AI Real Estate True Price Prediction Using on-Ground Video

Nithin G.[1]; Pramatha K. P.[2]; Naveen Kumar C.[3]; Pawanraj S. P.[4]

[1,2,3]Department of Computer Science and Engineering, East West Institute of Technology, Bengaluru
[4]Faculty of Computer Science and Engineering, East West Institute of Technology, Bengaluru

Publication Date: 2026/09/08

Abstract: Accurate real estate and land valuation is a challenging task due to the influence of multiple spatial, environmental, physical, infrastructural, and socioeconomic factors. Conventional property valuation methods often depend on manual inspection, historical transaction data, and expert judgment, which can be time-consuming, subjective, and difficult to scale. This paper proposes Estate Vision AI**, an intelligent, video-first real estate and land analysis framework that integrates computer vision, machine learning, geospatial intelligence, and environmental feature analysis to provide automated property assessment and valuation. The proposed system accepts on-ground property or land videos as a primary input and extracts visual and contextual features related to land characteristics, built structures, accessibility, surrounding infrastructure, drainage conditions, vegetation, and other observable property attributes. These features are combined with location-based and market-related information to estimate property price, land value, and development potential. The framework additionally provides analysis of soil-related visual characteristics, water-flow and drainage indicators, flood-risk factors, transportation accessibility, nearby facilities, and utility availability. An automatic valuation mode enables prediction without requiring the user to provide a reference price, while a comparison mode supports comparative analysis between multiple properties. The system also incorporates location-based video retrieval to enhance contextual understanding of properties and surrounding areas. By combining visual, spatial, environmental, and market features within a unified framework, Estate Vision AI aims to reduce dependence on manual valuation, improve scalability and consistency, and provide an explainable decision-support tool for property buyers, sellers, investors, valuers, and planning authorities. The proposed framework demonstrates the potential of multimodal artificial intelligence for developing more efficient, data-driven, and geographically aware real estate valuation systems.**

Keywords: *Real Estate Valuation, Computer Vision, Machine Learning, Geospatial Intelligence, Property Price Prediction.*

How to Cite: Nithin G.; Pramatha K. P.; Naveen Kumar C.; Pawanraj S. P. (2026) AI Real Estate True Price Prediction Using on-Ground Video. *International Journal of Innovative Science and Research Technology*, 11(8), 3020-3026. https://doi.org/10.38124/ijisrt/26aug1211

I. INTRODUCTION

The real estate sector plays a significant role in economic development, investment planning, urban growth, and land resource management. Accurate estimation of property and land value is therefore essential for buyers, sellers, investors, financial institutions, government agencies, and urban planners. However, conventional property valuation methods primarily rely on manual site inspection, expert judgment, historical transaction records, and comparison with similar properties. These approaches can be time-consuming, subjective, difficult to scale, and may not effectively capture the continuously changing physical, environmental, infrastructural, and geographical characteristics of a property.

Recent advancements in Artificial Intelligence (AI), Machine Learning (ML), Computer Vision, and Geospatial Information Systems have created new opportunities for developing intelligent and automated property valuation systems. Existing machine learning-based valuation models mainly use structured information such as property size, number of rooms, location, historical prices, and economic indicators. Although these approaches can provide useful predictions, they often lack direct analysis of the actual physical condition of the property and its surrounding environment. Important factors such as visible land characteristics, accessibility, drainage conditions, nearby infrastructure, environmental conditions, and development potential may require manual inspection and expert interpretation .

To address these limitations, this research proposes Estate Vision AI, a video-first intelligent real estate and land analysis framework that combines computer vision, machine learning, geospatial intelligence, and location-based data analysis. Unlike conventional systems that depend mainly on manually entered structured data, the proposed framework uses on-ground video of a property or land as a primary source of information. The system extracts relevant visual features from video frames and combines them with

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

geographical, environmental, infrastructural, and market-related information to generate a comprehensive assessment.

The proposed system is designed to estimate property price and total land value while also providing additional insights related to land characteristics, visible soil conditions, water-flow and drainage indicators, potential flood-risk factors, transportation accessibility, nearby facilities, utility availability, and development potential. An Automatic Valuation Mode enables users to obtain an estimated property value without providing a reference price, whereas a Comparison Mode allows multiple properties to be evaluated and compared based on relevant features and predicted values. The framework can also incorporate location-based video retrieval to provide additional contextual information about a selected geographical area.

