, CA

Bedrooms: 5

Bathrooms: 3.0

Square Feet: 2406

Figure 6 (A).

Similar Property #3

	Price: \$669,000 Similarity: 78.98% City: Anaheim, CA Bedrooms: 3 Bathrooms: 2.0 Square Feet: 1603
---	---

Similar Property #4

	Price: \$567,000 Similarity: 78.50% City: Anaheim, CA Bedrooms: 3 Bathrooms: 2.0 Square Feet: 1358
---	---

Similar Property #5

	Price: \$939,000 Similarity: 78.00% City: Anaheim, CA Bedrooms: 5 Bathrooms: 3.0 Square Feet: 3001
---	---

Figure 6 (B).

Figure 6: Top five visually similar properties retrieved by Clarivision AI using cosine similarity: (a) the first three retrieved properties and (b) the fourth and fifth retrieved properties.

For testing the Similar Property Retrieval algorithm, a property query image taken from Anaheim, CA was used. It is observed that the top five visually similar residential properties were retrieved by filtering out the search based on the location of properties in the same city. It is observed that cosine similarity measures for these properties range between 78.00% to 81.37%.

The top ranking property had a similarity score of 81.37%, followed by 80.39%, 78.98%, 78.50%, and 78.00% for the remaining retrieved properties. Even though the retrieved properties varied in terms of their prices, number of bedrooms, bathrooms, and sizes, they have visual features in common, captured by the ResNet50 feature representation. This illustrates that the retrieval process is mainly driven by visual similarity and not price similarity or similarity in any other structured feature.

Filtering the results to a single city ensures that the geographic relevance is preserved because residential property values and characteristics may differ substantially depending on the cities. Hence, the city filter limits the chance of

retrieving visually similar properties but from a geographically different location and which cannot be used as local reference property.

Cosine similarity scores should be treated as measures of similarity between the ResNet50 visual feature representations and not percentage similarity in physical/economic features. Higher scores mean that the visual representations of two property images are closer to each other in the learned feature space. The retrieved properties serve as additional visual references which could complement the numerical valuation performed by Clarivision AI.

8.7 Property Information Interface

The Property Information Interface is the place where users can enter the required structured data related to the property, including city, number of bedrooms, number of bathrooms, size of the property (square feet), and uploading the image of the residential property. After entering the required data, users can click on the “Predict Property Price” button to find the predicted price of the property using the proposed Clarivision AI framework.

The proposed interface serves as the primary interface to access the system and is convenient for performing the process of residential property valuations.

The Clarivision AI framework design contains all the elements of the proposed framework in a decision support environment that allows users to perform the following actions: perform residential property valuations, view the SHAP explanations, retrieve visually similar properties, access the visualization dashboards, segment the market of residential property, and generate PDF reports.

In addition to that, the combination of all those abilities into one system improves the ease of use by delivering an effective, clear, and informative way of making decisions regarding real estate. As opposed to the common practice when AVMs only deliver numeric values, Clarivision AI delivers an interactive system combining valuation prediction, visual analysis, and intelligent decision-making service.

Hybrid Real Estate Valuation | Clarivision System

** Clarivision System **

** XGBoost + ResNet50 = Hybrid Model **

City: Salton City, CA

Bathrooms: 1

Bedrooms: 1

Square Feet: 100

Upload Property Image

Upload (200MB per file • JPG, PNG)

Predict Property Price

Figure 7. Main interface of the proposed Clarivision AI decision-support platform.

The user interface used to enter structured property attributes and upload a residential property image before performing property valuation using the proposed Clarivision AI framework.

8.8 Property Information and Price Prediction Interface

The first step in the proposed Clarivision AI-based method is to gather both the structured attributes of the property as well as the image of the residential property using a user interface. The users provide all the necessary details regarding the city, number of bedrooms, number of bathrooms, size of the property in square feet, and the property image. After gathering the necessary information, the system then evaluates the input data and estimates the value of the residential property based on the proposed hybrid model.

The screenshot shows a web interface for property valuation. At the top, there are dropdown menus for 'City' (set to Bakersfield, CA) and 'Bathrooms' (set to 10). Below these are input fields for 'Bedrooms' (16) and 'Square Feet' (1250). An 'Upload Property Image' section features a file upload button and a preview of a house image. A 'Predict Property Price' button is located below the image. A green banner at the bottom of the form area states 'Prediction Completed Successfully'. The final result, 'Estimated Property Price', is shown as \$428,391.

