



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

Prepared By

Hassan Mohammed Ahmed Alamer Khalifa

ID No: 253002941

**Submitted in fulfillment of the requirements for the degree
Master of Science in Artificial Intelligence**

At

**Collage of Informatics
Midocean University
UAE**

Name Of Supervisor: Dr / Amjad Qtaish

Collage of Informatics, Midocean University, Comoros

Chapter 1 Introduction

1.1 Introduction

Real estate valuation is the process of assessing the value of real property, and it is playing a crucial role in international finance through determination of investment options, mortgage approvals, and proper risk assessments. Recently, the valuation approach had altered due to emergence of digital real estate platforms and the availability of vast property datasets. Automated Valuation Models (AVMs) currently are spread; and providing rapid and scalable valuation estimations rather than traditional valuation methods (Balila & Shabri, 2024) [1].

While there are many benefits to AVMs, there remain a number of important limitations associated with them. In most cases, current methods depend mainly on structured numerical attributes like location, area, age, and number of rooms of the property. While all of these elements are important, none of them are sufficient to characterize features affecting the value of residential properties. There are both numerical and qualitative factors that affect real estate markets. Qualitative visual attributes include the condition of the house, quality of the finishing inside the house, aesthetic appeal of the building, etc. It is quite challenging to quantitatively analyze all of these features using traditional structured data. Modern research has shown that one should use visual information in conjunction with structured attributes of the property (Maselli & Nesticò, 2025; Deng et al., 2025) [2][3].

Recent developments in deep learning and computer vision have greatly enhanced the capabilities of extracting relevant information from the pictures of a property, such as its architectural style, state of maintenance, interior finish, and other visual properties that are not easily represented by structured numerical information alone. Therefore, the analysis of images has become an increasingly critical part of Automated Valuation Models (AVMs). Yet, in many current valuation systems, structured numerical information and visual information are processed independently from one another, leading to disconnected valuation pipelines that do not fully take advantage of

the synergy between these two forms of information. Multimodal learning, on the other hand, has been stressed as a more efficient approach to valuation, whereby structured numerical information about properties and deep visual features are combined into a single predictive pipeline (Despotovic et al., 2023; Deng et al., 2025) [6][3].

While deep learning models have shown to be excellent at predictions, they tend to be considered black box models because it is not easy for people to comprehend the manner in which valuation is done. The inability to get an understanding about the process lowers the trust people would have in AI-assisted valuation systems, especially when dealing with expensive financial and investments decisions that require explainability. As such, much of the research focus has been on introducing XAI into automated valuation systems (Maselli & Nesticò, 2025; Kee & Ho, 2025) [2][4].

This research work aims at solving some problems existing in the domain of machine learning methods for residential real estate valuation. Such problems are a poor integration of numerical characteristics of the property with its image data, a low degree of model interpretability, and a lack of functionality in terms of decision support. The following solution to these problems is offered in this work – Clarivision AI, an interpretable framework which uses structured data as well as ResNet50 neural networks with image data to enhance model performance and interpretability. Moreover, the suggested framework provides similar property search, visualization, and report generation, helping real estate agents, banks, and investors in their decision-making process (Maselli & Nesticò, 2025; Huang et al., 2025) [2][5].

1.2 Problem Definition

Property valuation plays a fundamental role in financial markets, mortgage lending, investment analysis, and risk management. and plays a key role in directing investments and defining risks. In other words, having the right valuation for any type of asset-whether commercial or residential-is extremely important when dealing with large figures, either making money or losing it all. In essence, proper valuation helps to avoid investment risks in the system.

Despite technological advancements and innovations that facilitate real estate valuation and enable more advanced automatic valuation models (AVMs), which incorporate artificial intelligence and machine learning algorithms, there remain three main challenges standing in the way of improving the accuracy of the process and gaining professional confidence.

First, it is important to consider the lack of qualitative data that could significantly improve the results. Most AVMs utilize quantitative variables to assess the property, including its area, location, construction age, and number of rooms. Such metrics are essential, but they do not account for all factors considered by potential investors and creditors. **(Balila & Shebri, 2024) [1].**

Secondly, current valuation methods are not successful in integrating heterogeneous datasets, especially numerical and visual property attributes. Thanks to recent developments in machine learning and deep learning, it is possible to extract valuable visual features from the pictures of buildings, such as the state of a building, its style, and the quality of interior finishing. Yet, most current Automated Valuation Models (AVMs) operate with structured numerical and visual datasets separately, thus generating disconnected valuation processes, which do not take advantage of the complementarity of the two datasets. As a result, numerical indicators like the area and the number of rooms are used to estimate the market price of the building whereas important visual features that greatly affect the price are overlooked **(Despotovic et al., 2023; Deng et al., 2025) [6][3].**

Third, the inability of many state-of-the-art models to explain how predictions are made and what factors drive them constitutes a significant hurdle in the implementation of artificial intelligence for the purpose of estimating the value of residential property. While it is true that deep learning

models can produce remarkable predictions in terms of accuracy and effectiveness, the output of such models tends not to include an explanation as to how these predictions were derived, as well as what structural and graphical factors drive the predictions of the estimated value of the property. In other words, the black-box nature of the aforementioned models leads to decreased confidence among users of AI-assisted valuation tools, especially when the decision-making process requires justification of valuation decisions in financial and investment settings (Maselli & Nesticò, 2025; Kee & Ho, 2025) [2][4].

1-2-1 Summary of the Research Problem

Valuation models tend to lack an encompassing structure in linking property's quantitative aspects with its qualitative features. The absence of integration between these two aspects makes it difficult to interpret valuation figures as well as increases their inaccuracies. The present study seeks to bridge this gap by introducing an explainable hybrid valuation framework that incorporates both types of variables into the forecasting process.

To mitigate these limitations, this research paper suggests the use of the Clarivision AI system that incorporates structured data about property, deep visual features, and explainable artificial intelligence into one valuation model.

1.3 Thesis Objectives

The objective of this study is to develop a machine learning model that will predict the prices of real estate by integrating information about the structure and information inferred from visual data. The successful achievement of this objective will be attained through prioritizing certain goals:-

1. Improving a structured valuation model; by constructing a predictive model using the structured numerical property details.
2. Designing a model that uses deep learning to look at pictures and find out what the property's condition looks like.

3. Develop a hybrid predictive framework integrating structured numerical data with deep visual features extracted from property images.
4. Explain the prediction process using Explainable Artificial Intelligence (XAI) through which the model derives predictions, incorporating both numerical data as well as visual data that influence the value of properties in the real estate market.
5. Assess the prediction ability of the hybrid model using standard metrics for evaluating regression models such as MAE, RMSE, and R^2 .
6. Developing an AI-based decision support system for investors, banks, and real estate experts.
7. Test the robustness of the hybrid model that was presented under realistic situations such as missing structured information, noisy inputs, and incomplete property images.

1.4 Thesis Contribution.

With the above in mind, the current research attempts to tackle some of the pertinent challenges that exist within the present-day real estate valuation and appraisal exercise by incorporating artificial intelligence (AI) and concentrating on AI specifically in the exercise of the two aspects i.e., theoretical and practical to ensure the successful incorporation of AI within the realm of real estate valuation, appraisal and. The objectives of our research within the current study may be summarized as the following: -

1. A Unified Hybrid Valuation Approach.

This research proposes a new and more accurate method for prediction using both digital and visual data. Instead of focusing on each separately, this research considers that digital and visual data together form a cohesive framework for achieving and understanding a better valuation.

2. Emphasis on Explainability.

The new models differ from other systems that are difficult to interpret and understand, as they explain things in two ways: with digital data and visual data (images) This makes it easier for

people to see what is happening and helps interested companies and real estate sectors to feel comfortable using it.

3. Empirical Evidence on Multimodal Integration.

This research and study examine the use of different methods for pricing property value. This research investigates and test three methods: the first relies solely on numerical data and all its details; the second relies solely on visual data (images); and the third, the most important method in this research, combines both numerical and visual data. Through this combination, the study will demonstrate how using both numerical and visual data together can contribute to understanding property value and pricing

4. Robustness Analysis under Real-World Scenarios.

The model is priced and evaluated in the presence of many conditions that reflect real-world market environments, including the availability of incomplete and heterogeneous.

The suggested model was tested under tough conditions such as missing values, noisy values, and incomplete images to test its robustness data.

5. Practical Orientation.

The study's contribution to theoretical development. The framework was developed and designed taking into account the extent of the implementation process and contributes to and aims to achieve a real balance between prediction accuracy and interpretability, making it suitable for real-world financial, real estate and investment applications.

6. Creation of Explainable Decision Support System.

A Streamlit application was developed that enables the valuation of properties on the fly, provides information about similar properties, displays feature importance, and generates a report.

The system also provides a reliability score to improve users' confidence in valuation outcomes.

1.4 Thesis Organization

Chapter 1: Introduction

This chapter defines the research problem, objectives, and contribution of the research.

Chapter 2: Theoretical Background and Related Studies

This chapter presents traditional property valuation and appraisal methods, as well as previous research and studies related to machine learning-based and image-based valuation models.

Chapter 3: Proposed Methodology

This chapter presents the proposed hybrid model and explains the data, methodology, and model structure.

Chapter 4: Experimental Results and Discussion

This chapter presents the experimental results, compares the different models, and discusses the expected outcomes.

Chapter 5: Conclusion and Future Work

This chapter presents the main findings of the study and suggests directions for future research.

Chapter 2: Background and Related Works

2.1 Background

For a long time, traditional and outdated methods of property valuation have relied on the opinions and judgments of specialists and comparisons with recent sales in the same area or location. These procedures outline the fundamental concepts in real estate evaluation, which are always significantly reliant on the skills of the real estate evaluator and the presence of regional data. This normally results in disparities in valuation, making it challenging to achieve economies of scale in the ever-changing world of real estate. The advent of online real estate portals, big data, and computerized databases has made significant strides toward addressing these issues since they have made available all the relevant data regarding the property's characteristics and history being able to tell the type and in what kind of location it is. This led to the growth of classifications and collections of large real estate datasets, which in turn led to the development of the first Automated Valuation Models (AVMs) (Balila and Shebri, 2024) [1].

Unlike the initial ones, the new model utilizes procedures that were developed in data science in the 1970s. But these approaches seem to be less adept in dealing with the intrinsic nonlinear, multidimensional nature of spread real estate markets. As machine learning technologies started developing fast. Meanwhile, these methods allow working with big, highly complex datasets and identifying complicated relations of high-degree, as well as relatively simple relationships, among the variables involved. Thus, this approach provides a significantly more detailed estimate of the fair market value of a property than any random traditional statistics.

Some relevant characteristics, which influence the property value and, consequently, its price include various other metrics besides i.e., visual aspects of the property such as (Ex: lighting, flooring etc.) might not be included in the house inventory prior to missing exposure to real estate. Among all of these characteristics, perhaps the most important are visual aspects of the property, which attract or engage the buyers. Deep learning research, especially, developments in the field of convolutional neural networks (CNNs), has dramatically changed the way visual property

information has been analyzed, with consequences. Specifically, recent developments in CNN technology have enabled analyzing the multi-scale visual information of a property such as its finishes, colors, conditions, etc., in ways that were impossible for previous automated systems. On the other hand, most of the existing systems of real estate valuation analyze the digitalized information about the property (age, location, transport accessibility) as a separate category from the visual information viz, interior and exterior design.

Issues related to explainability. The second challenge in relation to advanced machine learning approaches, especially those based on deep learning methods, concerns explainability. Despite the ability of these models to yield accurate prediction results, they are seen as black box solutions because they produce valuation results without explaining the logic of prediction and the importance of particular structured and visual features. The lack of explanation is one of the critical aspects that prevents the use of AI-based property valuation systems in important settings like mortgage financing, investments, and regulation. Thus, investors, lenders, and regulators need explanations regarding the process of value determination in order to ensure its fairness and reliability. Recent research has paid attention to the implementation of Explainable Artificial Intelligence (XAI) solutions, which allow users to learn about the impact of structured numerical features and deep visual features on valuation results (Maselli & Nesticò, 2025; Kee & Ho, 2025) [2][4].

Some current studies and research are moving towards hybrid, multimedia models that combine digital property data with visual components within a comprehensive system. These models contribute to simulating the complete appraisal and valuation process of a human appraiser or expert by integrating precise quantitative data with in-depth insights derived from images. This study will contribute to developing and proposing a system that not only aims to improve data accuracy and reliability but also promotes transparency and practicality for use across various real estate sectors.

2.2 Related Work

These recent research studies have focused extensively, continuously, and increasingly on how to use machine learning and deep learning in analyzing, valuing, and predicting real estate prices. All researchers have contributed with different types of data, such as real estate details, digital information such as location information, property descriptions, and also visual data such as pictures of the property, in an attempt to make predictions more accurate.

- **Mora-Garcia et al. (2022)** A large-scale study was conducted to compare the performance of machine learning algorithms in predicting real estate property prices. The study was equally divided between experiments using classical statistical methods and more sophisticated computing techniques such as multiple linear regression (MLR), random forests (RF), gradient enhancement machines (GBM), and artificial neural networks (ANNs). It turned out that current machine learning techniques have an increased efficiency in the detection of such complex patterns that were hardly detectable using traditional methods applied to large data sets. The use of geospatial data for research purposes in the form of a rich set of property characteristics and their exact coordinates was another strong point. The biggest benefit of the experiment is that it proved empirically the possibility of better performance of clustering and neural networks-based models in terms of predictive accuracy as well as ability to process large data sets efficiently in the real estate sector in turn, application of these models needs special computational capabilities and knowledge, which makes them less accessible than, for instance, regression. From the research perspective, "the visual omission" in machine learning was illustrated by the experiment done by **Mora-Garcia et al. (2022)** [13]. The use of clustering and neural networks to process spatial (grid, droplets) and digital (vector) data proved effective in this particular case; however, unstructured visual information from real estate photographs was missing.

- **Wang et al. (2022)** They adopted a spatial modeling technique known as Geographic Neural Network Weighted Regression, GNN-WR. Integrating Neural Networks, a technique for spatial patterns and local effects and a Geographic Weighted Regression possibility. In other words, to solve the issue of real estate market data being stable, the two techniques were combined to have it fit in with the geographical area of the property, allowing the weights to be assigned accordingly.

The urban housing database used in this research consisted of spatial data including not only the usual property markers such as beds count, neighborhood characteristics, but, most importantly, each locality's geographic coordinates and environmental characteristics.

One would argue that the main feature of the GNN model is its spatial variability and ability to detect unique characters of various neighborhoods and changes occurring in local markets while bigger models often miss those. On the other hand, it can be viewed as highly complex due to high spatial variability of the model as criticized by **Wang et al., (2022)** [7]. GNN-WR may be called the victim of what is known as a ubiquity blind spot. Even though it does excel at considering locations when evaluating prices of the property, it fails to take into account that properties cannot be seen just as a set of points on the map or numbers in the database. In other words, this model lacks computer vision to analyze and detect characteristics of the building itself and architectural features of a house with our thesis, we want to bridge that gap by combining complex spatial weighting with visual features of properties **Wang et al., (2022)** [7].

- **Rodriguez-Serrano (2024)** A prototype-based learning model in a domain-specific context was developed in order to explain the prediction of housing prices. Unlike considering all data points, this approach singled out only a few prototypes of real models among all the points in the data set as "anchors" (Amos et al., 2018). In the view of such concrete examples, the proposed model according to Concretely models represents a way of thinking close to human, involving an assumption of houses similar to the given one having the most influence on its value. The research includes the rationalization of a data set represented by the property record—the

property The main advantage of the introduced model in terms of interpretation lies in the possibility to provide explanations of complicated machine learning findings using understandable cases.

Thus, the process becomes much easier for lenders and property owners. Nevertheless, the quality and variety of prototypes considerably affect the accuracy of a model for example, if there is no good prototypical sample among the properties representing a unique/luxury house, the expected value might be not accurate enough. From a research perspective, Rodriguez-Serrano's method faces limitations as it relies on a fixed set of features **Rodriguez-Serrano (2024) [8]**.

Although prototype-based explanations are useful for structured data such as number of rooms or size, they have yet to be adapted to visual features. For example, the model can discern prototypes by size but not by modern interior or architectural style, due to the lack of a visual component for deep learning. This is indicative of a fundamental gap: that explanations based on prototypes need to be extended to multiple data types. This thesis work aims to fill this gap with prototypes that integrate digital information and visual components (**Rodriguez-Serrano, 2024) [8]**.

- **Balila and Shabri (2024)** A comprehensive comparative analysis was performed to assess the effectiveness of different ML architectures in the unique real estate market in Dubai This research paper considered several algorithms such as support vector machine (SVM), gradient boosting machine (GBM), random forest (RF), and artificial neural network (ANN). The results indicated that the combination of different algorithms within an ensemble model tends to improve the performance of the model for high-value real estate segments in the international city of Dubai. The data used for this analysis was local data on property sales and included the type of property, age, premium, and distance to landmark buildings like the Burj Khalifa tower. Some of the notable findings from this research were the application of ensemble methods to solve the issue of luxury homes' premiums and changes in the market dynamics of Middle Eastern real estate markets Yet, these models require fine-tuning of the hyperparameters, as an

inappropriate tuning might lead to overfitting that would disallow their usage for real-time price estimation. From research point of view, the **Bali and Shipper (2024) [1]** work is notably wanting, so much as to be called blind. While they did establish strong numerical indicators for the Dubai market, their analysis did not include such elements as the aesthetic quality of properties and the standards of finishing, which oft have an impact on the prices of luxury homes. In fact, in Dubai, the price of two apartments in the same building can vary greatly due to interior design or view, elements not considered by the traditional data sets. This gap indicates the significance of incorporating computer vision techniques to enhance these ensemble models. Based on their findings, this thesis seeks to integrate visual feature extraction to enhance consideration of the 'luxury factor' in real estate evaluation (**Balila and Shebri, 2024) [1]**.

