Volume (1) No (7), 2025
Journal of Global Research in Multidisciplinary Studies (JGRMS)
Review Paper/Research Paper
Available online at https://saanvipublications.com/journals/index.php/jgrms/index

Survey of Machine Learning and Deep Learning Approaches for Housing Price Forecasting: Datasets, Features, and Challenges

Dr Manish Saraswat
Associate Professor
(CSE) and Controller of Examinations
Faculty of Science and Technology
The ICFAI University
Himachal Pradesh
manish.saraswat@iuhimachal.edu.in

***Abstract*—Accurately predicting housing prices is a key part of how buyers, sellers, investors, and lawmakers make decisions about real estate. Although the Hedonic Pricing Model and linear regression have seen extensive usage in the field of statistics, both methods frequently fail to capture the complexities of multidimensional data and intricate nonlinear interactions. Recent years have seen the rise of deep learning (DL) and machine learning (ML) as potent methods for improving predictive performance in housing data models that are multimodal, dynamic, and high-dimensional. Outlining the methodology, applicable situations, benefits, and limits of several ML and DL approaches to housing price forecasting, this paper offers a complete assessment of the field. Advanced DL architectures like LSTM and multimodal neural networks are studied, along with key models like Random Forest, XG Boost, and Support Vector Machines. In addition to popular datasets, feature engineering methods, assessment criteria, and the current trend towards combining visual, textual, and geographic data, the review delves into these topics and more. In addition, it delves into important obstacles such dataset constraints, computing demands, model interpretability, and data heterogeneity. In the last section of the paper, the authors suggest areas for further study, highlighting the importance of standardized multimodal datasets, explainable AI (XAI), domain adaptation, and causal modelling in improving the realism and practicality of models.**

Keywords—Housing Price Forecasting, Machine Learning (ML), Deep Learning (DL), Real Estate Valuation, Feature Engineering, Multimodal Data Integration, Model Interpretability.

I. Introduction

Housing was one of the first and most consequential businesses to attract the attention of the general public. The constant flux within it has earned it the reputation of being the least transparent sector of their ecosystem. Along with the rise of computational methods, numerous research have recently applied state-of-the-art machine learning models to the real estate market in an effort to determine an appropriate price or to forecast price changes [1]. By anticipating the market, desirable neighborhoods, and building structures of the future, these computerized models assist homeowners, builders, and business owners in making more informed decisions and increasing their profits. Moreover, they lend a hand to city planners and administrators in organizing a city's environmental and locational facilities in a manner that would benefit the city's broader populace.

Effective housing price forecasting is a critical aspect of the real-estate markets and it affects the actions made by the buyers, sellers, investors, and even policymakers. Upon the possibility to forecast the value of the property, it is not only possible to make relevant pricing efforts but also elevates the rights of negotiations and gives information on the vast economy of the work frames [2][3]. Demand of powerful and confident forecasting methodologies has never been more eminent than now when populations are growing in urban areas, and real estate markets are becoming more intricate in nature. The classical statistical models are interpretable and computationally effective but may frequently be insufficient in presence of nonlinear and high-dimensional data.

The accuracy of any forecasting model also depends on the features and datasets that are chosen. Research uses a wide variety of datasets, some of which are city-specific (e.g., Melbourne, Munich) and others regional (e.g., property listings with a plethora of features including location, size, number of rooms, and distance to amenities) [4]. To further guarantee the model's accuracy and reliability, new feature engineering methods, such as transforming categorical variables into target statistics, have been implemented.

The goal of machine learning is to generate predictions from new data sets by training a model. The output of an input-data-oriented algorithm differs most noticeably from that of a traditional algorithm, which is more concerned with a sequence of different sets of instructions. While supervised learning is concerned only with building models from labelled data sets, unsupervised learning is devoted solely to unlabelled data sets [5]. Various machine learning algorithms include clustering, regression, classification, support vector machines (SVMs), neural networks (NNs), and deep learning (DL). With a feature extraction-based model, must be able to foresee a specific result. Location, size, house type, city, nation, tax regulations, economic cycle, population mobility, interest rate, and a plethora of other factors that could affect supply and demand are all considered when making predictions about homes. The global real estate market is very volatile due to the many known and unknown factors that affect it. Values of these properties might rise or fall at different rates over time due to market and non-market factors.