Figure 8: Property Information and Price Prediction Interface

Figure 8: Property Information and Property Value Prediction Interface. Users can use the interface to enter information about properties in the form of city name, number of bedrooms, bathrooms, total square feet of the property, and a picture of the residential property. Clicking on Predict Property Value will result in prediction of the residential property value using Clarivision AI model.

8.9 Clarivision AI Prediction Dashboard

The Clarivision AI Dashboard provides an interactive visualization of the valuation results generated by the proposed framework. It presents the estimated market category, prediction confidence, AI model information, and a comparison between the evaluated property and the average market characteristics. This visualization enhances users' understanding of the valuation outcome and supports transparent decision-making through an intuitive graphical interface.

Prediction Confidence

94%

Market Segment

Premium

AI Model

Hybrid AI



Property Characteristics Compared with Market Average

	Feature	Your Property	Market Average
0	Bedrooms	6	3.5064
1	Bathrooms	9	2.4533
2	Square Feet	1250	2173.9132

Figure 9: Clarivision AI Prediction Dashboard.

As shown in Figure 9, the Clarivision AI prediction dashboard displays the predicted value of the residential property, along with the prediction confidence score, the market classification, AI model details, and an interactive comparison of the property to be evaluated with the average market attributes. The dashboard enables users to have a clear visualization of residential property valuation.

8.10 Clarivision AI Report

The proposed **Clarivision AI** framework automatically generates a comprehensive PDF valuation report after completing the prediction process. The report summarizes the valuation results in a structured format, enabling users to save, print, or share the valuation outcome. It serves as a practical decision-support document for real estate agencies, financial institutions, investors, and property owners.

CLARIVISION SYSTEM

Artificial Intelligence Property Valuation Report

Report ID: CV-20260723105111

Generated: 23/07/2026 10:51

AI Engine: Hybrid CLARIVISION SYSTEM AI (ResNet50 + XGBoost)

Property Information

City	Alpine, CA
Bedrooms	5
Bathrooms	3
Square Feet	980 Sqft

Artificial Intelligence Prediction

Estimated Property Value	\$459,788
Prediction Confidence	94%
Market Category	Mid Range

Artificial Intelligence Decision

The Hybrid Artificial Intelligence model predicts that this property belongs to the "Mid-Range Market Segment". The prediction is supported by the combination of image analysis and structured property data. Main Factors: ✓ Property Area ✓ City Location ✓ Bedrooms and Bathrooms ✓ Image Features extracted using ResNet50 Overal Confidence: Very Good.

Figure 10: Clarivision AI Report.

9. Discussion

From the experiment findings, it is evident that incorporating structured property attributes and deep visual representations can enhance residential property price prediction as compared to structured data only, as used in this study. The proposed Clarivision AI hybrid model gave better prediction errors and higher (R^2) as compared to the structured data XGBoost model. This implies that the visual features captured from images through the ResNet50 were more complementary since they added value to structured numeric data.

In this case, the proposed Clarivision AI hybrid model generated an (R^2) of 0.67375, meaning that about 67.4% of the variation of the transformed property prices was explained by the model. In this case, the generated (R^2) is promising and moderate rather than complete or high. About 32.6% of the variation was unexplained, which could be due to various reasons that are not captured in the current dataset like age, quality of construction, location, latitude and longitude, market conditions, and other property-related attributes.

The superior performance of the hybrid model over the structured-data-only XGBoost model lends weight to the possibility that multimodal learning can be valuable for residential property valuation. The R^2 score for the structured data only model was 0.62749, while for the proposed hybrid model, it was 0.67375. The improved score by approximately 0.04626, in addition to the decreases in MAE and RMSE scores, implies that the inclusion of both structured data and visual data improved the performance of the valuation process. This finding aligns with those of previous studies which have shown that the inclusion of structured property data and image data improves the accuracy of residential property valuations. These findings must nonetheless be considered within the context of the data set and experiments carried out in this research study.

The data-quality analysis showed that the dataset contained no missing values and no duplicate records. The property-price variable exhibited positive skewness, with a skewness value of approximately 1.35. Therefore, the target price was transformed using $(\log(1+\text{price}))$ before model training. This transformation reduced the influence of the right-skewed price distribution and provided a more suitable target representation for the regression models.