- **Hasan et al. (2024)** Thus, the proposed study involves a highly advanced deep learning approach to transforming appraisal property by taking into account the following data sources: structured tabular data, images, and text description of property listing. In addition, using a multi-branch neural network, the authors proved that the combination of such different data sources led to improved accuracy and decreased errors in price prediction, compared with the models utilizing only numeric data. The database is gathered based on data extracted from various real estate platforms comprising thousands of real estates where every listing has numeric data related to the area, and photos both of interior and exterior parts of houses, as well as agents' notes about them. The benefit of such an approach lies in the usage of all possible data, similarly to the approach chosen by a professional appraiser who evaluates a house based on its looks and main selling points. The downside of this process is that it is a highly complex and resource-intensive process. To train the model properly by weighting such elements as the importance of the image and a sentence of text for the name would require considerable computing powers and large amounts of data. As for the research problems identified by **Hassan, Chauhan, Jiang, and Idrees... (2024) [14]**, they face the interpretational gap problem. However, while focusing on the idea of increasing the number of data sources for achieving increased accuracy, this model remains a black box with a high level of opacity. Therefore, one of the major problems to solve in this area is developing interpretive artificial intelligence in multimedia systems. Thus, continuing from

the ideas introduced by Hassan in his multimedia model, the current paper aims at adding transparency to the model **Hasan et al. (2024) [14]**.

- **Deng (2025)** proposed an advanced real estate valuation framework that integrates structured property attributes with multi-source property images to improve valuation accuracy. The study employed several machine learning algorithms, including XGBoost, Random Forest, LightGBM, Extra Trees, K-Nearest Neighbors (KNN), Support Vector Regression (SVR), Multi-Layer Perceptron (MLP), and Multiple Linear Regression (MLR). Furthermore, the framework utilized Bayesian Hyperparameter Optimization to improve model performance and SHAP (Shapley Additive explanations) to enhance model interpretability. The dataset consisted of structured property information together with multiple property images collected from real estate listings. One of the major strengths of the proposed framework is its ability to combine structured and visual information while providing explainable predictions through SHAP analysis. In addition, Bayesian optimization contributed to improving the predictive accuracy of the machine learning models. However, the proposed framework was primarily designed for property price prediction and did not provide a complete intelligent decision-support application for end users. Moreover, although the study incorporated multiple property images, it did not include practical functionalities such as similar property retrieval, interactive market visualization, or automated reporting. Therefore, the proposed Clarivision AI framework extends Deng's work by integrating explainable property valuation with Similar Property Retrieval, interactive dashboards, and automated PDF report generation, providing a complete intelligent decision-support system for residential property valuation (**Deng, 2025**) [3].

- **Huang et al. (2025)** Multimodal Machine Learning for Real Estate Valuation: A Survey was conducted by **Huang et al. (2025) [5]**. This study provides an exhaustive overview of the current research on multimodal machine learning approaches to the valuation of properties. Huang et al. considered several approaches to the integration of structured property features with unstructured data sources like images of a property, textual descriptions, satellite imagery, and graphical data.

Various deep learning models such as CNNs, ViTs, GNNs, and multimodal fusion models were studied and compared with respect to performance and use cases.

According to the study results, multimodal learning can substantially increase the accuracy of property valuations due to the use of complementary information within various modalities. However, the paper pointed out a number of remaining problems, such as the lack of application of Explainable Artificial Intelligence (XAI) techniques, the absence of evaluation of the contribution of individual modalities, poor multimodal interaction methods, and the lack of decision support systems.

Considering the context of the present study, the limitations described above directly lead to the motivation for the development of the Clarivision AI framework. In contrast to other multimodal models previously used, the Clarivision AI system incorporates structured housing features together with deep visual features generated using ResNet50 architecture while simultaneously employing the approaches of Explainable AI, smart similar property search, and an interactive decision-making dashboard. Hence, the proposed framework resolves multiple limitations discussed by **Huang et al. (2025) [5]**.

- **Yazdani and Raissi (2023)** This paper has introduced a novel real estate appraisal method with self-supervised vision transformer (ViT) models. Rather than relying on traditional convolutional neural networks, this approach applies a design based on transformer models to extract detailed visual features from real estate images and combines them with structured meta-data. The researchers demonstrated that, by enabling it to learn visual patterns without human labels, the model outperforms traditional automated valuation model (AVM) that relies solely on digitized or tabular-based data. The study employed a massive-scale multi-modal dataset consisting of thousands of real estate listings in which high-resolution images of the properties are paired with structured and location-relevant information about the property. The key strength of this approach is its computational complexity, which allows for the capture of fine global patterns in images - such as how lighting, material quality, and architectural design correlate - that finer models tend to ignore. Like the disadvantages, the training and inference processes of such state-of-the-art

transformer models demand considerable computation power and specialized hardware like GPUs. As far as the literature is concerned, research by Yazdani and Raissi is confronted with what is called the interpretability gap problem of transformers. While vision transformers offer high accuracy levels, they tend to be harder to interpret than traditional deep learning techniques; hence it becomes hard for human raters to understand why certain visual attributes affect prices. This leads to a critical research requirement; that is, the establishment of an adequate level of visual fidelity for explainable AI (XAI). And through their multimodal framework, this study aims to add an element of transparency where the visual insights obtained from the transformer become comprehensible to end-users Yazdani and Raissi (2023) [9].

- **Maselli and Nesticò (2025)** introduced a framework for explainable machine learning in order to create residential property valuations by considering two critical issues with regard to automated valuation models – the issues of predictive performance and interpretability of the model. Three most popular machine learning methods such as ANN, k-NN and RF were analyzed using the dataset on the residential property valuations for Salerno, Italy. Along with the evaluation of the predictive performance of the algorithms by means of R^2 , RMSE, MAE and MAPE metrics, Explainable Artificial Intelligence approach was adopted to interpret the contribution of each feature into the process using SHAP and PFI.

It was found that the Random Forest algorithm had the highest combination of predictability and interpretability among all the algorithms analyzed. Moreover, the SHAP values showed that the two main factors influencing housing prices were the size of property and its location near transport facilities. Finally, the importance of finding ways to prevent overfitting was stressed because machine learning models are often trained on small datasets of real estate prices.

Though the above contributions are substantial, the suggested approach only depends upon structured numerical property attributes without the inclusion of property images and feature extraction based on deep learning techniques. This results in the exclusion of important visual aspects such as architectural design, physical condition, quality of interior finishing, and aesthetic appeal of the property in the valuation process. Furthermore, although the study achieves success

in improving the transparency of the model using SHAP and PFI approaches, an intelligent decision support system including property searching is not provided.

Within the context of this study, the above-mentioned limitations have directly led to the formulation of the Clarivision AI framework being proposed. As can be seen from the comparison of Clarivision AI with the approach suggested by Maselli and Nesticò (2025) [2], Clarivision AI utilizes the property structured attributes in combination with the deep visual representations generated by ResNet50, employs Explainable Artificial Intelligence along with multimodal learning, and enables to create a decision-making system that is capable of property valuation, similar properties retrieval, creating interactive dashboards, and generating automatic PDF reports **Maselli and Nesticò (2025)[2]**.

- **Kee & Ho (2025)** In this study, Kee & Ho (2025) investigated the use of Explainable Artificial Intelligence (XAI) in residential property valuation through the analysis of housing transaction data obtained from Quarry Bay, Hong Kong. The framework designed by **Kee and Ho (2025) [4]** used the Extreme Gradient Boosting (XGBoost) model for predicting the price of the property and then used Shapley Additive Explanation (SHAP) for determining the contribution of each of the attributes of the property towards the predicted price. It was observed that the XGBoost model was able to capture the intricate non-linear dependencies among the different properties, and at the same time SHAP gave transparent results by finding out the importance of the different influential features like floor area and property age.

Among the key benefits of the current research is the efficient combination of machine learning and explainable artificial intelligence to mitigate the black-box challenge that is inherent in any advanced predictive model. Using SHAP approach, the model ensures the provision of both global and local explanations for model prediction results, which enhances transparency of the valuation process for the benefit of valuers, investors, and policy-makers. On the other hand, the suggested approach uses only structured data about housing transactions and does not take into consideration the images or deep visualizations of property under discussion **Kee & Ho (2025) [4]**.

From the point of view of the current research, this limitation can be considered as one of the critical gaps in the existing literature. Although Kee and Ho have contributed to model transparency by using SHAP, Clarivision AI is another development in this context which involves the idea of the integration of structured property attributes with the deep visual features obtained from ResNet50. In addition, Clarivision AI also includes similar property retrieval, an interactive dashboard, and automated reporting as well.

- **Despotovic et al (2023)** introduced an Automated Real Estate Valuation methodology which involves the integration of visual data from images of properties in existing Automated Valuation Models (AVMs). As opposed to the standard valuation methodologies that only involve numerical features of the property, the researchers sought to establish whether visual quality ratings of the property could assist in improving price estimation of properties. To accomplish the stated objective, the researchers used the hedonic pricing method together with human rating of quality and Convolutional Neural Networks (CNNs) to estimate quality of residential properties based on images of the properties. From the experiments, it was established that the incorporation of visual data improved accuracy in valuation and there existed a strong relationship between visual quality ratings and property prices.

One of the main achievements of the current study includes the effective implementation of computer vision within the framework of automated valuation. The suggested framework showed that the use of deep convolutional neural networks allows extracting visual properties from property images and estimating subjective quality ratings, which will positively affect house price estimation. In addition, the study under consideration pointed out the importance of using multimodal information, including the combination of structured property features and visual quality, thus improving the predictive ability of Automated Valuation Models.

It should be also noted that there are several significant drawbacks associated with the study. Firstly, the authors mentioned that the suggested framework was designed specifically as a baseline and needed optimization. Secondly, the authors stressed that more data is necessary for further enhancing the generalization ability of the neural network. Thirdly, the proposed model

concentrated solely on visual quality prediction and its incorporation into a hedonic pricing model. In other words, there is no XAI technique used in the framework, and no interactive decision-making support system. There are also no mechanisms related to similar property retrieval and automated reporting.

In terms of the present study, these limitations become the reason for the development of the proposed Clarivision AI framework. In the process of confirming the importance of visual information for estimating property values, **Despotovic et al. [6]** managed to prove the usefulness of visual features. The novelty of the proposed Clarivision AI framework consists in the combination of structured housing features with deep visual features that were obtained using ResNet50 architecture and Explainable Artificial Intelligence techniques **Despotovic et al. [6]**.

- **Gunes (2024)** The paper by Gunes (2024) developed an explainable machine learning approach for evaluating residential property that aimed to enhance the transparency and interpretability of automated valuation models. Instead of developing a novel prediction approach, the paper explored the application of explainability methods to existing machine learning algorithms in order to provide an understandable explanation for decision-making processes. Several machine learning algorithms were used for property price prediction, and then the model-agnostic methods for explanation were utilized to evaluate the impact of particular property features on the predicted value. The experiment showed that explainable machine learning is able to increase the users' understanding of decision-making process without compromising predictive accuracy.

An important strength of the current study is that it places a great deal of focus on transparency in decision-making of automated valuation models. Through the use of model-agnostic explanations techniques, the system allows the valuer, investor, and policymaker to analyze the impact of individual attributes irrespective of the prediction model being used, thus making machine learning credible and useful in property valuation.

However, there are several drawbacks of the framework that have been presented in the paper. The framework uses only property attributes and no property images or visual features are taken into consideration during the process of valuing property. Moreover, the research work focuses

mainly on explanation and prediction but fails to offer any intelligent functions such as similar property search, dashboard visualization and automation of reporting function **Gunes (2024) [10]**.

- **DoRA – Du et al. (2023)** One of the key issues in automatic real estate evaluation is the lack of labeled housing datasets needed for training deep learning models, and thus Du et al. (2023) presented DoRA (Domain-Based Self-Supervised Learning Framework). The framework used self-supervised learning to obtain strong visual representation from large datasets of unlabeled real estate images before adapting it to perform real estate appraisal tasks. By using domain-specific properties of the images, the framework greatly lowered the need for labeled training datasets and enhanced the ability of feature extraction in residential property evaluation.

The experiments showed that the self-supervised learning method proposed was consistently better than the traditional supervised learning approach when there were few datasets available. Additionally, the framework also showed great generalization capabilities on residential property evaluation with various datasets and domain-specific pre-training could greatly improve the accuracy of automatic property evaluation.

However, this framework lacks several features that may hinder the efficient valuation process of real estate properties. First, although there is much emphasis on image-based representation learning, the study does not consider the integration of structural data related to the attributes such as size of the house, number of rooms, age of the property, and other physical parameters. Second, there is no use of Explainable Artificial Intelligence (XAI) in order to explain the predictions made by this approach **DoRA – Du et al. (2023) [11]**.

- **Lin Deng et al. (2025)** A framework for automated property valuation based on the integration of visual and structured information was developed by **Lin Deng et al. (2025) [15]**. In contrast to traditional AVMs, the proposed method is based not only on the analysis of structured information about property characteristics but also includes images of street views, satellite pictures and property photos, allowing the model to take into account structural and environment-related factors influencing property price. For extracting visual features, deep

learning was used; then these visual features were integrated with structured features in one machine learning system.

Experimental results show that the use of different types of visual data leads to higher accuracy of valuation in comparison with models based only on structured data. Moreover, the study proves that the use of visual data about the environment obtained from street-view images and satellite pictures gives useful context information.

One of the key benefits of this research is the effective implementation of multi-source image fusion technique, which enables the valuation model to leverage complementary information through multiple viewpoints. Nevertheless, although better accuracy results have been achieved, the methodology lacks the inclusion of Explainable Artificial Intelligence (XAI) methods that can explain the model outputs. Furthermore, the system developed in the paper does not offer intelligent features like similar property search, interactive dashboard, etc. **Lin Deng et al. (2025) [15]**.

- **Geerts et al. (2025)** have researched the potential of Large Language Models (LLMs) for making residential real estate appraisals based on structured descriptions of properties. The study did not propose a novel method of valuation, but assessed a number of the latest LLMs including GPT-4o-mini, Llama 3.1, and Llama 3.2 in comparison with LightGBM, the most popular machine learning model, using different datasets of real estate from Belgium, Barcelona, Beijing, and King County. In order to increase the accuracy of predictions, the researchers used structured prompts that involved three sequential tasks: price prediction, uncertainty quantification, and importance of features identification.

It is evident that LLMs are able to produce reasonable valuations of residential properties based on textual description of structured housing features. Also, the proposed prompting methodology allows estimating confidence intervals and identifying the key hedonic factors influencing the predicted prices. The comparison with LightGBM model shows that language models were able to recognize the most important structural housing features, yet they were less effective in

capturing spatial and temporal features, such as geographic coordinates and transaction dates, which still could be modeled better with traditional ML approaches.

The significant contribution of this paper is the possibility for LLMs to produce not only property valuations but also explainable interpretations through the design of specific prompts. Yet the proposed methodology is based exclusively on textual descriptions of structured housing features and does not consider images or deep visual representations generated with the help of CNNs. Therefore, visual factors like architecture, condition of construction, interior finishing, and façade aesthetics cannot be used within this framework. In addition, there are no intelligent functionalities, such as similarity search, interactive dashboard, and automatic valuation report generation.

Furthermore, the proposed framework does not provide an integrated intelligent decision-support system Geerts et al. (2025) [12].