Machine learning Deep learning methods, based on the principles of machine learning, support the further enlargement of the boundaries of housing price modelling through the use of multidirectional data, such as images, texts, and time series [6]. Nonetheless, this makes deep learning models so dependent on the access to large and good quality data and extensive computing resources. Regardless of these difficulties, the harmonization of various data forms using deep learning provides new opportunities to obtain more subtle and context-sensitive prediction of prices.

A. *Structure of the Paper*

The structure of this paper is as follows: Section I introduces housing price forecasting rise of ML and DL methods. Section II provide overview of housing price forecasting. Section III covers ML and DL approaches. Section IV discusses datasets, feature selection, and challenges. Section V summarizes recent literature. Section VI concludes with key insights and future directions.

II. OVERVIEW OF HOUSING PRICE FORECASTING

Traditional statistical methods have slowly given way to more modern machine learning and deep learning techniques for predicting house prices. Linear regression and the Hedonic Pricing Model were the mainstays of home price projections in the early days of the industry. These traditional methods make price estimates by analyzing a home's structural characteristics, geographic location, and neighborhood attributes. The Hedonic model is a classical method that realizes the prediction by establishing a linear relationship between the various characteristic attributes of a house and the price [7][8]. These methods have good interpretability, and the model parameters are easy to understand, but there are significant shortcomings in dealing with complex nonlinear relations and high-dimensional characteristics, and it is difficult to effectively deal with various influencing factors and market fluctuations in the real estate market.

A. *Applicable Scenarios of Different Housing Price Forecasting Methods*

The traditional method is suitable for housing price forecasting tasks with small data amount and relatively simple feature relationship, such as the housing price assessment of some specific communities. Machine learning methods are suitable for larger data sets, especially when there are complex interaction relationships between features, and ML methods are better able to capture these relationships. DL methods are more suitable for the fusion of multi-source data, such as the simultaneous use of numerical, image and text features to predict housing prices [6]. Even while these methods excel at handling high-dimensional data and complicated non-linear interactions, they nevertheless necessitate a substantial amount of data and computer resources to train the model.

B. *Approaches in Forecasting*

The importance of accurate forecasts in making well-informed decisions has led to a dramatic expansion in their understanding of the subject as a result of recent years' worth of cutting-edge scientific inquiry. Instead, then approaching forecasting from a deterministic standpoint, statistical and stochastic methods offer significant benefits when it comes to constructing forecasts based on probability distributions. In addition, these techniques form the backbone of any forecasting model. Quantitative and qualitative approaches are the two main schools of thought when it comes to forecasting. In order to predict the future state of an activity, quantitative methods organize data from the past [9]. When non-quantifiable information influences the events under forecasting or when past data pertaining to the events in question is scarce or unavailable, qualitative methods are employed. In mortality forecasting, strategies such as expectation, extrapolation, and explanation are commonly employed. The use of conventional statistical approaches for predicting future outcomes has grown in popularity among demographers and actuaries. Extrapolative, Explanatory, judge mental, and Composite are the subcategories of forecasting methodologies recognised by the Society of Actuaries (SOA), a global organization for all things actuarial.

C. *Factors that Affect House Pricing*

Figure 1 shows the components that determine house pricing. Understanding these aspects is the first step in making house price predictions.