By integrating multimodal data from visual, spatial, environmental, and market sources, Estate Vision AI aims to move beyond traditional single-factor or structured-data-based property prediction approaches. The proposed framework seeks to provide a scalable, data-driven, and intelligent decision-support system that can assist property buyers, sellers, investors, real estate professionals, and planning authorities. This research contributes toward the development of next-generation real estate valuation systems in which on-ground visual information and geographical intelligence can be combined to support more comprehensive and context-aware property assessment.

II. LITERATURE REVIEW

The application of artificial intelligence and machine learning to real estate valuation has received increasing attention as researchers seek to improve the accuracy, efficiency, and scalability of traditional property appraisal methods. Conventional valuation approaches generally rely on expert judgment, comparable property transactions, hedonic pricing models, and statistical regression techniques. Although these methods remain widely used, they may have limitations in representing complex nonlinear relationships among property characteristics, geographical conditions, neighbourhood attributes, environmental factors, and market dynamics. Recent research has therefore explored machine learning, deep learning, computer vision, remote sensing, and spatial data analysis for automated property and land valuation..

A significant area of research focuses on machine learning-based automated valuation models (AVMs**)**. Jafaryetal. developed and compared machine learning and deep learning models for automated land valuation using a broad range of physical, geographical, socioeconomic, environmental, legal, and planning factors. Their findings demonstrated that incorporating diverse influencing factors can improve land valuation performance, with XG Boost outperforming several alternative models in their study. This work highlights the importance of integrating heterogeneous data sources rather than relying only on basic property attributes. However, the approach primarily depends on structured datasets and does not directly use on-ground visual or video information as an input for valuation

Spatial and geographical information has also emerged as an important component of property valuation. Lee et al. proposed a machine learning framework combining a convolutional neural network with traditional numerical and categorical property features to represent geographic variation in residential valuation. The model incorporated geographical layers and neighbourhood information and achieved a lower mean absolute percentage error than the baseline model in their Greater Sydney case study. This research demonstrates that spatial information can improve automated valuation by reducing the loss of geographical context. Nevertheless, the framework is based primarily on structured property and spatial data rather than continuous on-ground video analysis of the property environment.

Remote sensing has provided another important direction for property assessment. Research on property valuation using remote sensing data has investigated the potential of high-resolution imagery and other spatial data sources to identify physical property characteristics relevant to valuation and land taxation. Such approaches demonstrate the value of Earth observation data for large-scale and geographically distributed property assessment. However, satellite, aerial, or remote sensing imagery may not always capture detailed ground-level characteristics such as local drainage conditions, visible accessibility, surrounding infrastructure, and other features that may be observed through on-site video.

Recent literature has also emphasized the growing importance of multimodal learning in real estate appraisal. Reviews of machine learning and AI applications in real estate indicate a transition from models based mainly on structured transaction and property data toward approaches that combine multiple data types, including geographical, visual, textual, and other contextual information. At the same time, issues related to data quality, model interpretability, transparency, and practical deployment remain important research challenges.

Therefore, the proposed Estate Vision AI framework aims to address this research gap by developing a multimodal, video-first approach to property and land assessment. The system is intended to extract observable visual features from property videos and integrate them with geographical and market information to support automated price estimation. In addition to property valuation, the framework investigates contextual features such as accessibility, surrounding infrastructure, drainage and water-flow indicators, visible land characteristics, nearby facilities, utilities, flood-related environmental indicators, and development potential. The proposed approach also includes automatic valuation and property comparison capabilities. By integrating these components into a unified decision-support framework, Estate Vision AI seeks to extend existing automated valuation research toward a more comprehensive, context-aware, and ground-level analysis of real estate and land.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

III. METHODOLOGY

The proposed Estate Vision AI framework follows a multimodal machine learning methodology for automated real estate and land assessment. The central concept of the proposed system is to use an on-ground video of a property or land parcel as one of the primary sources of information and combine the extracted visual information with structured property data, geospatial information, environmental indicators, and market-related features.

The overall methodology consists of the following stages:

- Data acquisition
- Video preprocessing
- Frame selection and visual feature extraction
- Geospatial and contextual feature generation
- Environmental and land-condition analysis
- Multimodal feature fusion
- Machine learning-based property valuation
- Automatic valuation and comparison analysis
- Explain ability and result generation

The use of multiple data modalities is motivated by recent research showing that property valuation can benefit from combining visual, geographical, structured, and contextual information rather than relying only on conventional tabular attributes.