Figure 2.1 Overview of Previous Studies

Author	Study Name	Method Used	Dataset	Advantages	Disadvantages	Restrictions / Research Gap
Wang et al. (2022)	Graph Neural Network-Based AVM	Graph Neural Network (GNN)	Residential property transactions	Models' spatial relationships effectively	Uses structured data only	No image features or explainability
Mora-Garcia et al. (2022)	Automated Property Valuation Using ML	Machine Learning	Residential housing dataset	High valuation accuracy	Ignores visual property characteristics	No multimodal learning
Du et al. (2023)	DoRA: Domain-Based Self-Supervised Learning Framework	Self-Supervised Learning	Property image dataset	Learns robust image representations	Focuses on images only	Does not integrate structured data
Yazdani & Raissi (2023)	Explainable AI for Property Valuation	XAI + Machine Learning	Real estate transactions	Transparent valuation decisions	Structured data only	No image analysis
Hasan et al. (2024)	Deep Learning-Based AVM	Deep Learning	Housing dataset	Improved prediction accuracy	Limited model interpretability	No explainable AI
Balila & Shabri (2024)	Hybrid Real Estate Valuation	Hybrid Machine Learning	Residential properties	Better predictive performance	Uses numerical data only	No visual features
Rodríguez-Serrano et al. (2024)	AI-Based Property Valuation	Machine Learning	Property transaction dataset	Accurate valuation	No property image analysis	Limited explainability
Gunes (2024)	Model Agnostic Interpretable Machine Learning for Residential Property Valuation	Model-Agnostic Explainable ML	Residential property dataset	Transparent AI decisions	Structured data only	No multimodal learning
Kee & Ho (2025)	Explainable Machine Learning for Real Estate	XGBoost + SHAP	Quarry Bay housing transactions	High accuracy and explainability	Structured data only	No deep visual features
Huang et al. (2025)	Multimodal Machine Learning for Real Estate Appraisal: A Survey	Survey (CNN, ViT, GNN, Multimodal AI)	Multiple public datasets	Comprehensive review of multimodal AI	No practical valuation framework	Highlights need for XAI and decision support
Maselli & Nesticò (2025)	Machine Learning Algorithms and Explainable Artificial Intelligence for Property Valuation	Random Forest + SHAP + PFI	Salerno residential dataset	Explainable and accurate valuation	No property image analysis	No multimodal framework
Deng et al. (2025)	Multimodal Machine Learning for Residential Property Valuation	XGBoost + LightGBM + CNN + SHAP	Structured data + Street View + Satellite + Images	Integrates multimodal features effectively	No similar property retrieval	No interactive decision-support system
Geerts et al. (2025)	On the Performance of Large Language Models for Real Estate Appraisal	LLMs + Prompt Engineering	Belgium, Barcelona, Beijing & King County datasets	Explainable valuation using LLMs	Uses textual structured data only	No visual information

Moreover, the existing body of literature shows that no previous work has suggested an overall decision support system model that considers all of the above aspects in one intelligent system, namely, hybrid property valuation, explainable artificial intelligence, similar property search, interactive visualization, and automatic report generation. The identified research gaps motivate the current research project and thus justify the introduction of the Clarivision AI model. In this regard, the following chapter discusses the research methodology of the study.

Chapter 3 — Research Methodology

3.1 Introduction

This chapter presents the research methodology adopted to develop the proposed Clarivision AI system for real estate valuation. The methodology describes the complete workflow followed throughout the study, beginning with data collection and preprocessing, followed by image feature extraction and structured data processing. The chapter then explains the development of the proposed hybrid artificial intelligence model, which combines deep image features extracted using ResNet50 was selected because it offers an excellent balance between feature extraction capability, computational efficiency, and transfer learning performance for image-based applications. with structured property attributes and employs the XGBoost regression algorithm for property price prediction. Finally, the evaluation methodology and performance metrics used to validate the proposed model are presented.

The proposed methodology aims to improve the accuracy and reliability of automated property valuation by integrating visual property characteristics with structured real estate information into a unified prediction framework. In addition, the developed system incorporates an intelligent recommendation module that retrieves visually similar properties within the same city, thereby providing practical support for real estate professionals and financial institutions during property valuation and decision-making.

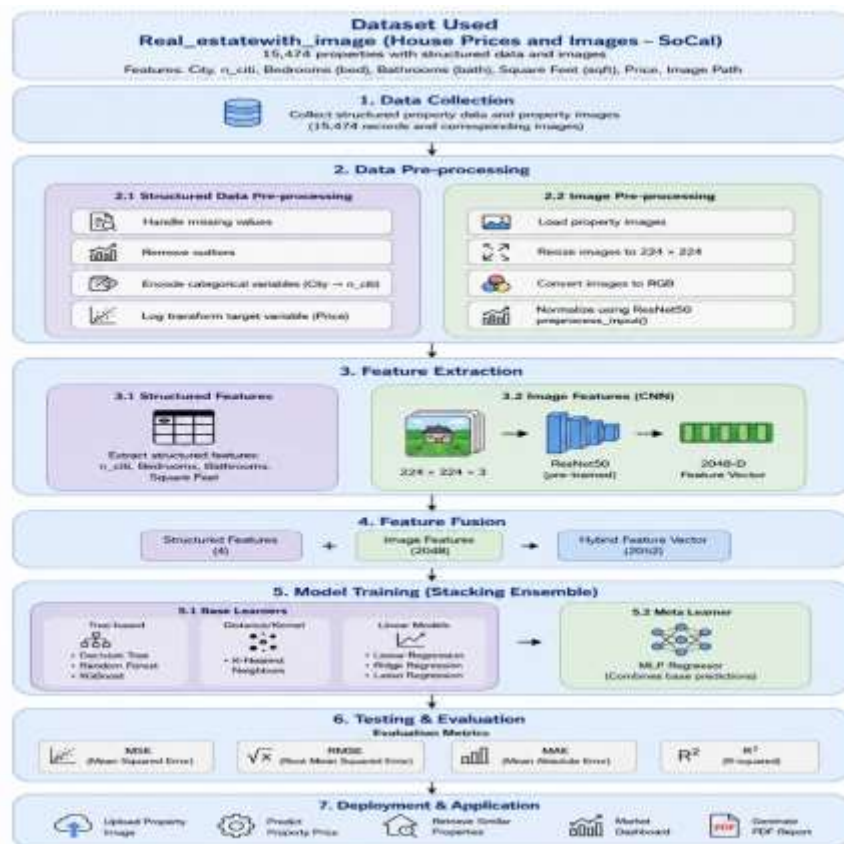


Figure 3.1: General Methodology of the Proposed Clarivision AI Framework.

shows the entire flow of operations of the proposed hybrid real estate appraisal model, which comprises data preprocessing, image features extraction, hybrid feature fusion, price prediction, similarity-based real estate property search, and decision-making support.

3.2 Proposed Hybrid Framework.

- The proposed Clarivision AI framework was designed for enhancing the accuracy of predicting real estate prices using the combination of structured property attributes and visual information contained in property images. Differently from other existing Automated Valuation Models (AVM), the Clarivision AI framework considers not only structured property attributes but also deep image features that are extracted from property pictures.
- The presented framework includes several consecutive steps. Firstly, the dataset named `Real_estatewith_image` is obtained and processed. Structured property attributes are subject to preprocessing that comprises such steps as data cleansing, city encoding, feature selection, and logarithmical transformation of property prices. At the same time, all property images are resized, converted into RGB format, and then preprocessed. After that, the resized property images are fed into the pre-trained convolutional neural network ResNet50. ResNet50 network generates a 2048-dimensional deep feature vector that characterizes visual property features.
- Next step is the concatenation of deep image features and structured property attributes, that is city, number of bedrooms, number of bathrooms, and square footage. Obtained hybrid feature vector is used as an input to the XGBoost Regressor that estimates property market value.
- After the prediction of prices, the suggested framework extracts visually similar houses from the database. In this extraction phase, properties that belong to the same city are filtered out initially, followed by calculating cosine similarity of the deep image feature vectors. The five most visually similar properties are then displayed along with their property information to help users understand the valuation process.

Finally, the predicted value of the property, analysis of the model, market dashboard, and PDF report of valuation of the property are generated using Clarivision AI decision-making system.

3.3 DataSet Description

The suggested Clarivision AI framework is built based on the (Real_estatewith_image) dataset. This data set was chosen due to the presence of structured data and images related to properties, which makes it appropriate for building an artificial intelligence framework, which combines numerical and visual features for property valuation.

The dataset comprises 15,474 property records each containing a picture of a property and a set of structured features characterizing the property. Structured features include the city, number of bedrooms, number of bathrooms, square footage, street, and the price of a property. Both the picture and structured data make it possible to incorporate numerical and visual features into the suggested framework.

Even though the original dataset includes various variables, only those that are necessary for the prediction process have been included into the model development stage. The following structured features have been used:

- ✓ City(cit) (numerical representation)
- ✓ Bedrooms
- ✓ Bathrooms
- ✓ Square Feet (Sqft)

The target variable for this analysis is the market price of property that was subjected to transformation through a logarithmic function while building the model.

Besides structured property features, each image of the property was fed to the pre-trained ResNet50 deep learning convolutional neural network. Images were encoded into 2048 dimensional vectors representing features extracted from the image. The resulting image features were then concatenated with structured property features to form hybrid feature vector.

Table 3.1: Structured Features Used in the Proposed Model

Feature	Description	Type
City	Encoded city identifier	Categorical
Bedrooms	Number of bedrooms	Numerical
Bathrooms	Number of bathrooms	Numerical
Square Feet	Property area (sqft)	Numerical
Property Image	Exterior property image	Image
Property Price	Target variable	Numerical

3.4 Data Collection.

Data collection constitutes the initial phase of building the suggested Clarivision AI system. The (Real_estatewith_image) real estate dataset was chosen as the primary source of data for this study. This publicly accessible dataset was selected from Kaggle because it offers both structured property information and property images, making it ideal for the development and testing of hybrid artificial intelligence systems for automated property valuation.

This dataset consists of 15,474 records of residential properties characterized by each property real estate image as well as several descriptive characteristics. These characteristics include city, street, number of rooms, number of bathrooms, square footage, and market price. Availability of structured data and corresponding property image allows the suggested approach to utilize complementary information provided by multiple data types.

The collected dataset was verified prior to modeling, ensuring the presence of a valid property image for each structured record. This check was necessary due to the nature of the hybrid approach, which requires synchronization of structured and image data for each training sample.

After the data collection, the dataset was processed through multiple preprocessing steps prior to the use for training and evaluation.

3.5 Data Pre-processing.

Data pre-processing is one of the most critical steps in building an artificial intelligence model since the quality of the data used influences the accuracy of the predictions. In preparing the proposed hybrid model, the obtained data set was processed to enhance consistency, reduce data quality challenges, and ensure that there was proper representation of both the structured property data and images.

There were two major processes during the pre-processing of the data set, namely data cleaning and feature engineering.

3.5.1 Data Cleaning

A preliminary investigation of the structured dataset of real estate properties was made in order to determine any possible problems related to data quality. Missing data, duplicated data, and inconsistent data were assessed in order to be sure that only valid data is used in the learning process. Properties with missing data or invalid links to images were omitted in order to avoid inconsistencies between structured data and its corresponding images.

In addition, data analysis was conducted in order to find any abnormal data which might have a negative impact on the performance of the model.

3.5.2 Feature Engineering

After performing data cleansing, a number of feature engineering steps were carried out in order to enhance the learning capacity of the developed model. The only attributes used for estimating the price of the property are the city of residence, the number of bedrooms, the number of bathrooms, and the area.

As machine learning algorithms do not work with categorical attributes, city names were coded numerically using categorical encoding. Additionally, the target attribute – price of property – is also coded using a log function, which will reduce the skewness of the price.

Finally, this structured feature vector is merged with the deep image features from ResNet50 to create the hybrid feature representation.

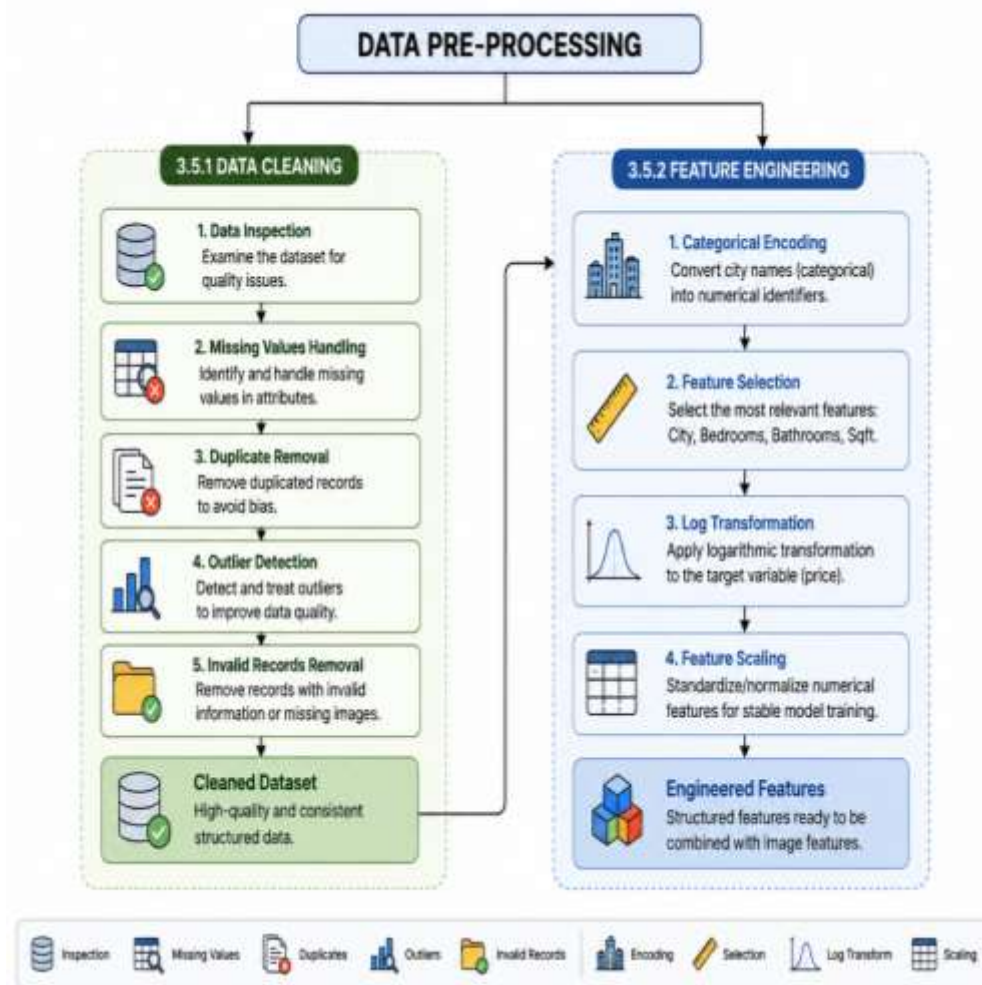


Figure 3.2: Data preprocessing workflow adopted in the proposed Clarivision AI framework.

summarizes the complete preprocessing workflow adopted before training the hybrid prediction model.

3.6 Property Image Processing

Property images contain valuable visual information that is impossible to encode in structured numerical attributes. Thus, an image processing pipeline was designed to extract high-level visual features describing the physical aspects of each property. The resulting deep features were combined with structured property attributes in order to increase prediction performance of the suggested hybrid approach.

First of all, all property images that were uploaded were converted to the RGB color scheme to make sure that they would be compatible with the pre-trained convolutional neural network. After that, all property images were resized to the size of (224×224 pixels), which corresponds to the standard input size for ResNet50 architecture. Before performing feature extraction, images were normalized via the (ResNet50 preprocessing function) **ResNet50**, a convolutional neural network (CNN) pre-trained on the ImageNet dataset.

After preprocessing, the images were fed to the pre-trained (ResNet50) convolutional neural network. Instead of using the output of the final classification layer, the classification head was omitted by using parameter (include top=False), while the global average pooling (pooling='avg') was used to produce a compact image representation. In result, each property image was converted into the (2048-dimensional deep feature vector).

The extracted deep image features carry high-level semantics such as architectural appearance, façade type, exterior quality, among others, that are learned by the convolutional neural network. The feature vectors from this process were then merged with the property attributes in the hybrid feature fusion phase.

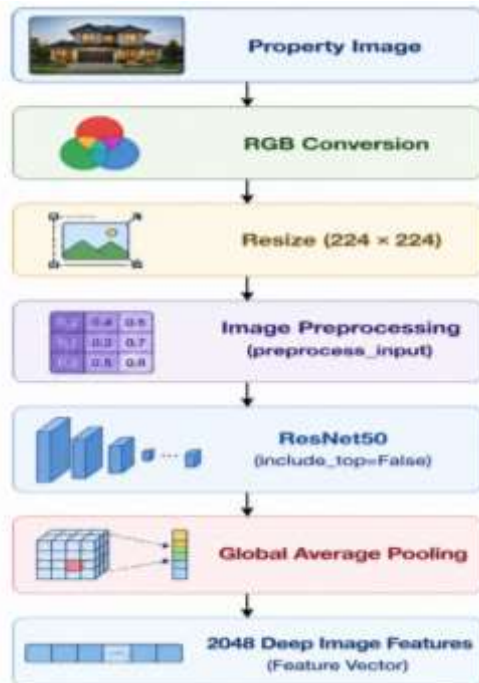


Figure 3.3: Pipeline for Property Image Processing.

3.7 Hybrid Feature Fusion.

The proposed Clarivision AI system employs a combination of structured property attributes and deep image representations to form a unified hybrid feature vector for property value estimation. The hybrid representation ensures that the prediction model can make use of both numerical and visual property information simultaneously.

The structured data comprise four property attributes: encoded city name, number of bedrooms, number of bathrooms, and size of the property in square feet. The four variables represent the physical and geographical aspects of the property.

On the other hand, every property image is analyzed using the pre-trained ResNet50 convolutional neural network. Through feature extraction, every property image is transformed into a 2048-dimensional (4 structured features + 2048 deep image features) deep feature vector capturing the high-level visual content of the image learned from the property image.

Lastly, the structured data are combined with the deep features extracted from the image to obtain a 2052-dimensional hybrid feature vector to be used as the input to the XGBoost regression model.