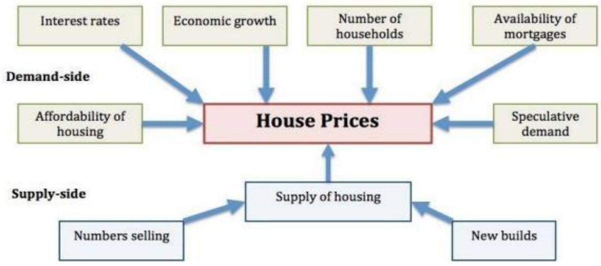

Fig. 1. Factors of housing price

- **Economic Growth:** A person's ability to pay affects the housing market. People will have more disposable income to spend on housing as a result of improved economic conditions, which will lead to more demand and higher prices. Actually, housing is generally said to be a luxury commodity with income elastic demand, meaning that as wages rise, a larger percentage of income goes towards housing costs.
- **Unemployment:** Unemployment is associated with economic growth. Fewer people will be able to purchase a home when the unemployment rate rises. However, nobody wants to go into the real estate market if they're scared of losing their jobs.
- **Interest Rates:** Monthly mortgage payments are impacted by interest rates. During an interest rate hike, homebuying becomes less appealing due to the higher cost of mortgage payments. Renting is preferable to buying due to high interest rates.
- **Consumer Confidence:** One of the most essential factors in deciding whether or not to take out a mortgage is the borrower's level of confidence. Particularly relevant is the public's outlook on the housing market; if they anticipate a decline in home prices, they will put off making a purchase.
- **Mortgage Availability:** Home loans were readily available from various financial institutions during the boom years of 1996–2006. Large income multiples (e.g., five times income) could be borrowed from them. Banks also demanded extremely small deposits (for example, a full mortgage). Since more individuals could afford to buy homes thanks to the easing of lending requirements, demand for housing surged.

- **Supply:** Prices rise when supplies are low. As a result of an oversupply, prices will decline. When the Irish housing market boomed from 1996 to 2006, for instance, about 700,000 new houses were constructed.
- **Affordability/House Prices to Earnings:** House prices to income ratios affect demand. House prices are expected to rise in relation to income, which means fewer individuals will be able to purchase them.
- **Geographical Factors:** The housing market is often rather regional.

III. Machine and Deep Learning Approaches for Housing Price Forecasting

Machine learning's incorporation into home price forecasting has revolutionized market analysis and predicted accuracy. This section delves into important ML algorithms, such as Random Forest, XG Boost, and Cat Boost, showcasing their prediction power and capacity to adapt to various housing datasets. Performance evaluation metrics that guide model reliability, such as MAE, RMSE, and R^2, are given emphasis. In addition, a comparison of ensemble and tree-based approaches is shown to highlight their advantages and disadvantages in different real estate scenarios [10][11]. This identifies critical insights into model selection, scalability, and the balance between interpretability and performance in real-world property valuation contexts.

A. *Overview of Key Machine Learning Models*

ML has taken a central role in the housing price prediction as it models complex patterns and accommodates various housing datasets. Some supervised learning algorithms have been tested upon real properties valuation in the course of the previous years, and each of them with its advantages or disadvantages regarding either accuracy, interpretability, and data independence. This section provides ML algorithms that are capable of analysing the dataset and forecasting home values:

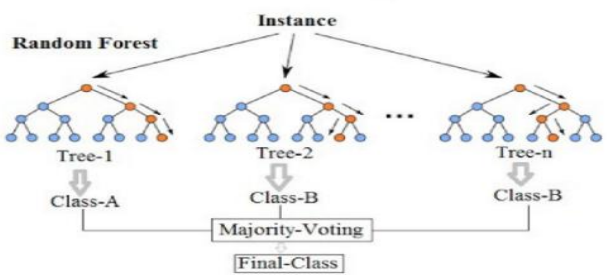

Fig. 2. Random Forest Model

- **Random Forest (RF):** A common ensemble algorithm involving the combination of a number of decision trees to lessen overfitting and enhance foretelling sensitivity., indicated that RF provided resounding performance on prediction of house prices, diversely on medium-sized datasets. Figure 2 shows the RF model.
- **Gradient Boosting Machines (GBM) & XGBoost:** These boosting algorithms repetitively transform weak learners into considerably high-accuracy interpretations of predictions [12]. They have proven that XGBoost coupled with hedonics pricing and outlier filtering, is more effective in modeling the property values as opposed to the methods used in the past.
- **Support Vector Machine Model:** The first step's detected SVM is then used as classifiers to predict future market moves. SVM can transform linear or non-linear input attribute spaces into high dimensional feature spaces, which makes it helpful for modelling housing prices in Figure 3.