➢ *Data Acquisition*

The proposed system collects information from multiple sources. The primary input is an on-ground video of the property or land. Additional information may include property attributes, geographical coordinates, land area, property type, historical transaction or market data, and location-related contextual information. The input data can be represented as:

[D = {V, P, G, E, M}] (1)

Where:

- (V) = property or land video data
- (P) = structured property attributes
- (G) = geospatial and location information
- (E) = environmental and land-condition information
- (M) = market and historical valuation information

Structured attributes and geographical information are important components of automated valuation models, while visual information can provide additional characteristics that may not be explicitly represented in conventional datasets.

➢ *Video Preprocessing and Frame Extraction*

The uploaded property video is first processed to remove irrelevant, duplicated, blurred, or low-quality frames. Instead of processing every frame, representative frames are selected at predefined time intervals or through a key-frame selection mechanism.

Let the input video be represented as:

[V = {f_1, f_2, f_3, ..., f_n}] (2)

Where (f_i) represents an individual video frame.

A quality score can be assigned to each frame based on factors such as:

- Image sharpness
- Brightness
- Motion blur
- Scene uniqueness
- Object visibility

Only frames satisfying a predefined quality threshold are retained:

[F_{selected} = {f_i \mid Q(f_i) > T}] (3)

Where (Q(f_i)) is the quality score and (T) is the selection threshold.

This stage reduces computational complexity and ensures that the subsequent computer vision model receives informative visual representations.

➢ *Computer Vision-Based Feature Extraction*

The selected frames are analyzed using a computer vision model to identify relevant property and environmental features. Depending on the implementation, a convolutional neural network, vision transformer, or object detection model may be employed.

The visual analysis aims to identify observable features such as:

- Buildings and constructed structures
- Roads and access paths
- Vegetation and open land
- Water bodies or drainage channels
- Visible utility infrastructure
- Land surface characteristics
- Surrounding development
- Accessibility indicators

For each selected frame, the visual feature extractor generates an embedding:

[z_v = CNN(f_i)] (4)

Or, in the case of a transformer-based architecture:

[z_v = ViT(f_i)] (5)

The frame-level representations are then aggregated to obtain a video-level feature representation:

[Z_V = Aggregate(z_{v1}, z_{v2}, ..., z_{vn})] (6)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

Aggregation can be performed using mean pooling, maximum pooling, attention-based pooling, or another suitable temporal aggregation technique.

The use of visual representations in property valuation is supported by research showing that image-based and multimodal information can capture property and environmental characteristics that may not be fully represented by structured variables.

➢ *Geospatial and Location Intelligence Analysis*

The geographical location of the property is used to generate spatial and contextual features. Geographic coordinates can be used to calculate relationships between the target property and surrounding infrastructure.

Important geospatial features may include:

- Distance to major roads
- Distance to transportation facilities
- Distance to schools and hospitals
- Distance to commercial areas
- Nearby points of interest
- Neighbourhood development
- Terrain characteristics
- Land-use context

The geospatial feature vector can be represented as:

$$Z_G = [d_1, d_2, ..., d_k] \tag{7}$$

Where (d_k) represents a spatial attribute or distance-based feature.

Previous research has demonstrated that geographical and spatial information can significantly improve property valuation performance by preserving location-specific and neighbourhood characteristics.

➢ *Environmental and Land-Condition Analysis*

The proposed system further analyzes observable environmental conditions from visual and geospatial information. This component is designed as a decision-support indicator, rather than a replacement for laboratory soil testing, professional surveying, hydrological modelling, or official environmental assessment.

The analysis may generate indicators related to:

- Visible soil or land-surface characteristics
- Vegetation coverage
- Drainage patterns
- Visible water-flow paths
- Waterlogging indicators
- Proximity to water bodies
- Terrain slope
- Potential flood-related environmental indicators

A simplified environmental feature vector can be expressed as:

$$Z_E = [s, w, v, t, f] \tag{8}$$

Where:

- (s) = observable soil or surface characteristics
- (w) = water and drainage indicators
- (v) = vegetation-related features
- (t) = terrain-related features
- (f) = flood-related environmental indicators

These variables are intended to provide additional contextual information that can influence land usability and development potential.