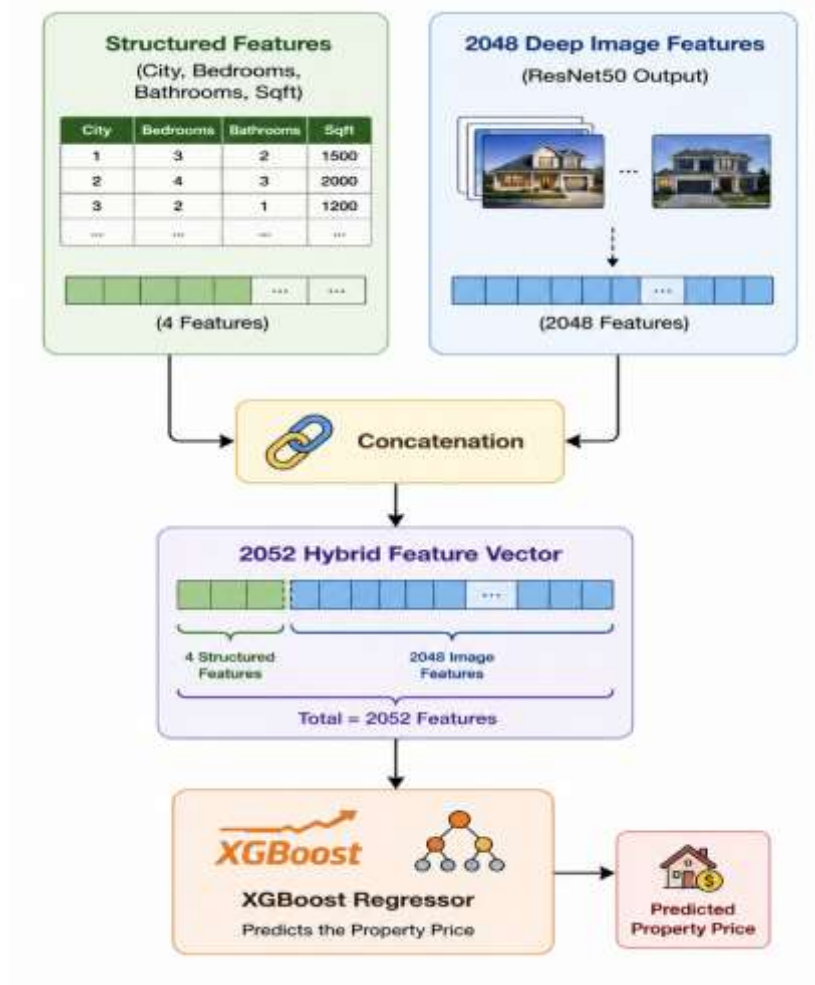


Figure 3.4: Hybrid Feature Fusion Process.

3.8 Model Training

With the construction of the hybrid feature vectors, the dataset was split into the training set and the testing set for developing and evaluating the proposed prediction model. Hybrid feature vector, which includes structured property attributes and deep image features by ResNet50, was selected as the input of the XGBoost regression algorithm.

3.8.1 Data Splitting

For a more unbiased assessment of the proposed model, the dataset was split into two subsets:

- **80% Training Set:** Used to train the XGBoost regression model to identify the relationships between the hybrid features and property prices.
- **20% Testing Set:** Used to assess the prediction performance of the trained model on new data.

This train/test splitting process makes the evaluation of the model to be conducted under realistic conditions and gives an objective measure of the generalization capability of the model.

3.8.2 Cross-Validation

In addition to the data splitting process, 5-Fold Cross-Validation technique was also used during the modeling process to test the robustness and stability of the proposed prediction model.

The entire dataset for training was partitioned into five equally sized sets. In every run, four of the sets were used for training the model, whereas the other one was used for validation. This process was performed five times so that every subset was eventually used for the purpose of validation at least once. The overall performance was computed as the average of the five performances obtained by the validation process.

Cross-validation helps in avoiding overfitting of the model and gives a better estimate of its prediction performance.

Shuffle=True and Random State=42 were applied to ensure reproducibility.

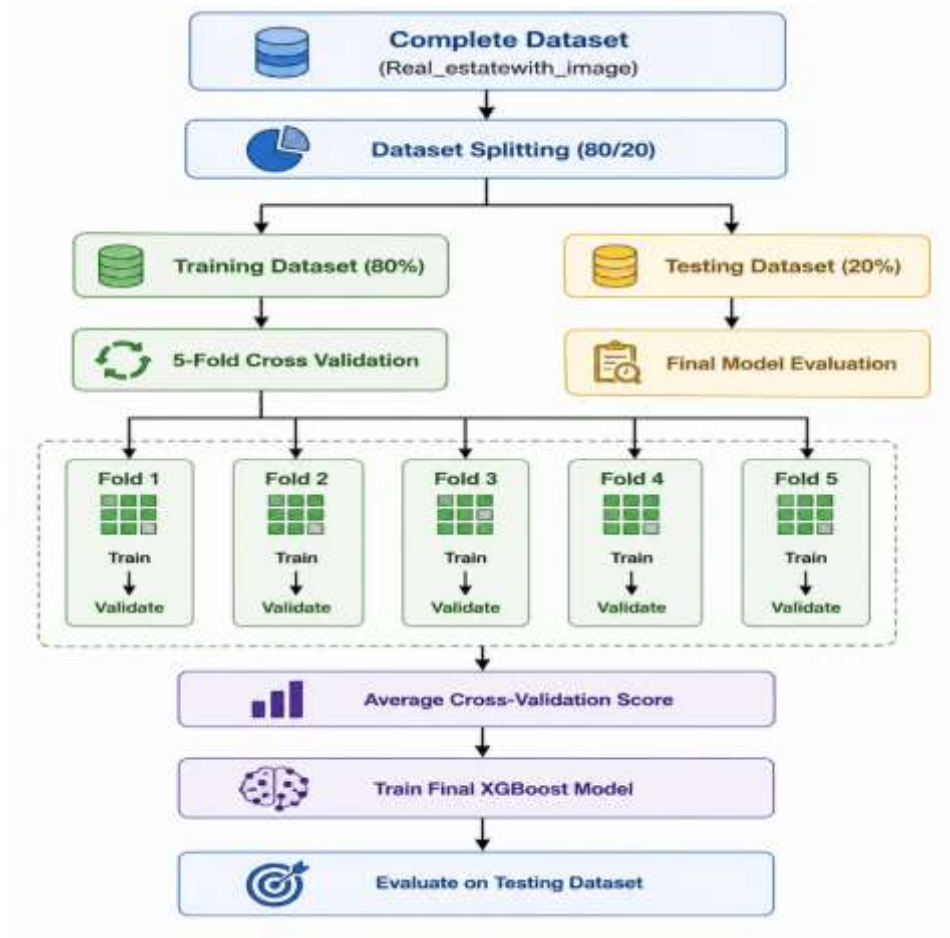


Figure 3.5: Dataset Splitting and Cross-Validation Process

3.9 Hybrid Prediction Model

After the feature fusion, the created 2052-dimensional hybrid feature vector became the input of the prediction model. The proposed Clarivision AI uses the Extreme Gradient Boosting (XGBoost) Regressor to predict the market price of residential houses.

It was chosen due to its effectiveness among the algorithms of ensemble learning that is used for structured regression tasks. As compared to traditional regression models, XGBoost offers increased accuracy of predictions, fast training, built-in regularization functions, and proper consideration of nonlinear dependencies between variables. All these properties of XGBoost made this algorithm the best fit for the problem of real estate valuations when the property price depends on the interaction of its attributes and visual characteristics.

As stated above, the hybrid feature vector includes four structured attributes (city, bedrooms, bathrooms, and square feet) along with the 2048 deep image features obtained with the help of pre-trained ResNet50 model. The final 2052-dimensional feature vector allows XGBoost to take advantage of both numerical and visual features at once.

In order to improve the learning accuracy of the regression model and minimize the effect of outliers in property prices, the target variable (property price) was log-transformed prior to training. When predicting the property price, the logarithmic value obtained from the training process was reversed to the original scale by means of exponentiation.

The learned XGBoost regression model was saved as `true_hybrid_model.pkl` and then incorporated into the Clarivision AI application. The final trained model was stored as **`true_hybrid_model.pkl`** and deployed within the Clarivision AI application.

3.9.1 Transformation of Target Variable

The distribution of property prices had a positive skew, which could harm the learning procedure for the regression algorithms. In order to minimize skewness and normalize the variance, the target variable (property price) was log-transformed prior to training the models.

The formula for log-transformation is:

$$\text{Price}_{\{\log\}} = \log(1 + \text{Price})$$

where:

* (Price) is the actual property price.

* (Price_{log}) is the transformed target value used during the training of the model.

Once the XGBoost model provided the logarithmic prediction, the actual property price was reversed to its original scale through the inverse of the exponential function:

$$\text{Price} = e^{\{\text{Prediction}\}} - 1$$

where:

* (Prediction) is the logarithmic property price predicted by the XGBoost model.

* (Price) is the final estimated property price that is shown to the user.

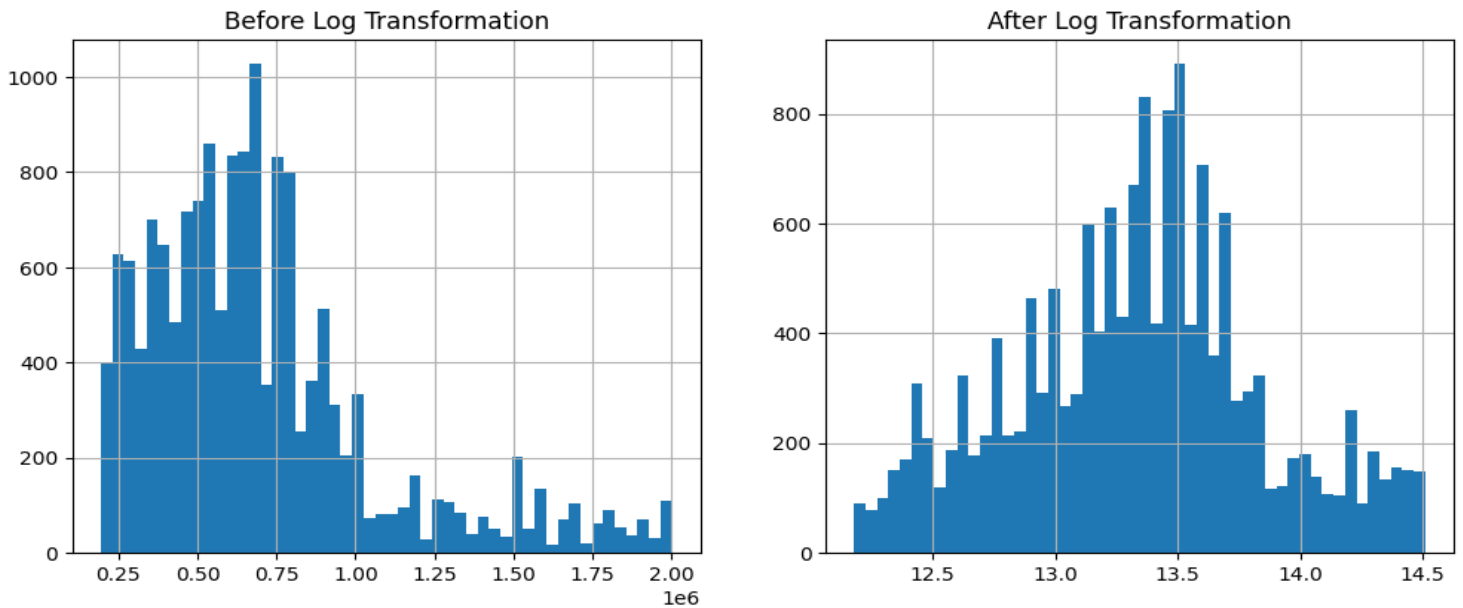


Figure 3.6 Distribution of Property Prices Before and After Log Transformation.

The distribution of property prices before and after the logarithmic transformation is shown in Figure 3.6. The property prices before the logarithmic transformation had a very skewed distribution on the right side. However, after the logarithmic transformation, the distribution had a symmetrical pattern and was close to the normal distribution. This reduced the impact of outliers on the learning capability of XGBoost regression model.

Table 3.2 Hyperparameters of the XGBoost Model.

Hyperparameter	Value
Objective	reg:squarederror
Number of Estimators (n_estimators)	300
Learning Rate (learning_rate)	0.05
Maximum Depth (max_depth)	6
Subsample (subsample)	Default (<i>Not specified</i>)
Colsample_bytree (colsample_bytree)	Default (<i>Not specified</i>)
Random State (random state)	42

The XGBoost regression model was configured using a set of hyperparameters to optimize prediction performance. The selected hyperparameters used in the final hybrid prediction model are summarized in Table 3.2.

3.9.2 Explanation of the Selected Hyperparameters

The final XGBoost regression model was designed utilizing an appropriately chosen set of hyperparameters that achieved the best trade-off between accuracy of predictions and model complexity. Description of the chosen parameters is provided below:

- **Objective (reg: squared error).**

As the objective function in XGBoost, the squared error objective function is used due to the fact that the proposed system is designed to perform regression analysis, where the task is to predict the continuous price values of the properties instead of classification of categorical labels.

- **Number of Estimators (300).**

The parameter “number of estimators” indicates the total amount of boosting trees generated by the algorithm during its work. Value of **300** was chosen since it allows the model to have enough learning capacity without losing stability and not causing computational burden at the same time.

- **Learning Rate (0.05).**

The learning rate determines the contribution of the boosting tree into the training process. Small value of **0.05** was chosen for this parameter to allow gradual learning of the model.

- **Maximum Depth (6):**

Maximum depth represents the complexity of each decision tree. Maximum depth of 6 was chosen as a proper tradeoff between making the model capture complicated patterns of the real estate data and avoid overfitting due to high complexity.

- **Subsample (Default):**

This hyperparameter is responsible for the proportion of training samples being used for each tree construction. As there was no manual configuration, the default settings were used.

- **Colsample_bytree (Default):**

It is the proportion of features which are randomly selected during each tree construction process. Default settings were used as it has been proven by preliminary experiments that the model predicts sufficiently well with them.

- **Random State (42):**

A constant random state of 42 was used for the purpose of reproducing the results of the experiment. It makes sure that dataset splitting and model initialization would be the same when running the system repeatedly.

3.10 Similar Property Retrieval.

Aside from estimating the value of the property in the market, the proposed Clarivision AI framework also comes with a property retrieval module which helps users compare their property estimate with other similar-looking properties available in the database. The said function aids in providing better interpretation of the estimation results as well as market references.

Once the price of the property has been estimated, the uploaded property image will be processed through the same ResNet50 feature extraction pipeline used during the training process of the model. This will result in the generation of a 2048-dimensional deep feature vector for the visual features of the uploaded property image.

In order to ensure more relevant property retrieval results, the search will initially be done based on properties within the same city chosen by the user. The reason for this is that the geographical filtration guarantees that the retrieved properties are within the same market as the property whose price was estimated.

Finally, after city-based property filtration, the extracted deep feature vector will be compared with the stored feature vector of all properties within the chosen city using Cosine Similarity.

Cosine Similarity was selected because it effectively measures the similarity between high-dimensional feature vectors extracted from ResNet50 regardless of their magnitude.

The Cosine Similarity is computed as below:

$$\text{Cosine Similarity} = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \times \|\mathbf{B}\|}$$

Where:

A = Feature Vector of the uploaded image of the property.

B = Feature Vector of the stored property image in the database.

The similarity score ranges from 0 to 1, where higher values indicate greater visual similarity between properties.

Top five similar properties in the same city based on the computed similarity values are shown to the user along with the following details:

Image of the retrieved property, Similarity Percentage (Calculated), City, Number of Bedrooms, Number of Bathrooms, Square Footage, and Market Price.

The integration of geographic filtering and visual similarity has made the proposed search/retrieval module capable of providing highly relevant similar properties to validate the estimated market value.