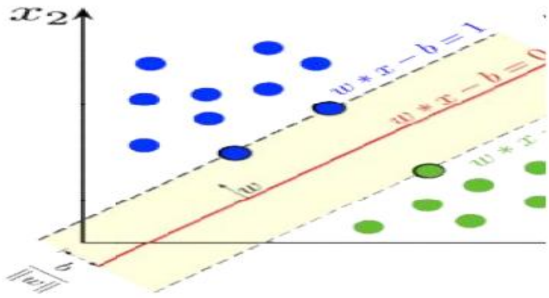

Fig. 3. Support Vector Machine Model

The capacity to simulate complicated, nonlinear correlations in high-dimensional property records is what makes deep learning such a game-changer when it comes to predicting home prices. Unlike traditional statistical or rule-based models, deep learning techniques such as CNNs, RNNs, and multimodal architectures leverage structured and unstructured data to improve forecasting accuracy [13]. This section explores core DL models, integration of multimodal sources like text, images, and geolocation, and highlights their growing impact on real estate valuation. Additionally, it outlines key limitations including data dependency, black-box interpretability, and computational demands, emphasizing the balance between predictive power and model transparency.

1) *Introduction to Deep Learning Models*

A potent method for predicting home prices in recent years has been deep learning, which can automatically understand complicated patterns from huge datasets. For the simple reason that deep learning can pick up on these patterns on its own. Heuristics and rule-based systems are examples of the manual procedures that the real estate sector has historically used for house price prediction.

Housing prices are significant for many real estate sector players since they provide valuable information for making informed decisions. Predicting the future value of a home in a certain area by analyzing historical data and applying statistical models is known as housing price prediction. A common issue in real estate is the need for reliable valuation forecasts to help both buyers and sellers make educated decisions.

A regression model is a statistical tool for predicting a continuous result variable (like home prices) from a set of discrete predictor variables. Economists, survey takers, and anybody else interested in data analysis can benefit from these models. This type of model is employed to infer future outcomes by fitting a line to the raw data.

These models help in fitting a line to the data, which allows for better prediction making [14][3]. Their underlying premise is that the variables serving as predictors and those serving as outcomes are linearly related. DL is an algorithm used in ML that is based on characterizing learning data. DL and shallow learning are related concepts.

- **RNN (Recurrent Neural Network):** One of the ANN models made to handle sequential data, including text, audio, and time series, is the RNN. In contrast to classic neural networks that analyze inputs in a single forward pass, Recurrent connections in RNNs enable them to remember information from previous inputs and utilize it to affect how current inputs are processed. RNNs have recurrent connections, as shown in Figure 4.

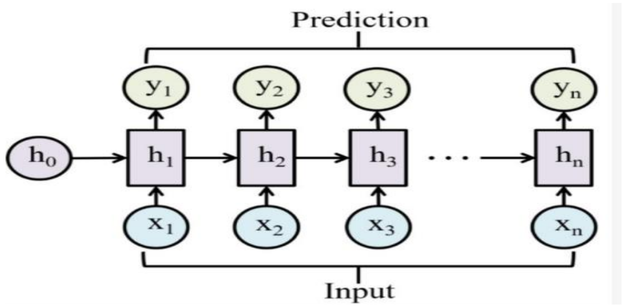

Fig. 4. Basic RNN Architecture.

- **LSTM (Long Short-Term Memory):** The TensorFlow and Kera's libraries were used to build an LSTM prediction model in Figure 5. LSTM was created specially to manage sequential data's long-term dependencies. The model may regulate the information flow and decide whether to accept or reject data according to its applicability to the current job by utilizing gate mechanisms [15]. This leads to an improved ability to manage long data sequence.

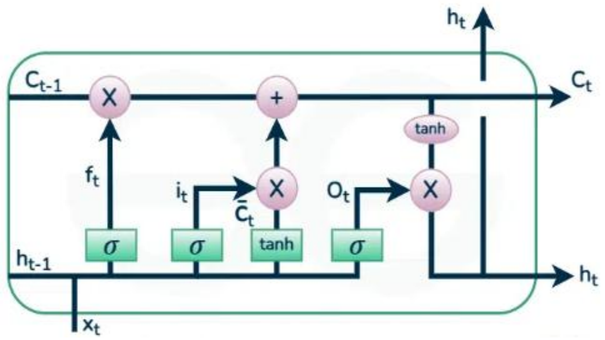

Fig. 5. LSTM architecture

2) Integration of Multi-Modal Data Sources

Multi-modal deep learning methods in computer vision and natural language processing (NLP) have made us want to look into how to use more house-related features to make house price predictions even more accurate. Typical real estate websites, as far as it can tell, include details on the house and its neighborhood (including its location and areas of interest), a brief written description, and a gallery of photos. All of these things work together to help people choose the right house for them.