The SHAP results yielded further explanations regarding the influence of structured property attributes. The property area (sqft) was found to be the most important structured attribute, while other important attributes include numerical city representation (n_citi), number of bathrooms (bath), and number of bedrooms (bed). The influence of property area and location variables is expected given the overall findings in the literature on residential property evaluation which considers property size and location as crucial factors influencing the value of the property. However, the SHAP explanation currently developed can explain only the structured attributes and not the individual visual attributes generated from ResNet50.

The five-fold cross-validation experiment revealed relatively little variability of the (R^2) metric scores in case of structured-data models. In particular, the mean (R^2) score of the structured XGBoost was found to be 0.66200 with the standard deviation of 0.01175, whereas that of the Random Forest model was equal to 0.67651 with the standard deviation of 0.01011. This suggests relatively high stability of the model's performance across the validation folds. It is worth mentioning that the cross-validation experiment has been carried out based on the structured property data only, and thus cannot be considered as the evidence of robustness of the hybrid model under consideration. Furthermore, the obtained results do not provide sufficient grounds to argue that the proposed hybrid model does not suffer from overfitting problem.

Apart from property price prediction, Clarivision AI also offers SHAP explanations, Similar Property Retrieval, visualization capabilities, market-oriented decision support outputs, and reporting in one platform. The Similar Property Retrieval system finds the Top-5 visually similar properties in the city chosen by the user based on cosine similarity of the deep visual features. This capability makes the system go beyond mere price prediction since it also gives the user some properties to compare. Nonetheless, retrieval was only assessed in terms of function and quality as there were no similarity ground truths for the data set used. Thus, the quantifiable measures of Precision@5, Recall@5, and NDCG@5 were not computed.

Combining the prediction of the valuation, explainability of structured features, similarity search via image retrieval, visualization, and reporting illustrates the real-world value of Clarivision AI as an Explainable Decision-Support Platform. However, the suggested framework should not be viewed as a substitute for professional property appraisal. The results are based on the properties of the dataset used, model setting and attributes of the property.

In general, the results presented above prove that multimodal learning and Explainable Artificial Intelligence represent a promising approach for enhancing property valuation systems. In particular, the suggested framework performed better than the structured-data-only approach in terms of predictive accuracy and provided additional functionality for decision-making. However, further studies with a larger variety of datasets, additional attributes of property, image-level explanation techniques, and quantitative retrieval assessment should be conducted before further generalizations can be made.

10. Conclusion

This paper presented the development of Clarivision AI, an interpretable hybrid model for residential property valuation that combined structured property data along with visual features obtained through ResNet50. The hybrid features have been used with the XGBoost regression algorithm while SHAP is used for the explanation of the contributions of the structured property features.

The proposed framework resulted in an MAE of 0.20985, RMSE of 0.29112 and (R^2) of 0.67375 on the transformed property-price scale. This indicates decent and promising predictive performance and shows that adding visual features increases the performance compared to the structured-data-only approach.

Apart from this, the Clarivision AI framework also includes Similar Property Retrieval, interactive visualization and automated reporting modules that can be helpful in real estate decision making. However, the framework performance depends on the quality of the dataset and its geographical diversity, availability of property features and the application of log-transformed prices. Hence, the framework must be seen as the decision support tool and not a replacement for professional property appraising.

Future work will involve evaluation of the framework based on larger and more geographically diverse datasets, additional property features, external validation, visual explanations and retrieval metrics.

11. Recommendations

Based on the findings and limitations of this study, the following recommendations are proposed:

1. For future research, structured attributes like property age, property type, environmental characteristics, socio-economic information, accessibility, etc., need to be added to enhance the valuation performance of residential property valuation.
2. In future versions of the framework, multiple images of the property, which can be indoor, outdoor, aerial, and street view images, will be used to offer a better visual representation of the properties.
3. The proposed framework needs to be tested on residential property datasets from different geographic regions and market situations to evaluate its robustness and regional validity.
4. To get explanations of visual and structured features of the property, other approaches to explainability like LIME, Grad-CAM, and Integrated Gradients need to be used.
5. Similar Property Retrieval module needs to be tested using quantitative retrieval metrics like Precision@K, Recall@K, and NDCG@K, supported by the validation of experts or users. In addition, the process of retrieval can be improved using visual similarity along with other factors like location, property type, construction year, market trends, and property condition.
6. Upcoming developments can involve live data from the market and online real estate listings to provide timely property valuations and make the framework more practical.
7. Potential future updates of Clarivision AI could involve categorization of properties according to their market segmentation, including low-value, medium-value, and high-value property categories.
8. The proposed framework can be implemented as a cloud-based decision support system using web and mobile applications for real-estate experts, financial institutions, investors, and other legitimate users.
9. Upon future research, potential biases and fairness problems in property valuations done by an automated system need to be explored. The performance of the model should be analyzed in different cities, neighborhoods, property classes, and pricing levels to see whether prediction mistakes target specific areas or market segments.
10. The ethical use of automated valuation systems should be considered during future development. The framework should support responsible data use, privacy protection, secure handling of property information, and appropriate governance when automated predictions are used in financial or high-impact decisions.