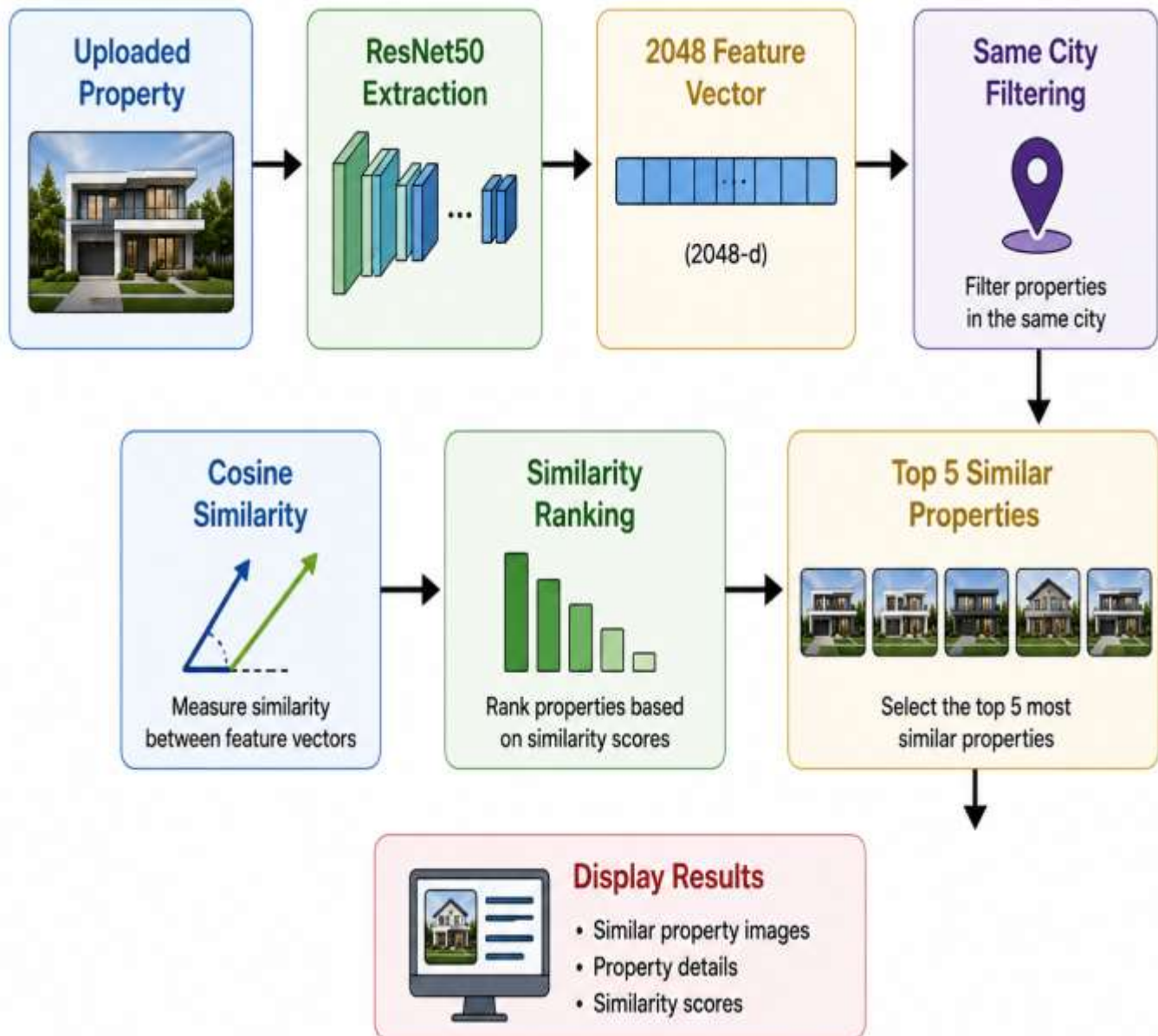


Figure 3.7 Similar Property Retrieval Process

3.11 Model Evaluation.

For the purpose of measuring the efficiency of the proposed hybrid predictive model, a number of commonly accepted regression performance criteria were utilized. These criteria help in obtaining a thorough evaluation of the accuracy of predictions and the reliability of the model as a whole.

This evaluation was carried out based on the testing dataset which had been used after the stage of model training had been accomplished. Given the nature of the task which involves predicting continuous property prices, regression criteria were chosen to measure the performance of the model. **The following evaluation criteria were used within this research:**

- **Mean Absolute Error (MAE)**

Mean Absolute Error (MAE) represents an average absolute deviation of the predicted property prices from the real market prices. The lower the MAE, the higher prediction accuracy and model performance are.

- **Root Mean Squared Error (RMSE)**

Root Mean Squared Error (RMSE) is the average magnitude of prediction errors but with larger weight being assigned to the larger errors. RMSE criterion is especially helpful in identifying those models which make large prediction mistakes.

- **Coefficient of Determination (R^2)**

The coefficient of determination (R^2) provides an index that indicates the proportion of the variability in the property price data set that can be explained by the prediction model. The value of R^2 close to one shows the success of the model in describing the relationship between the input variables and the output variable.

- **Cross-Validation**

In order to test the robustness and generalization ability of the developed model, the process of K-Fold Cross-Validation was used while developing the model. This method tests the model performance on several subsets of data sets and minimizes the possibility of over-fitting the model. The results derived from these evaluation methods are described in Chapter 4.

Table 3.3 Evaluation Metrics Used in This Research.

Metric	Purpose
MAE	Measures the average absolute prediction error
RMSE	Measures the magnitude of prediction errors
R ²	Measures the proportion of explained variance
Cross Validation	Evaluates model robustness and generalization

3.12 System Workflow.

The complete workflow of the proposed Clarivision AI is shown in Figure 3.X. The process starts with the user providing structured information about the property's attributes in terms of the city name, number of bedrooms, number of bathrooms, and size of the property. Furthermore, a property image is provided by the user using the graphical user interface.

After the inputs have been obtained, the structured attribute data are encoded, and the uploaded property image is preprocessed with operations that include converting the image into RGB, resizing the image, and normalizing the image. Subsequently, the preprocessed image is fed into the pre-trained ResNet50 convolutional neural network model, which extracts 2048-dimension vector for the property.

The image-based feature vector is concatenated with the structured attributes of the property to form a 2052-dimension hybrid feature vector. Then, the feature vector is used as an input for the trained XGBoost regression model to predict the market price of the property.

After the prediction of the property price, the system identifies visually similar properties in the database. In order to retrieve visually similar properties, first, all the properties which fall in the same city as chosen by the user are filtered out. Cosine similarity is then used to calculate the similarities between the uploaded property photo and all the photos of the property available in the same city. The top five similar properties based on the similarity score are then displayed to the user along with their respective details.

Finally, the entire valuation report provided by the Clarivision AI system includes predicted property price, Reliability Score, market category, similar properties, feature importance analysis, market dashboard, and the automatic generation of PDF property valuation report.

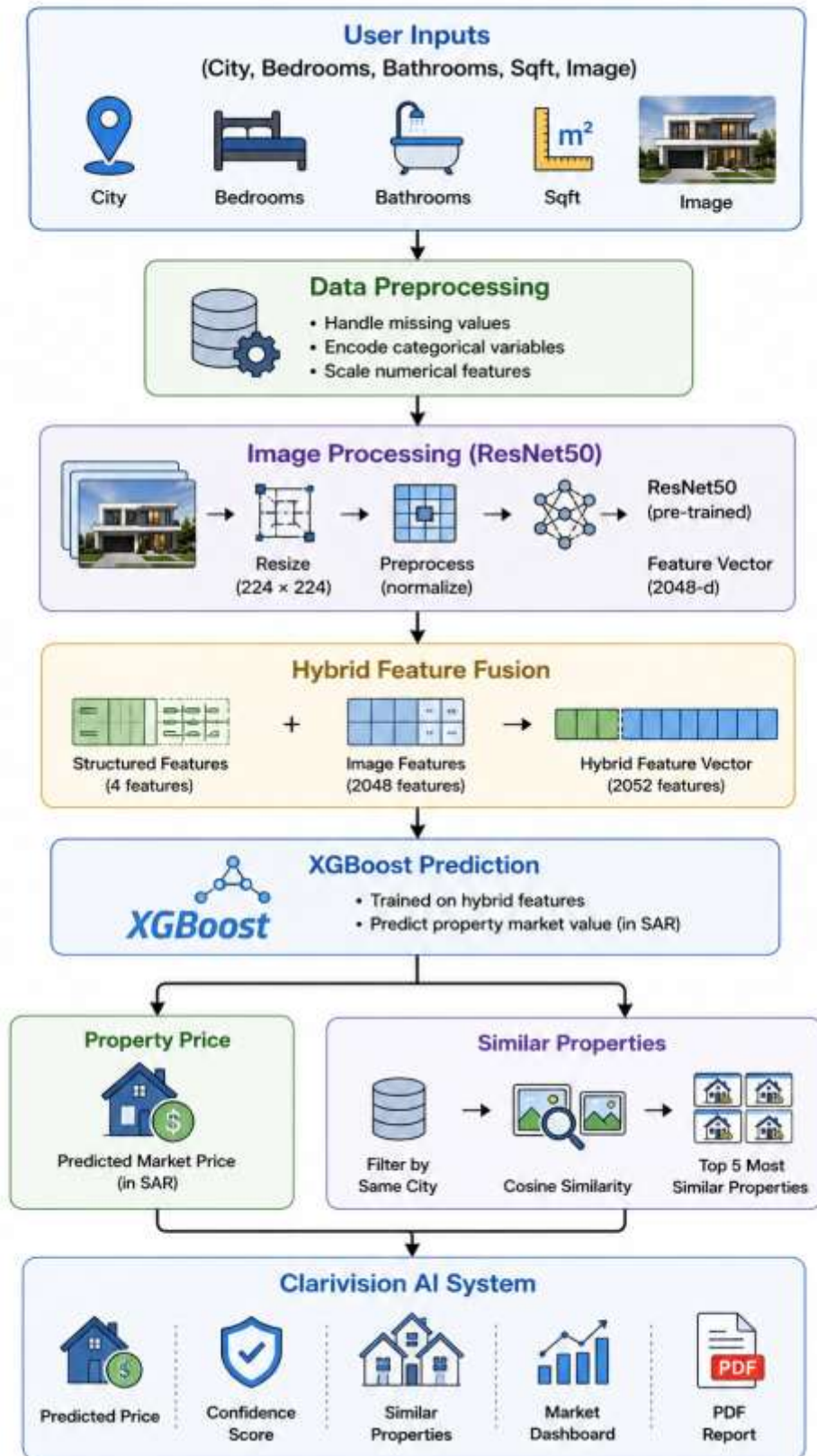


Figure 3.8: Overall Workflow of the Proposed Clarivision AI System.

Chapter 4 — Experimental Results and Discussion

4.1 Introduction:

The following chapter is dedicated to the presentation of the experimental results gained through the Clarivision AI framework that was designed to support real estate valuation. The assessment of the framework will cover exploratory data analysis, model evaluation, cross-validation, feature importance analysis, explainable artificial intelligence (SHAP), search for similar properties, and the dashboard that was created for the presented framework. Lastly, the received results will be discussed and compared with previous research. For the implementation process, Python programming language and some artificial intelligence libraries were used. The developed model provides a user-friendly web interface where users can upload property images, input property details, get market value estimation, find similar property based on visual similarity, visualize results using intelligent dashboard, and create an automatic PDF report.

4.2 Exploratory Data Analysis (EDA).

Exploratory Data Analysis (EDA) has been performed in order to gain an understanding of the attributes of the Real_estatewith_image dataset prior to the evaluation of the proposed hybrid prediction model. The objective of performing EDA was to analyze the structure of the data, look for any quality issues in the dataset, understand the distribution of the main variables, and study the correlations between the structured attributes. The results from this step will help us understand the dataset better and interpret the results from the experiment.

4.2.1 Dataset Overview.

Real_estatewith_image dataset (**House Prices and Images - SoCal**) [16], which is employed in the present research, comprises of 15,474 instances of real estate listings, in which every single listing has a corresponding image. Both structured numeric features as well as the image data are included in the dataset, which makes it possible to create the required hybrid model for real estate valuation. The numeric features comprise of city ID, number of bedrooms, number of bathrooms, and area of the property in square feet, whereas the target feature is the price of every property.

Table 4.1 Dataset Summary.

Attribute	Value
Total Property Records	15,474
Total Property Images	15,474
Number of Features	8
Structured Features	City, Bedrooms, Bathrooms, Sqft
Target Variable	Property Price

Table 4.1 captures the essential features of the Real_estatewith_image dataset, which is utilized extensively in the study. It incorporates structured property data and images, and therefore makes it possible to develop the hybrid prediction model proposed.

4.2.2 Missing Values Analysis.

The missing values were inspected prior to the model evaluation phase to determine the data completeness of the dataset. It is important to detect and manage any missing values due to their influence on the effectiveness of the machine learning algorithms.

The Missing Value Distribution in Structured Features from Figure 4.1 is displayed above. The findings reveal that there were no missing values in any of the selected attributes after the preprocessing step. Thus, it is evident that the dataset was complete and was suitable for training.

the proposed model without applying other missing value imputation methods. Hence, all observations were used for further experiments.

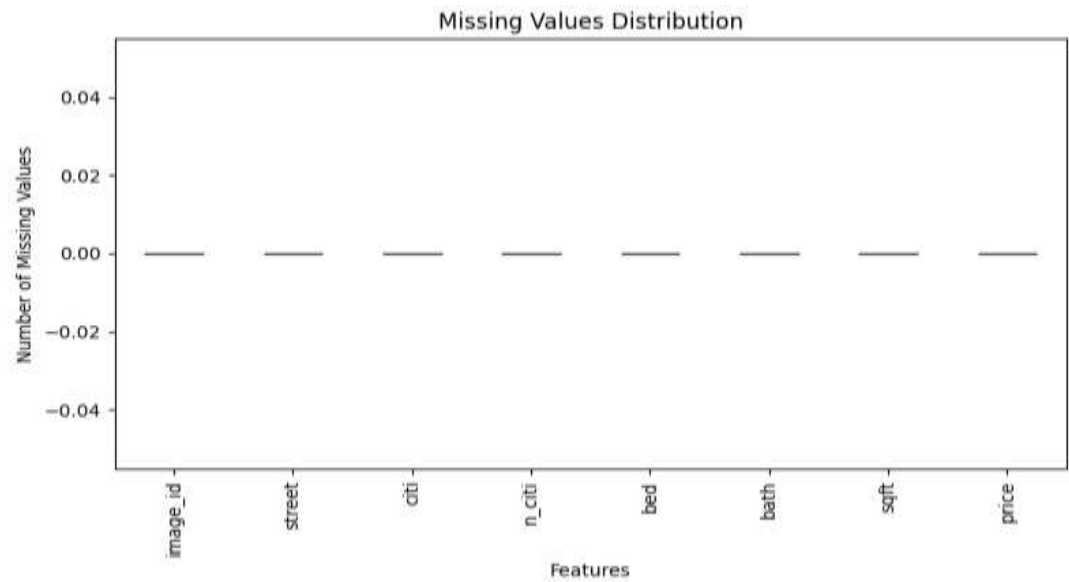


Figure 4.1 Missing Values Distribution in the Real_estatewith_image Dataset.

4.2.3 Distribution of Property Prices.

Distribution of property prices has been studied to get an insight into the features of the target variable prior to assessing the proposed predictive model. The study of distribution of prices offers useful information about the variation in property prices and hence explains the need for using log-transformation in the pre-processing phase.

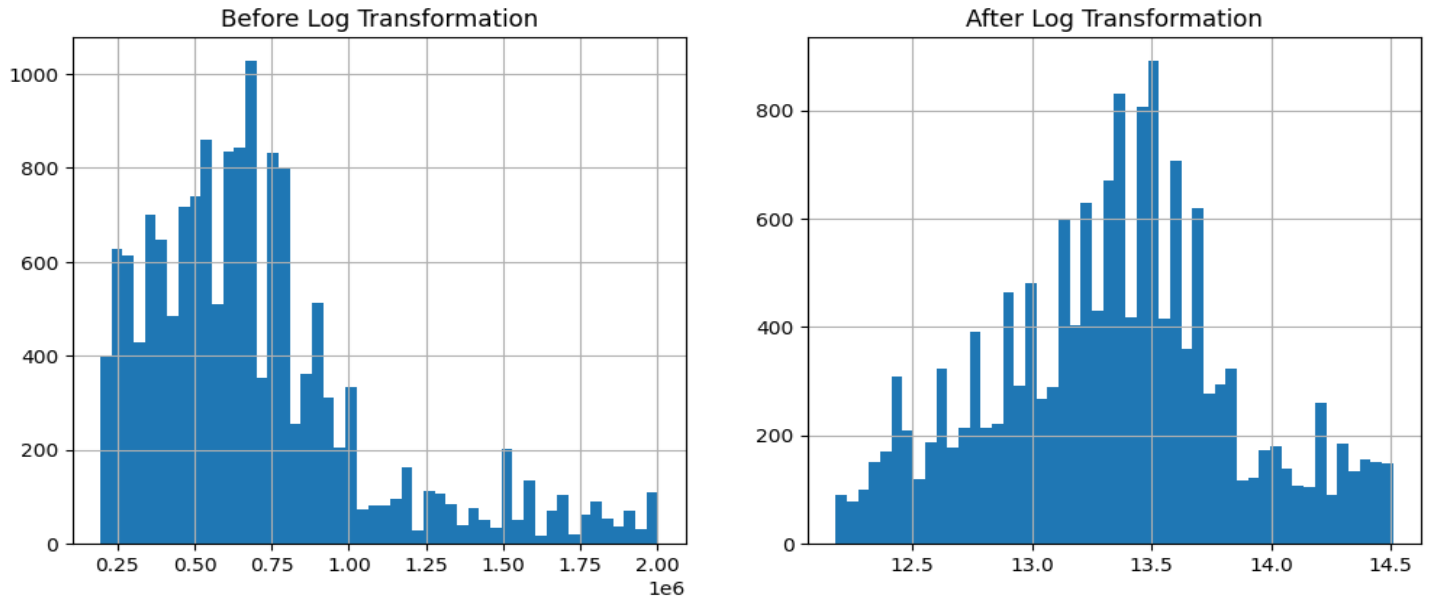


Figure 4.2 Distribution of Property Prices Before and After Log Transformation.

As illustrated in **Figure 4.2**, the original property prices exhibited a highly right-skewed distribution with a long tail caused by expensive properties. Such a distribution may negatively affect the performance of regression algorithms because extreme values dominate the learning process. After applying logarithmic transformation, the price distribution became more symmetric and closer to a normal distribution. This transformation reduced the influence of outliers and produced a more balanced target variable, improving the stability and predictive capability of the XGBoost regression model.