➢ *Multimodal Feature Fusion*

The extracted features from different sources are combined to create a unified property representation.

The final feature vector is defined as:

$$Z = [Z_V \oplus Z_G \oplus Z_E \oplus Z_P \oplus Z_M] \tag{9}$$

Where:

- ($\oplus$) represents feature concatenation or another fusion operation.
- (Z_V) = video-derived features
- (Z_G) = geospatial features
- (Z_E) = environmental features
- (Z_P) = structured property features
- (Z_M) = market-related features

➢ *Property Price Prediction Model*

The fused feature vector is supplied to a supervised regression model to estimate the market value of the property.

The general prediction function is:

$$\hat{y} = F(Z)$$

Where:

- (Z) = multimodal feature representation
- (F) = trained machine learning regression model
- ($\hat{y}$) = predicted property price

Multiple algorithms should be evaluated during experimentation. Suitable candidate models include:

- Random Forest Regressor
- XGBoost Regressor
- LightGBM Regressor
- Artificial Neural Network
- Multimodal Deep Neural Network

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

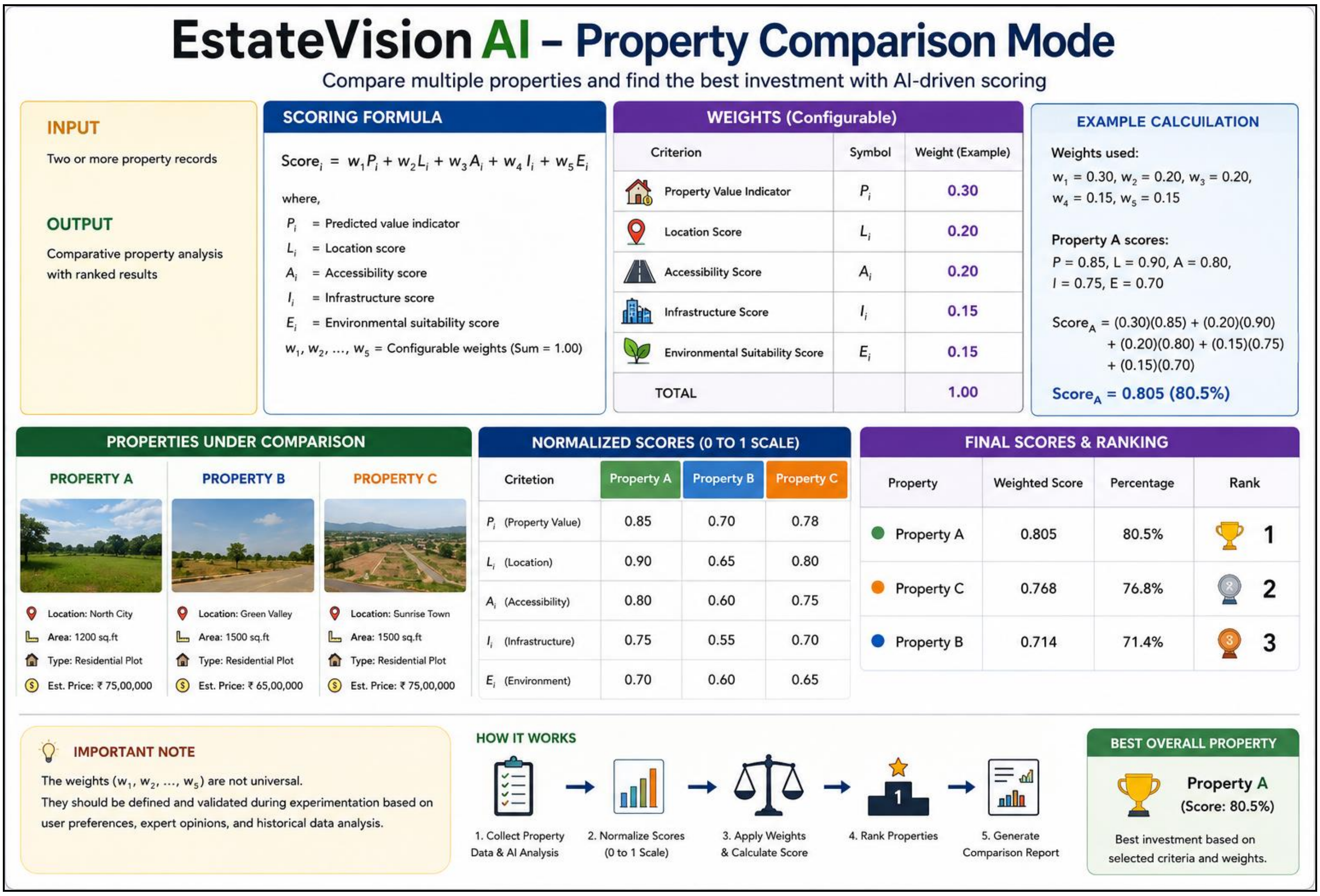

Fig 1 Estate Vision AI Algorithm Chart

➢ *System Architecture Diagram*

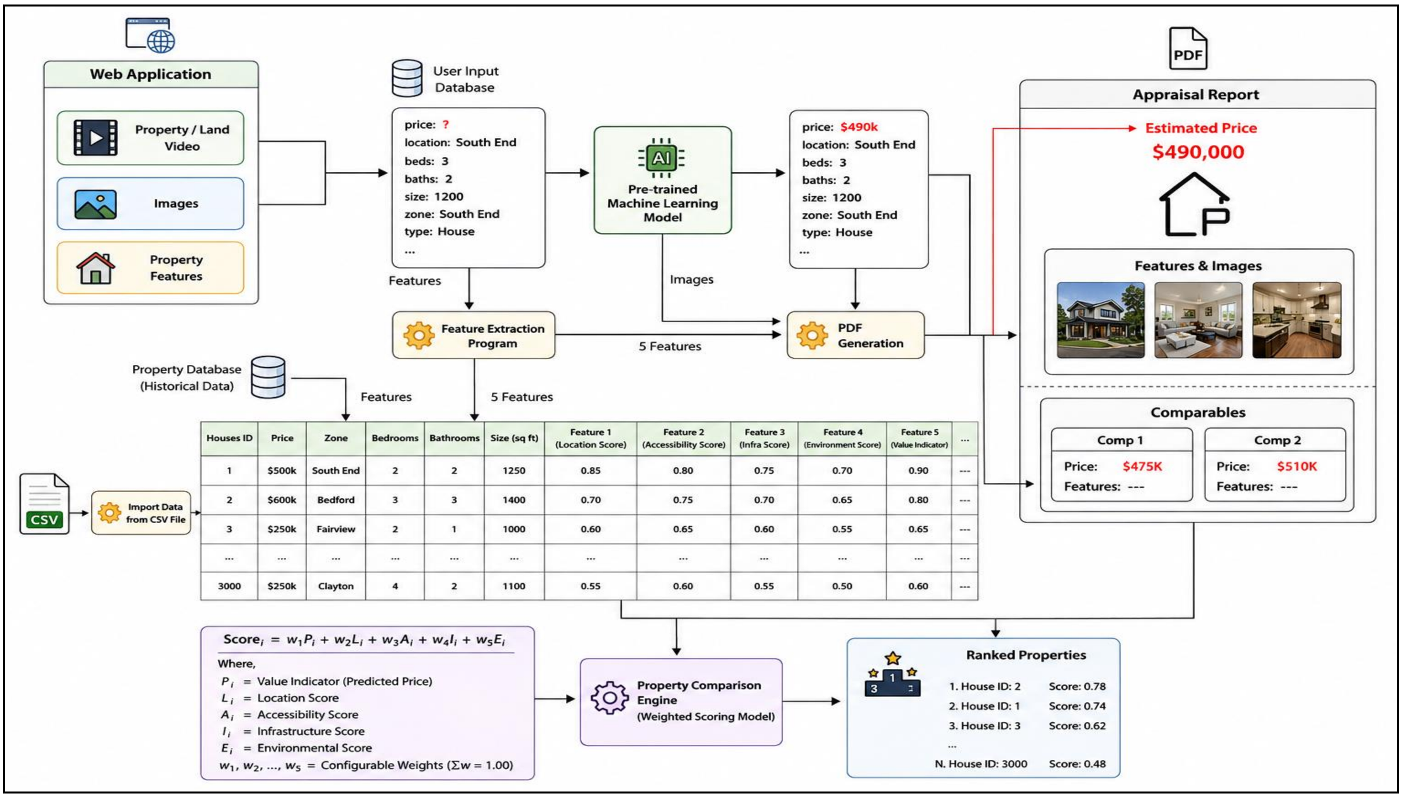

Fig 2 Estate Vison AI System Architecture

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

IV. RESULTS AND DISCUSSION

The dataset was divided into training, validation, and testing subsets. The training dataset was used to learn the relationship between property characteristics and observed market values, while the validation dataset was used for model selection and parameter tuning. The independent testing dataset was used to evaluate the final predictive performance.