11. Clear transparency measures should be provided for users who rely on automated property-price predictions. The system should communicate the model's limitations, the uncertainty associated with predictions, the main factors influencing the estimated value, and the appropriate role of human professional review.

References

- Balila, A. E., & Shabri, A. B. (2024). Comparative analysis of machine learning algorithms for predicting Dubai property prices. *Frontiers in Applied Mathematics and Statistics*, 10, Article 1327376. <https://doi.org/10.3389/fams.2024.1327376>
- Deng, L. (2025). Real estate valuation with multi-source image fusion and enhanced machine learning pipeline. *PLOS ONE*, 20(5), Article e0323351. <https://doi.org/10.1371/journal.pone.0323351>
- Despotovic, M., Koch, D., Stumpe, E., Brunauer, W. A., & Zeppelzauer, M. (2023). Leveraging supplementary modalities in automated real estate valuation using comparative judgments and deep learning. *Journal of European Real Estate Research*, 16(2), 200–219. <https://doi.org/10.1108/JERER-11-2022-0036>
- Du, W.-W., Wang, W.-Y., & Peng, W.-C. (2023). DoRA: Domain-based self-supervised learning framework for low-resource real estate appraisal. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management (CIKM '23)* (pp. 1–7). Association for Computing Machinery. <https://doi.org/10.1145/3583780.3615470>
- Geerts, M., Reusens, M., Baesens, B., vanden Broucke, S., & De Weerd, J. (2025). *On the performance of LLMs for real estate appraisal* (arXiv:2506.11812). arXiv. <https://doi.org/10.48550/arXiv.2506.11812>
- Hasan, M. H., Jahan, M. A., Ali, M. E., Li, Y.-F., & Sellis, T. (2024). *A multi-modal deep learning-based approach for house price prediction* (arXiv:2409.05335). arXiv. <https://doi.org/10.48550/arXiv.2409.05335>
- Huang, C., Liang, B., Li, Z., & Chen, F. (2025). Multimodal machine learning for real estate appraisal: A comprehensive survey. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 345–361). Springer. https://doi.org/10.1007/978-981-96-8183-9_26
- Kee, T., & Ho, W. K. O. (2025). eXplainable Machine Learning for Real Estate: XGBoost and Shapley Values in Price Prediction. *Civil Engineering Journal*, 11(5), 2116–2133. <http://dx.doi.org/10.28991/CEJ-2025-011-05-022>
- Maselli, G., & Nesticò, A. (2025). Machine learning algorithms and explainable artificial intelligence for property valuation. *Land*, 14(1), Article 132. <https://doi.org/10.3390/land14010132>
- Mora-Garcia, R.-T., Cespedes-Lopez, M.-F., & Perez-Sanchez, V. R. (2022). Housing price prediction using machine learning algorithms in COVID-19 times. *Land*, 11(11), Article 2100. <https://doi.org/10.3390/land11112100>
- Rodríguez-Serrano, J. A. (2024). Prototype-based learning for real estate valuation: A machine learning model that explains prices. *Annals of Operations Research*, 344(1), 287–311. <https://doi.org/10.1007/s10479-024-06273-1>
- ted8080. (n.d.). *House prices and images – SoCal* [Data set]. Kaggle. <https://www.kaggle.com/datasets/ted8080/house-prices-and-images-socal>
- Wang, Z., Wang, Y., Wu, S., & Du, Z. (2022). House price valuation model based on geographically neural network weighted regression: The case study of Shenzhen, China. *ISPRS International Journal of Geo-Information*, 11(8), Article 450. <https://doi.org/10.3390/ijgi11080450>
- Yazdani, M., & Raissi, M. (2023). *Real estate property valuation using self-supervised vision transformers* (arXiv:2302.00117). arXiv. <https://doi.org/10.48550/arXiv.2302.00117>