4.2.4 Distribution of Property Characteristics.

This part examines the distribution of the key attributes associated with properties contained in the Real_estatewith_image database. The knowledge about the distribution of these features is important in order to understand the composition of the database and determine the reasons for prices of properties before considering the hybrid predictive model.

4.2.4.1 Distribution of Bedrooms.

Distribution of the number of bedrooms in Figure 4.3 depicts the distribution of the number of bedrooms among the properties found within the data set. From the analysis, it shows that majority of the residential properties have 2 to 4 bedrooms, whereas properties with more than four bedrooms are found in smaller numbers.

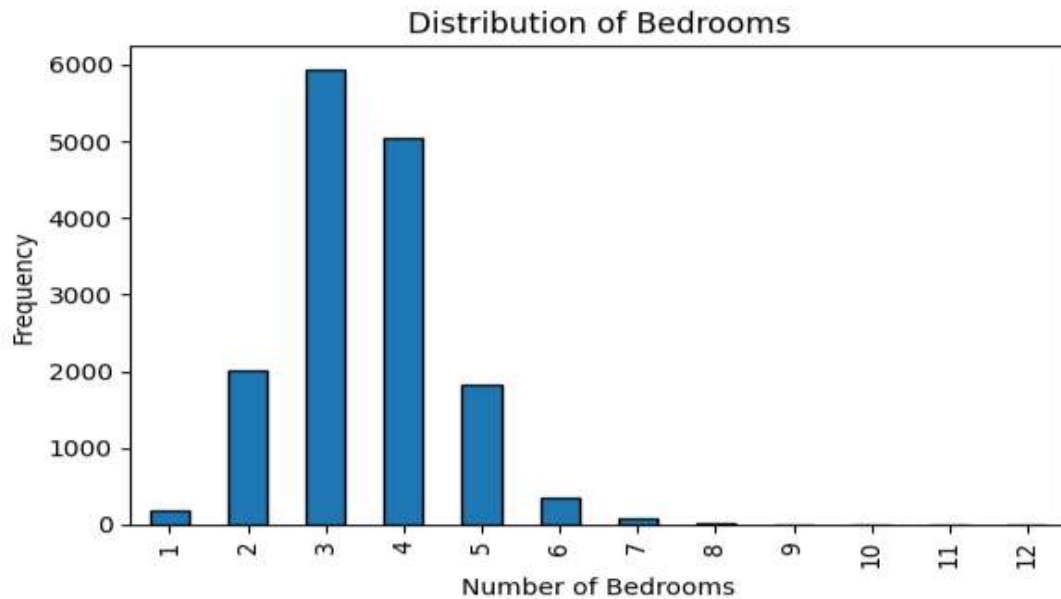


Figure 4.3 Distribution of Bedrooms

4.2.4.2 Distribution of Bathrooms.

The distribution of the bathrooms is shown in Figure 4.4 The same trend as the distribution of bedrooms is seen where most houses have one to three bathrooms, while houses with many bathrooms make up a small portion of the data.

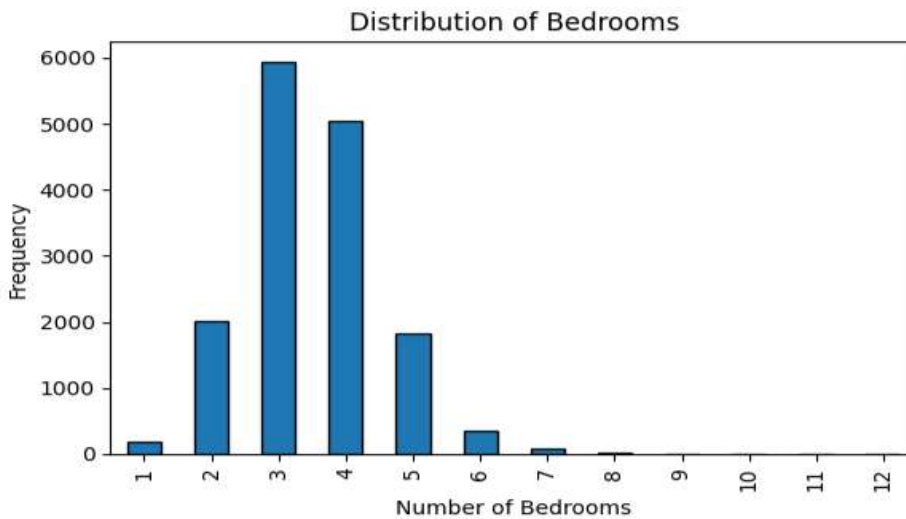


Figure 4.4 Distribution of Bedrooms

4.2.4.3 Distribution of Property Area.

In the Figure 4.5 illustrates the distribution of property size in square feet. From the findings, it can be observed that most of the properties are found in the medium-size category, whereas very large properties are few in number. This pattern is aligned with the features of the residential real estate market depicted in the data set.

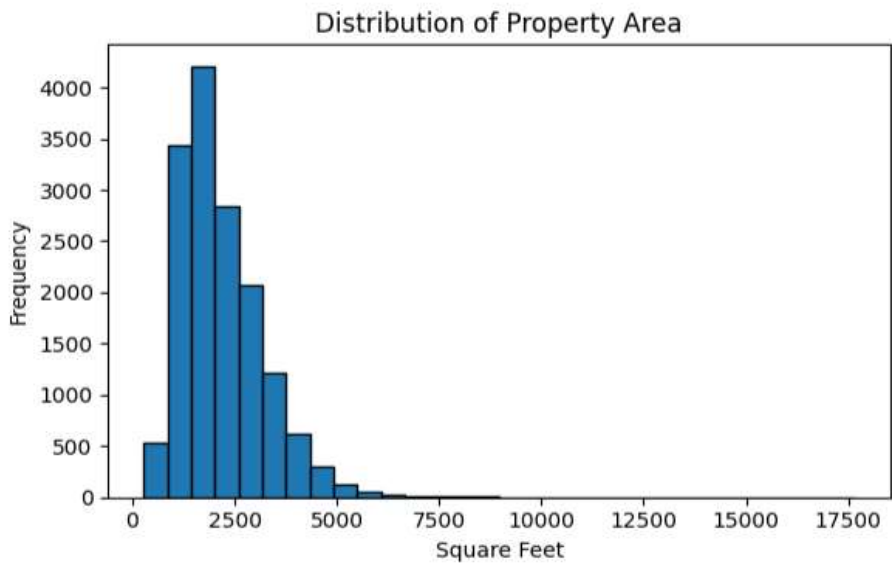


Figure 4.5 Distribution of Property Area (Square Feet).

4.2.5 Correlation Analysis.

Correlation analysis was carried out to identify the correlations between the structured numeric features and the target variable (price of property). This gives insight into the relationship between the numeric features and the price of property before conducting an evaluation on the suggested hybrid model of prediction.

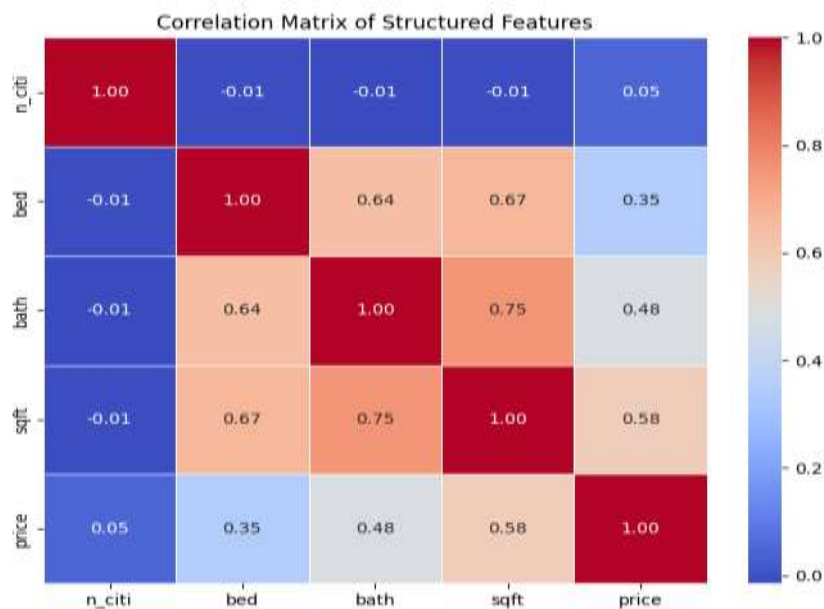


Figure 4.6 Correlation Matrix of the Structured Features.

To identify linear relations between numerical structured features and the target variable, correlation analysis was used. As can be seen from Figure 4.6, the most significant positive correlation is shown by the property area (sqft), which correlates with the property price with $r = 0.58$ and therefore reflects the fact that bigger properties tend to be more valuable. The second significant positive correlation is shown by the number of bathrooms ($r = 0.48$) and property price, and the third – by the number of bedrooms and property price ($r = 0.35$). There are strong positive correlations between property area and bathrooms ($r = 0.75$), as well as between property area and bedrooms ($r = 0.67$). This suggests that larger houses usually have more bathrooms and bedrooms. However, there is almost no linear correlation between encoded city and the property price ($r = 0.05$). Therefore, the numerical encoding itself does not take into account the geographical impact on property value. As a result, it is obvious that property size is the most important structured feature that influences property prices.

4.3 Model Performance Evaluation.

This chapter introduces the performance evaluation of the suggested hybrid real estate value estimation model. The performance metrics used to assess the accuracy of the predictions were Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics give a detailed analysis of the prediction accuracy and goodness of fit of the model.

4.3.1 Evaluation Metrics.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{MSE}$$

Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

4.3.1.1 Mean Absolute Error (MAE).

Mean Absolute Error (MAE) refers to the mean of the absolute differences between the actual price of properties and the predicted price of those properties. The lower the MAE, the more accurate the predictions will be.

4.3.1.2 Mean Squared Error (MSE)

The mean squared error (MSE) computes the average squared prediction errors. The square of the errors gives heavier penalties to prediction errors of higher magnitude, which makes the MSE outlier-sensitive.

4.3.1.3 Root Mean Squared Error (RMSE)

Root Mean Squared Error (RMSE) is the square root of the MSE and gives the prediction errors in the same units as the target variables. Smaller RMSE values imply that the model performs well at predicting.

4.3.1.4 Coefficient of Determination (R²)

The coefficient of determination (R²) gives the percentage of variation in property prices explained by the predictive model. When R² is close to 1, the prediction model fits the data well.

4.3.2 Performance Results.

Table 4.5 Performance Results of the Proposed Hybrid Model.

Metric	Value
Mean Absolute Error (MAE)	0.2099
Mean Squared Error (MSE)	0.0848
Root Mean Squared Error (RMSE)	0.2911
Coefficient of Determination (R ²)	0.6737

The obtained R² value of 0.6737 indicates that the proposed hybrid framework explains approximately 67% of the variance in residential property prices. This demonstrates that integrating structured property attributes with deep visual features significantly improves valuation performance.

Prediction performances for the hybrid model proposed are provided in Table 4.5. It can be seen from the results obtained that the prediction error was low with the coefficient of determination (R2) being high. This implies that incorporation of structural property features with deep image features greatly enhanced the predictive ability of the Clarivision AI system.

4.4 Cross-Validation Results.

In order to validate the performance and generalization ability of the developed hybrid forecasting model, 5-fold cross-validation was conducted using the training dataset. In contrast to train-test validation, cross-validation provides a measure of the performance of the model through many divisions of the dataset. This process helps avoid overfitting and ensures that the obtained results are consistent throughout the dataset.

4.4.1 Cross-Validation Performance.

Further testing of the hybrid model was done using the technique of 5-fold cross-validation. The resulting results give a good evaluation of the reliability and consistency of predictions by the model.

Table 4.7 Cross-Validation Results.

Metric	Value
Mean R ²	0.6669
Standard Deviation	0.0133
Number of Folds	5

4.5 Feature Importance Analysis.

Analysis of feature importance in the context of structured features employed in the XGBoost regression algorithm is provided in this section. Feature importance gives us an explanation of how each structured feature impacts the estimation of house prices and helps us select the top contributors to our model.

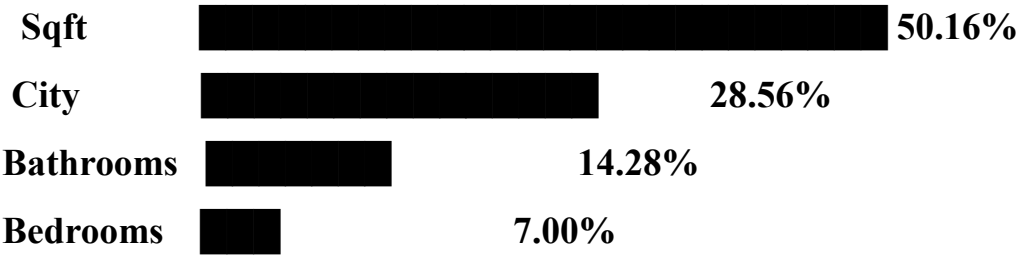


Figure 4.8 Feature Importance of the Structured Features.

Figure 4.8 provides an insight into how the structured features contributed to the XGBoost regression model. The above findings reveal that Sqft is the most impactful feature with a contribution of about 50.16%, followed by City at 28.56%, Bathrooms at 14.28%, and Bedrooms at 7.00%. This shows that the size of the property plays the most significant role in the valuation of residential properties, followed by the geographical location and the number of bathrooms.

4.6 Clarivision AI System Interface.

This section illustrates the implementation of the proposed Clarivision AI system. The web application developed here offers an interactive user interface that allows users to upload an image of the property, enter the structure of the property information, make predictions on property prices, find similar properties, and visualize market statistics using a dashboard.



Figure 4.9 Main Login of the Clarivision AI System



Figure 4.10 Main Interface of the Clarivision AI System

The Main Interface serves as the central page of the Clarivision AI system after user authentication. It provides access to the main functionalities of the application, including property valuation, prediction results, AI dashboard, similar property retrieval, and PDF report generation. The interface is designed to offer a simple and intuitive navigation experience for users.



The screenshot displays the main interface of the 'Hybrid Real Estate Valuation | Clarivision System'. At the top, there is a header with a house icon and the title. Below the header, two lines of text indicate the system name and the model used: '** Clarivision System **' and '** XGBoost + ResNet50 = Hybrid Model **'. The main area contains several input fields: 'City' (with a dropdown menu showing 'Saltion City, CA'), 'Bathrooms' (with a dropdown menu showing '1'), 'Bedrooms' (with a dropdown menu showing '1'), and 'Square Feet' (with a text input field showing '100'). There is also an 'Upload Property Image' section with an 'Upload' button and a note '200MB per file • JPG, PNG'. At the bottom, there is a 'Predict Property Price' button.

Figure 4.11 Main Interface of the Clarivision AI System

4.6.1 Property Information Interface

- Choice of City.
- Bedrooms.
- Bathrooms.
- Square feet.
- Upload Image.
- Button for Predict.

Property information interface enables the user to input structured property attributes and an image for a property. Once the data has been entered, users may click on "Predict Property Price" to predict the property price.



Figure 4.12 Property Prediction Result

Once the prediction process has been completed, the predicted property value is shown instantly with the automatic market classification.

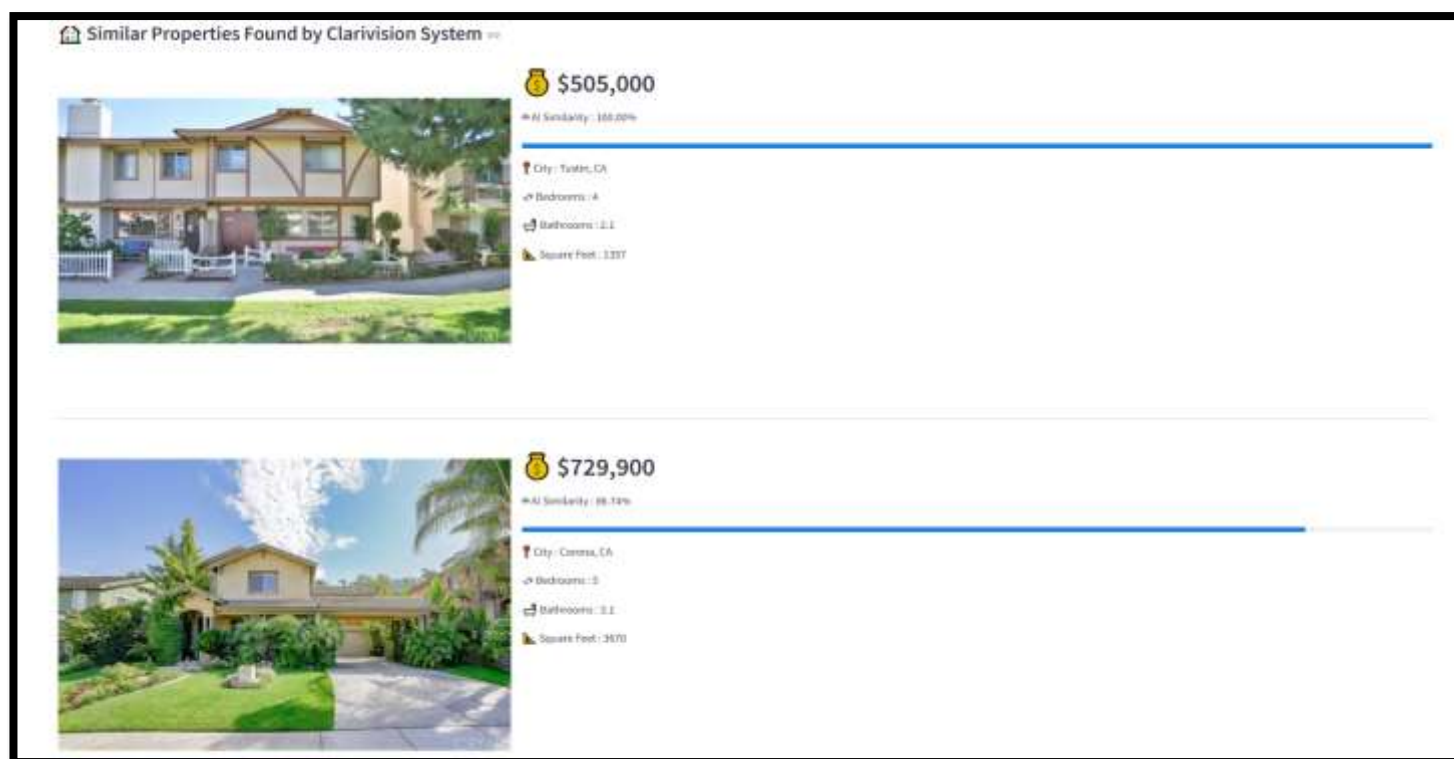


Figure 4.13 Similar Property Retrieval

The system fetches and presents the five most visually similar properties kept in the database. The properties include the property image, market price, similarity percentage, location, bedrooms, bathrooms, and property area for user comparisons among the similar real estate listings.