The models were evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and coefficient of determination ((R^2)). Lower MAE, RMSE, and MAPE values indicate better prediction performance, whereas a higher (R^2) indicates a stronger ability to explain the variation in property prices.

Several machine learning algorithms were considered, including Random Forest, XGBoost, LightGBM, and neural-network-based models. The proposed multimodal model was subsequently compared with models trained using only structured property attributes. The experimental results should be reported in the following format:

Table 1 Comparative Model Performance

Model	MAE	RMSE	MAPE	(R^2)
Random Forest	[value]	[value]	[value]%	[value]
XGBoost	[value]	[value]	[value]%	[value]
LightGBM	[value]	[value]	[value]%	[value]
Neural Network	[value]	[value]	[value]%	[value]
EstateVision AI Multimodal Model	**[value]**	**[value]**	**[value]%**	**[value]**

An important objective of the study is to determine the contribution of different information sources to property valuation. An ablation experiment can therefore be performed by progressively adding different feature groups.

Table 2 Multimodal Feature Integration

Feature Configuration	MAE	RMSE	MAPE	(R^2)
Property/Market Features Only	[value]	[value]	[value]%	[value]
Property + Geospatial Features	[value]	[value]	[value]%	[value]
Property + Geospatial + Environmental Features	[value]	[value]	[value]%	[value]
Video + Property Features	[value]	[value]	[value]%	[value]
Full Multimodal Model	**[value]**	**[value]**	**[value]%**	**[value]**

The proposed Estate Vision AI framework differs from conventional property-price prediction systems by treating property valuation as a multimodal analysis problem rather than solely as a structured-data regression problem. The combination of video-derived visual information, geospatial attributes, environmental indicators, and conventional property and market variables provides a broader representation of the property.

The expected benefit of this architecture is improved contextual awareness. Visual information can provide direct observations of the property and its surroundings, while geospatial features describe the relationship between the property and nearby infrastructure. Structured market variables provide the economic context required for price estimation.

Consequently, the reliability of Estate Vision AI should be evaluated not only using overall prediction metrics but also through regional validation, error analysis, ablation studies, and comparison with appropriate baseline models. Explain ability techniques should additionally be incorporated to identify the features that contribute most strongly to individual predictions.

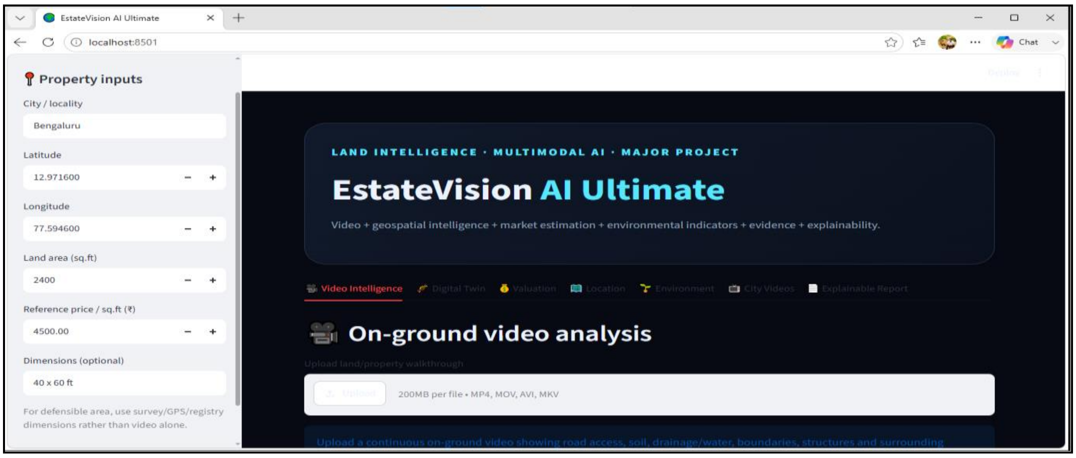

Fig 3 A Sample View of the Webpage Form to Upload Video

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26aug1211

V. CONCLUSIONS

This research presented Estate Vision AI, a proposed multimodal artificial intelligence framework for intelligent real estate and land valuation. The system addresses limitations of conventional property valuation approaches by integrating computer vision, machine learning, geospatial intelligence, environmental analysis, and structured property and market information within a unified framework. Unlike traditional approaches that primarily depend on manually entered property attributes or historical transaction data, the proposed system uses on-ground property or land video as an important source of visual information and combines the extracted features with geographical and contextual data.

The proposed methodology provides a comprehensive property assessment that includes estimated property price and land value, observable land and surface characteristics, water-flow and drainage indicators, flood-related environmental indicators, accessibility, nearby transportation and facilities, utilities, and development potential. The inclusion of Automatic Valuation Mode and Property Comparison Mode further extends the system from simple price prediction toward an intelligent decision-support platform.

The proposed architecture demonstrates the potential of multimodal learning for capturing different dimensions of property value. Visual information can provide ground-level observations, geospatial information can represent location and accessibility, environmental information can provide contextual land indicators, and structured market data can represent economic characteristics. The combination of these modalities is intended to produce a more comprehensive property representation than models based on a single data source.

Overall, Estate Vision AI provides a foundation for developing video-first, context-aware, and data-driven real estate intelligence systems. With further validation and integration of high-quality market, GIS, remote-sensing, and property datasets, the framework has the potential to support more efficient property assessment and informed decision-making for buyers, sellers, investors, valuers, and planning authorities.

REFERENCES

[1]. R. B. Abidoye and A. P. C. Chan, “Artificial neural network in property valuation: Application framework and research trend,” *Property Management*, vol. 35, no. 5, pp. 554–571, 2017, doi: 10.1108/PM-06-2016-0027.

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

[3]. H. Lee, H. Han, C. Pettit, Q. Gao, and V. Shi, “Machine learning approach to residential valuation: A convolutional neural network model for geographic variation,” *The Annals of Regional Science*, vol. 72, pp. 579–599, 2024, doi: 10.1007/s00168-023-01212-7.

[4]. M. Koeva, O. Gasuku, M. Lengoiboni, K. Asiama, R. M. Bennett, J. Potel, and J. Zevenbergen, “Remote sensing for property valuation: A data source comparison in support of fair land taxation in Rwanda,” *Remote Sensing*, vol. 13, no. 18, Art. no. 3563, 2021, doi: 10.3390/rs13183563.

[5]. A. J. Bency, S. Rallapalli, R. K. Ganti, M. Srivatsa, and B. S. Manjunath, “Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery,” in *Proc. IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2017, pp. 320–329, doi: 10.1109/WACV.2017.42.

[6]. P. Jafary, D. Shojaei, A. Rajabifard, and T. Ngo, “Automating property valuation at the macro scale of suburban level: A multi-step method based on spatial imputation techniques, machine learning and deep learning,” *Habitat International*, vol. 148, Art. no. 103075, 2024, doi: 10.1016/j.habitatint.2024.103075.

[7]. R. B. Abidoye and A. P. C. Chan, “Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network,” *Pacific Rim Property Research Journal*, 2018, doi: 10.1080/14445921.2018.1436306.

[8]. R. B. Abidoye and A. P. C. Chan, “Modelling property values in Nigeria using artificial neural network,” *Journal of Property Research*, 2017, doi: 10.1080/09599916.2017.1286366.

[9]. R. B. Abidoye and A. P. C. Chan, “Critical review of hedonic pricing model application in property price appraisal: A case of Nigeria,” *International Journal of Sustainable Built Environment*, 2017, doi: 10.1016/j.ijsbe.2017.02.007.

[10]. R. B. Abidoye and A. P. C. Chan, “Valuers’ receptiveness to the application of artificial intelligence in property valuation,” *Pacific Rim Property Research Journal*, 2017, doi: 10.1080/14445921.2017.1299453.

[11]. P. Jafary, D. Shojaei, A. Rajabifard, and T. Ngo, “Automated valuation models and machine learning approaches for property and land valuation,” related research on automated valuation and spatial data integration, 2024.

[12]. Y. Kang, X. Zhang, S. Yao, Y. Liu, and S. Huang, “GeoAI for real estate valuation: Integrating spatial and visual information for property price prediction,” related research direction in computer vision and geospatial analysis.