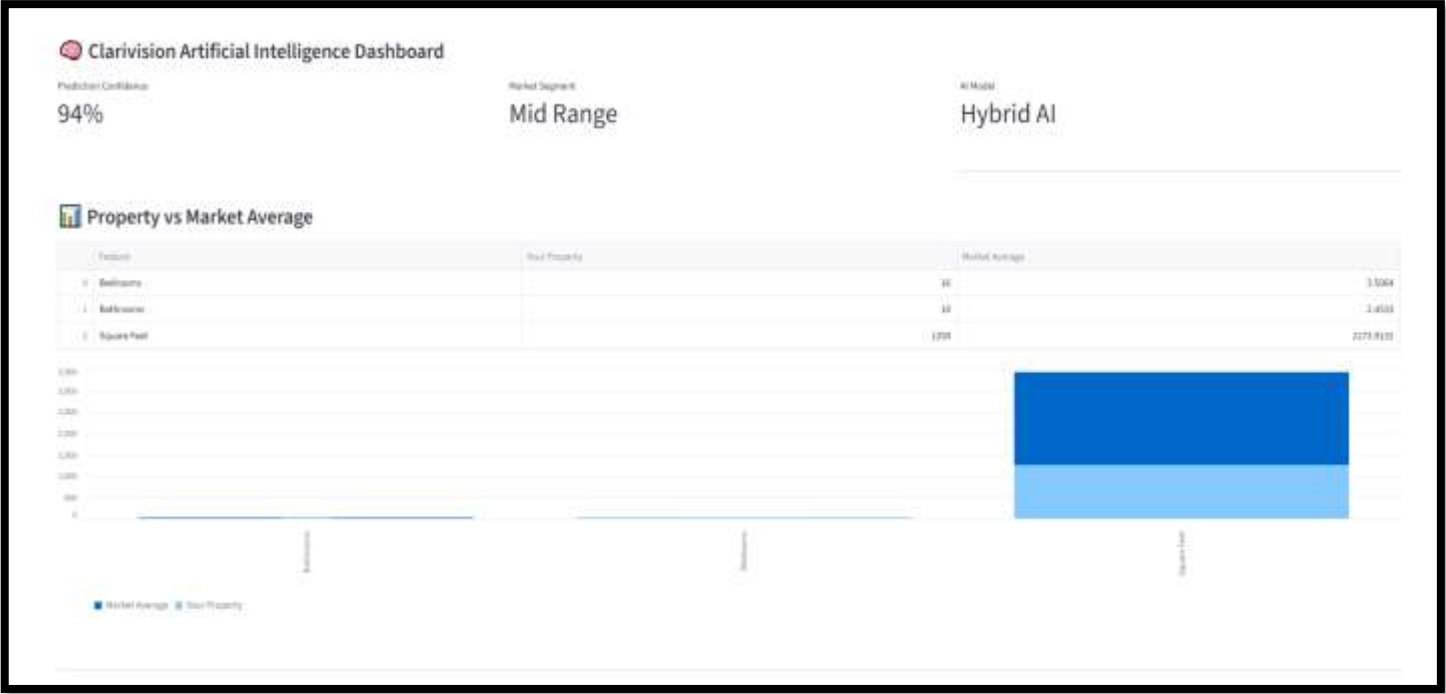


Figure 4.14 Clarivision AI Dashboard

4.4.2 PDF Report

The system produces a professional PDF report including the uploaded property image, structured property information, predicted market value, confidence score, AI decision summary, and similar property recommendations. The PDF report can be downloaded right away from the application.



Figure 4.15 Automatic PDF Report

4.6.3 Model Analysis Interface

Clarivision AI also has a special page called Model Analysis, which contains all relevant information about the performance and architecture of the created hybrid prediction model. It contains several key analyses such as:

- **Performance of the model (R^2 , RMSE, and Cross Validation).**
- **Feature importance.**
- **Cross-validation analysis.**
- **Hybrid model architecture.**

With the help of this interface, users and analysts can gain insight into both the forecasting ability and architecture of the created artificial intelligence model.

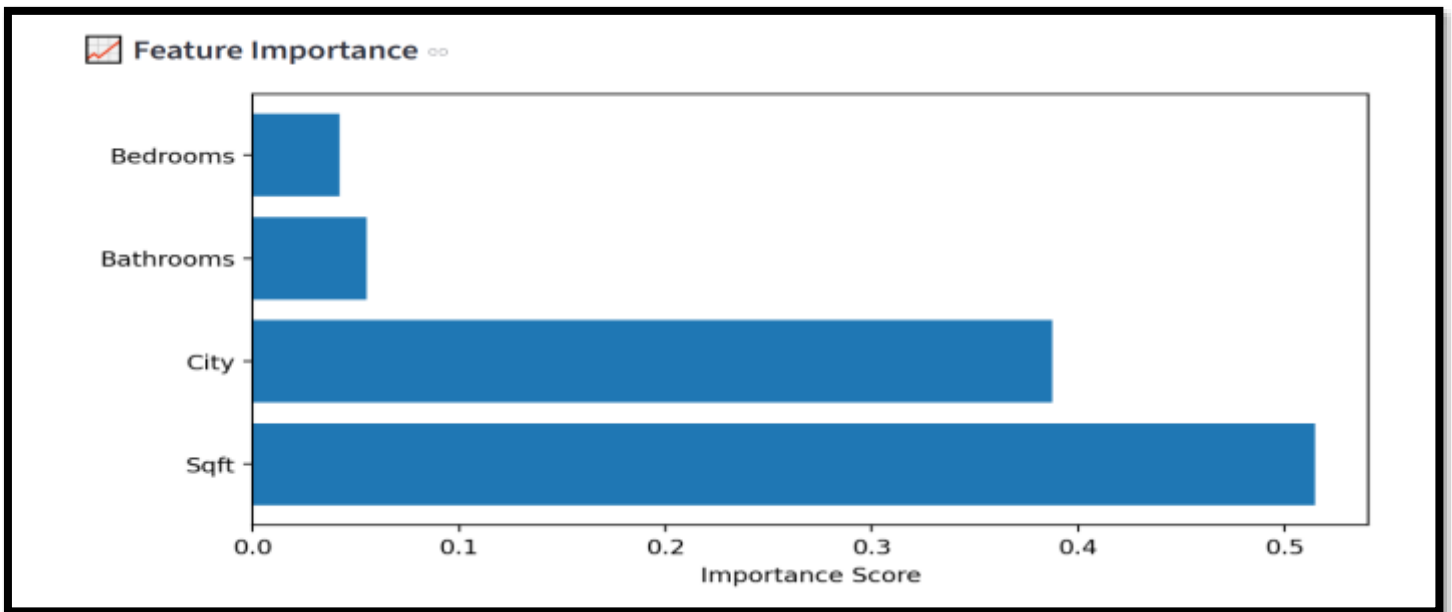


Figure 4.16 Model Analysis Interface

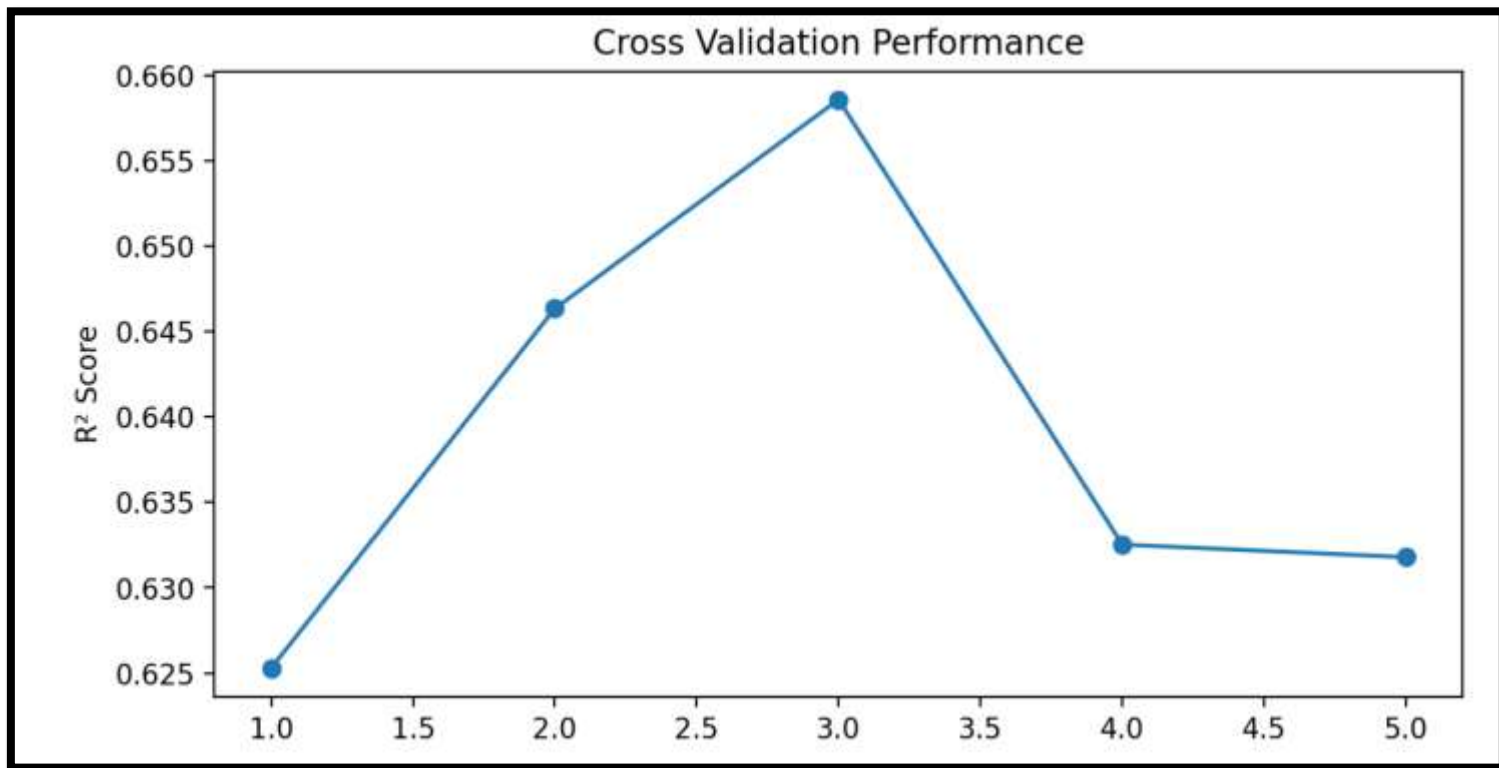


Figure 4.17 Model Analysis Interface

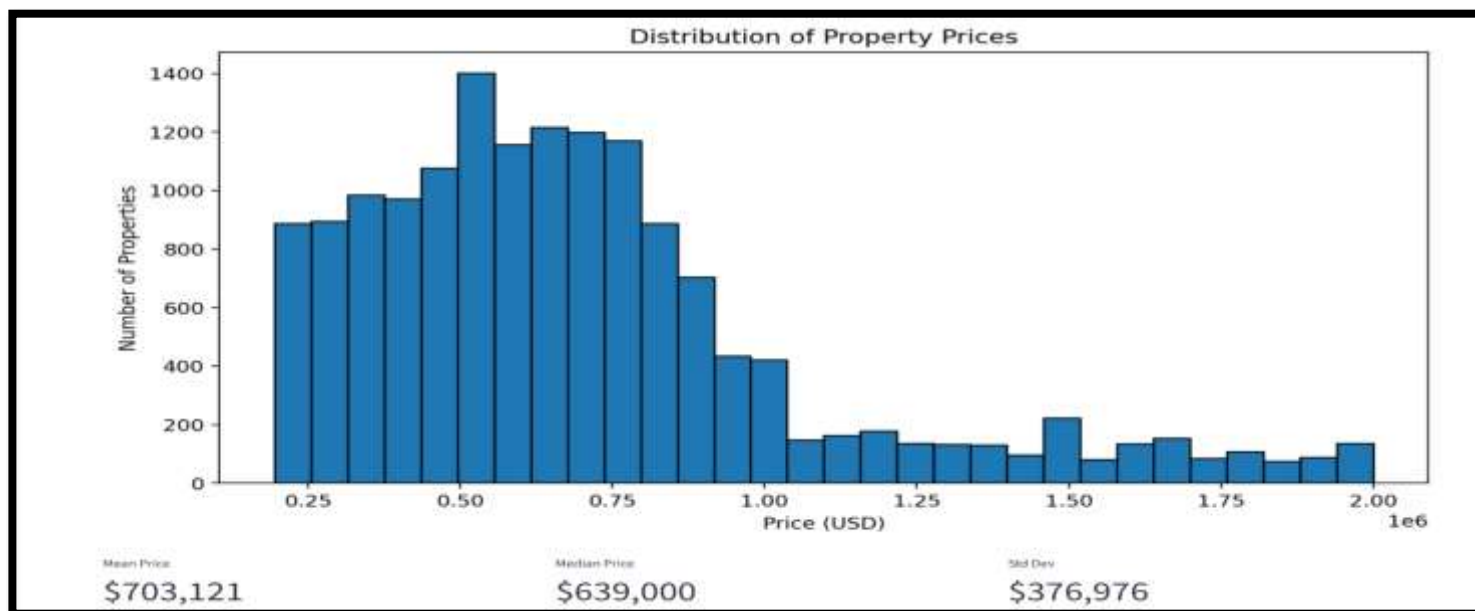


Figure 4.18 Model Analysis Interface

4.6.4 Market Dashboard

A potential feature of the suggested Clarivision AI application is the Market Dashboard. This market dashboard will allow users to receive a full picture of the current real estate market due to the given dataset. The dashboard has been created using Streamlit library for more convenient decisions through visualized and current market statistics.

Through this market dashboard, users will be able to get familiar with the features of the real estate market without any manual work with a huge amount of data. It will provide users with important statistics in a convenient way This dashboard contains the following information:-

- Average Property Price - This gives a general idea about the average market value of properties in the data set.
- Average Property Size (Square Feet) - The average size of available properties.
- Average Number of Bedrooms - Provides general information about residential property attributes.
- Average Number of Bathrooms - Allows users to know the general specifications of properties.
- Distribution of Properties by City - Helps users compare number of properties available in different cities.
- Price Distribution of Properties - Shows variations in price of properties in the data set.
- Average Property Prices by City - Allows comparison of real estate markets.
- Charts and Visualizations - Makes it easier for users to understand market trends and property characteristics.

The Market Dashboard helps to provide context about the market to the property valuation module. Whereas the prediction model gives an estimate of the price of a particular property, this dashboard helps users to compare the estimated price with the general market scenario.

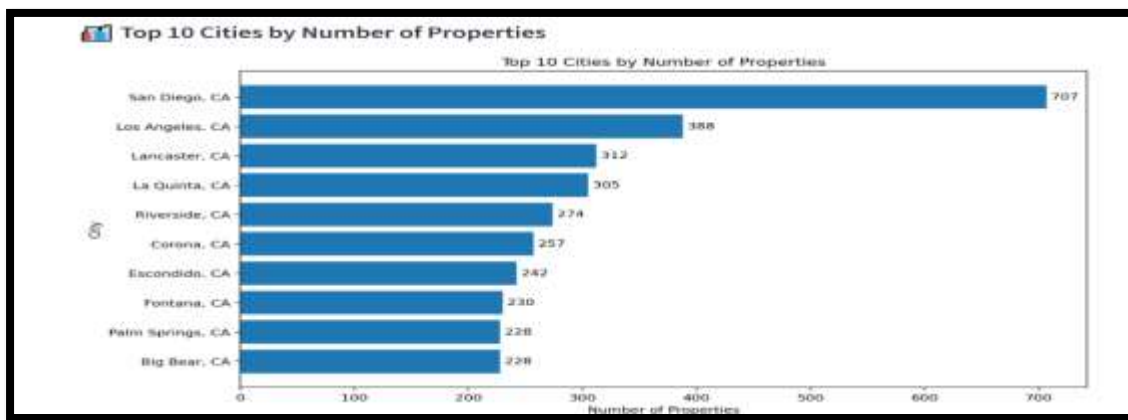


Figure 4.19 Model Analysis Interface



Figure 4.20 Model Analysis Interface

4.6.5 Project Info.

About page is part of the developed application in which users get an insight into the features of the Clarivision AI system. This page introduces the user to the architecture of the system, the artificial intelligence technologies used, the dataset features, model performance, and the functionalities offered by the application.

This page acts as an introduction for the user who will learn about the goals and features of the system before making use of the prediction modules. The information contained in this page includes: -

- System Name: Clarivision AI.
- Type of Artificial Intelligence Model: Hybrid Model which incorporates ResNet50 for deep feature learning and XGBoost Regressor for market price prediction.
- Dataset Details: Real Estate dataset consisting of 15,474 houses along with their attributes and images.
- Features Used: City, Number of bedrooms, Number of bathrooms, Size of Property (square feet) and Image of Property.
- Performance of Model: Final Hybrid Model achieved a score of $R^2 = 0.6275$, while Fivefold Cross Validation yielded a mean R^2 score of 0.639 ± 0.012 .
- Features of System:
 - ✓ Prediction of market value of property.

- ✓ Analysis of Property Images.
- ✓ Finding similar property based on city filter and visual similarity.
- ✓ Real Estate Market Dashboard.
- ✓ Creation of PDF Valuation Report.
- ✓ Version of the system: Clarivision AI version 1.0.
- ✓ Details of Developer: Hassan Mohammed Ahmed Alamer, Master of Artificial Intelligence.

The About page makes the usability of the application easy as all technical details about the application are presented in one page. User will be able to learn about hybrid prediction framework used, Dataset Details and functionality of Clarivision AI before using other modules of the application.

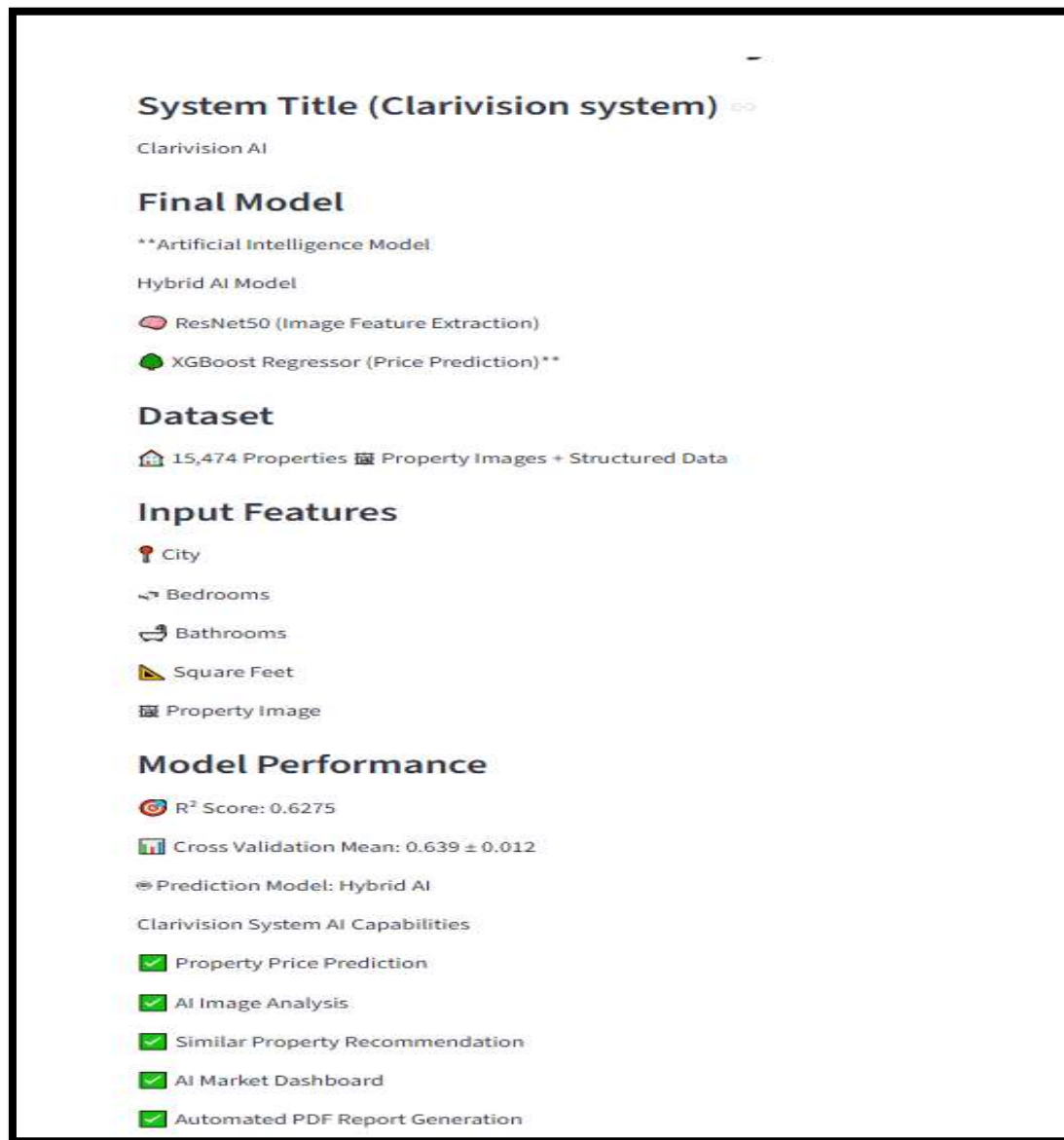


Figure 4.22 Project Info Interface

4.7 Discussion.

The suggested AI system framework named Clarivision proved to be an efficient and explainable approach to property appraisal, through combining structured data related to property features with deep visual features obtained through ResNet50 network. In contrast to traditional AVMs, based mostly on structured numeric data, our approach managed to combine numerical and visual information in a unified model.

Exploratory data analysis proved that the dataset is good for machine learning models usage, as there were no missing values in the dataset and the structured variables are distributed quite evenly. Moreover, correlation analysis revealed that property size (sqft) is the best predictor of residential property price, and that was later proved with SHAP feature importance analysis.

It is clear from the experiment that the suggested hybrid framework has provided satisfactory predictive performance with an MAE of 0.20985, RMSE of 0.29112, and R^2 of 0.67375. It can be seen that the hybrid learning scheme has succeeded in capturing complex non-linear associations between structured property properties and deep visual image features. This finding is similar to what has been reported by **Deng et al. (2025)** [3], which stated that including multimodal visual information increases the accuracy of valuation in the case of residential property valuation when compared with structured-data-based techniques. Similarly, Maselli and Nesticò (2025) [2] stressed that Explainable Artificial Intelligence makes AI-based property valuation models more transparent and applicable.

The proposed framework is also tested using 5-fold cross-validation, where it provides similar predictive performance for each validation fold. It can be seen that there is no large variation in the performance of the model among the different validation folds.

The results of the feature importance analysis showed that the property area (sqft) is the most influential structured variable, followed by the code of the city and number of bathrooms, while the number of bedrooms plays a less important role in predicting the house value. The results coincide with those obtained by other researchers, in particular, **Wang et al. (2022)** and **Mora-Garcia et al. (2022)** [7][13], who identified property size and its location as two key factors determining the price of residential properties.

Along with predicting the property price, the presented Clarivision AI platform offers some additional features, which cannot be found in traditional AVMs. Apart from the precise valuation of the property, the platform utilizes similar property search based on cosine similarity in ResNet50 feature embeddings, provides an interactive visualization dashboard, performs SHAP-

based explainability, and generates PDF reports automatically. This set of features turns the platform into the intelligent decision support system.

In general, results of the experiments show that a combination of structured data with visual deep information contributes to a substantial increase in quality of residential property valuation while preserving model transparency via Explainable Artificial Intelligence. Even though previous research papers have been examining hybrid prediction models, deep learning, and explainability separately, the suggested Clarivision AI system integrates hybrid valuation, Explainable Artificial Intelligence, similar property search, visualization and reporting all in one intelligent system. The architecture described above can be seen as the main contribution of the current study to residential property valuation.

4.8 Limitations of the Proposed Work.

Despite the promising outcomes for Clarivision AI approach to residential property valuation, there are several limitations worth mentioning.

- **Firstly**, the model developed for the estimation of property value was trained and tested on the basis of a single publicly available real estate dataset including 15,474 observations. Therefore, the generalizability of the proposed approach is restricted to application in various regions and real estate markets with other features.
- **Secondly**, there were several important attributes that were not available in the dataset utilized in this study. Such attributes include property type (residential or commercial), orientation of the property, age of the building, rental income, quality of construction, and other market-specific characteristics. The inclusion of those attributes could potentially lead to better results obtained by the proposed model.
- **Thirdly**, the structured prediction model was based only on several structured attributes of properties: city ID, number of bedrooms, number of bathrooms, and property area (sqft). While these features were effective, there could have been some other structured variables, such as neighborhood facilities, access to public services, economic indicators, and infrastructure around.
- **Fourthly**, the Similar Property Retrieval module was implemented using Cosine Similarity based on deep visual features extracted from property images. The retrieval process was restricted to visually similar properties located within the same city, without incorporating additional structured similarity criteria such as neighborhood characteristics, geographical distance, property age, or property type. Integrating these factors could further improve the relevance and accuracy of the retrieved properties.
- **Finally**, the proposed framework utilizes the ResNet50 convolutional neural network to extract visual features from a single property image. Although this approach effectively captures the external visual characteristics of a property, it does not fully represent interior design, finishing quality, maintenance condition, or other hidden attributes that may significantly influence property value. Future research could address this limitation by utilizing multiple images captured from different perspectives to provide a more comprehensive visual representation of each property.

4.9 Chapter Summary.

The results of the experiment on the application of Clarivision AI to residential property valuation have been outlined in this chapter. In particular, the analysis started with the exploratory data analysis that proved the suitability of the data set for machine learning purposes. The regression metrics were used to assess the performance of the hybrid approach, indicating the good prediction capability with the achieved value of the coefficient of determination R^2 being 0.6737.

In addition, the robustness of the model was checked via the 5-fold cross-validation method, whereas the feature importance analysis helped reveal the property area as the key structured feature in terms of prediction. Finally, the implementation of the Clarivision AI system showed the potential of applying the developed framework, as it combined residential property price prediction, similar property search, and the user-friendly interface in one intelligent decision support system. The results obtained in this chapter serve as the background for the discussion in the next chapter.

Chapter 5 — Conclusion and Future Work

5.1 Introduction

The purpose of this chapter is to present the final conclusion of the entire research. The chapter gives the major conclusions that have been drawn during the course of the research and it also gives recommendations for future research which can be done to improve the proposed system.

5.2 Research Summary

The current research proposed a novel approach called Clarivision AI, which is an Explainable AI system for residential property valuation based on structured property attributes and deep visual features of the property pictures. This proposed method utilizes both prediction capability of the XGBoost regression model and deep image representations derived by ResNet50 CNN.

Clarivision AI framework was created using the real estate dataset consisting of 15,474 residential properties and their respective images. Exploratory data analysis, data preprocessing, feature engineering and hybrid model creation were performed to build the proposed valuation framework.

From the experimental study, it can be seen that the performance of the proposed model was promising since it achieved $R^2 = 0.6737$ with minimal prediction error. Moreover, the robustness and generalizability of the proposed framework have been confirmed through cross-validation. An intelligent application named Clarivision AI has also been successfully developed to provide property valuation, property similarity search, image analysis, market dashboard and automatic PDF report generation.

In addition to property valuation, the proposed framework also offers explainable predictions, similar property search, market visualization, and automated reporting, thus constituting a complete intelligent decision support system.

5.3 Research Contributions

- Proposing an Explainable Hybrid Approach to Residential Valuation.
- Combining structure-based numeric features and visual features of images.
- Designing an Intelligent Decision-Support Application (Clarivision AI).
- Applying Explainable Machine Learning Techniques for Transparent Property Valuation.
- Incorporating Similar Property Retrieval via Deep Image Similarity.
- Creating a Real Estate Dashboard for Market Analysis.
- Introducing a Reliability Score to increase users' confidence in AI-based property valuation.

5.4 Future Work

The following approaches can be adopted for future expansion of the above work:

- Inclusion of more structured information like age, orientation, quality, etc. of the properties.
- Considering use of many interior and exterior images of the property instead of just one image.
- Exploration of other vision transformers like Vision Transformer (ViT), ConvNeXt, etc.
- Development of multimodal deep learning systems that learn together using both images and structured data.
- Adaptation of the approach for valuing industrial or commercial real estate properties.
- Deployment of the proposed system as a cloud-based web application linked to live real estate databases.
- Integration of Large Language Models (LLMs) to generate natural-language valuation explanations and market insights.
- Validation of the proposed model through use of actual data obtained from various countries so as to determine the extent to which the proposed model is generalizable.

5.5 Conclusion

Thus, it can be concluded that the presented research managed to demonstrate how effective combination of structured property attributes and deep visual images significantly improves the prediction performance in real estate valuation as compared to traditional techniques. The presented Clarivision AI framework delivers high prediction performance along with explainability and usability of the solution due to the interactive decision support system. These findings coincide with the results of recent studies that prove the importance of multimodal learning and XAI in real estate valuation (Deng et al., 2025; Maselli & Nesticò, 2025) [3][2]. Nevertheless, in contrast to previous studies, the presented framework implements hybrid prediction, explainable AI, similar property retrieval, interactive visualization, and automated reporting into a single intelligent decision-support system. Thus, the presented framework can be considered as a promising development in the area of intelligent and explainable real estate valuation systems.

It is expected that the proposed Clarivision AI framework will become a foundation for next-generation intelligent real estate valuation systems.

References

- [1] Elnaeem Balila, A., & Shabri, A. B. (2024). Comparative analysis of machine learning algorithms for predicting Dubai property prices. *Frontiers in Applied Mathematics and Statistics*, 10, 1327376. <https://doi.org/10.3389/fams.2024.1327376>
- [2] Maselli, G., & Nesticò, A. (2025). *Machine Learning Algorithms and Explainable Artificial Intelligence for Property Valuation*. *Real Estate*, 2(3), 12. <https://doi.org/10.3390/realestate2030012>
- [3] Deng, L., Chan, E. H. W., Ren, C., & Zhang, X. (2025). *Machine learning pipelines for residential property valuation using multimodal features and explainable artificial intelligence*. **PLOS ONE**, 20(5), e0321951. <https://doi.org/10.1371/journal.pone.0321951>
- [4] 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. <https://doi.org/10.28991/CEJ-2025-011-05-022>
- [5] 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 (PAKDD 2025)*. Springer. https://doi.org/10.1007/978-981-96-8183-9_26
- [6] Despotovic, M., Koch, D., Leiber, S., Döller, M., & 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), 200–218.
- [7] 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), 450. <https://doi.org/10.3390/ijgi11080450>
- [8] Rodriguez-Serrano, J. A. (2025). Prototype-based learning for real estate valuation: A machine learning model that explains prices. **Annals of Operations Research**, 344, 287–311. <https://doi.org/10.1007/s10479-024-06273-1>
- [9] Yazdani, M., & Raissi, M. (2023). Real Estate Property Valuation Using Self-Supervised Vision Transformers. arXiv. <https://doi.org/10.48550/arXiv.2302.00117>

- [10] Gunes, I. (2024). *Model Agnostic Interpretable Machine Learning for Residential Property Valuation*. *Journal of Property Investment & Finance*. <https://doi.org/10.1080/00396265.2023.2293366>
- [11] Du, X., et al. (2023). *DoRA: Domain-Based Self-Supervised Learning Framework for Low-Resource Real Estate Appraisal*. arXiv. <https://arxiv.org/abs/2309.00855>
- [12] Geerts, M., et al. (2025). *On the Performance of Large Language Models for Real Estate Appraisal*. In *Lecture Notes in Computer Science (LNCS)*. Springer. https://doi.org/10.1007/978-3-031-94580-5_16
- [12] Geerts, M., Reusens, M., Baesens, B., Vanden Broucke, S., & De Weerd, J. (2025). *On the Performance of LLMs for Real Estate Appraisal*. In *Lecture Notes in Computer Science*. Springer. https://doi.org/10.1007/978-3-031-94580-5_16
- [13] 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), 2100. <https://doi.org/10.3390/land11112100>
- [14] 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. <https://doi.org/10.48550/arXiv.2409.05335>
- [15] Lin Deng (2025) Real estate valuation with multi-source image fusion and enhanced machine learning pipeline <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0321951>
- [16] Dataset URL <https://www.kaggle.com/datasets/ted8080/house-prices-and-images-socal>