The emergence of machine learning has led to new hybrid models that combine several methodologies. These models outperform classic statistical models in terms of predicted accuracy. Newer models and methods have surpassed what early neural networks were capable of [16][17]. These include ANN-GIS, PSO-SVM, and others. Multimodal machine learning allows for a thorough examination of housing markets by combining various data sources such as text, photographs, and geographic information. This allows for the capture of both geographical and temporal dynamics. With this method, interpretability and forecast accuracy have been greatly enhanced, marking a significant advancement in the discipline. Figure 6 depicts the aforementioned tendencies and provides a synopsis of the two big milestones in real estate valuation and the continuing subfield of multimodality.

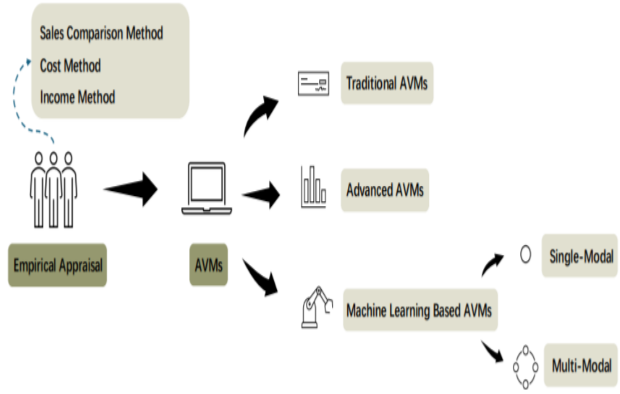

Fig. 6. Development Trends in Real Estate Appraisal

3) Challenges and Limitations of Deep Learning

DL methods have demonstrated potentials in capturing nonlinear, high-dimensional relationships in housing price forecasting, their effectiveness remains to encounter various obstacles in practice [18]. These constraints are due to data quality, interpretability of models, overhead costs and limitation into real estate analytics. Key Challenges Identified:

- **High Data Requirement:** Models of deep learning like LSTM, RNNs, and transformers require large amounts of labelled and high-quality data. Real estate in situations, such like small or emerging markets, where data is limited or noisy, DL can be hindered.
- **Lack of Interpretability (Black-Box Nature):** DL models tend to become more of black box, and it is hard reasoning behind a particular price forecast. This becomes a problem in sectors such as real estate where transparency is crucial to the stakeholders including buyers, agents and policy makers.
- **Modality Fusion and Feature Engineering:** Data types that are of different categories existing in the form of images, geospatial data, textual descriptions, etc., are very difficult to incorporate in DL models. The correct way of aligning, representing and meshing such data remains a developing field.
- **Computational Complexity and Resource Demand:** Deep models with efficient training usually demand strong GPUs and tuning of hyperparameters which are also computationally costly in comparison to traditional ML models.

IV. Datasets, Features, and Challenges in Housing Price Prediction

Accurate housing price forecasting relies heavily on high-quality datasets and robust feature engineering strategies. Public datasets such as Ames Housing and Kaggle's competition data have become standard benchmarks, aiding in the evaluation of various ML and DL models. Effective feature selection enhances model interpretability and predictive strength, with techniques ranging from filter-based methods to embedded regularization strategies. However, challenges like data heterogeneity, limited dataset diversity, and feature redundancy persist. This section explores widely used datasets, essential feature engineering practices, and prevailing limitations while also outlining future research directions such as multimodal data integration and explainable AI techniques to strengthen prediction models.

A. Overview of Commonly Used Datasets

Trainable, quality datasets are vital in attaining stable, highly accurate predictive ML and DL models of housing prices. Generalizability and sensitivity of models are directly affected by the availability, granularity and richness of data [19]. During past ten years numerous benchmark and property-specific datasets have been popular in the literature to test and learn predictive models in the property field. The widely Used Public Datasets are given below:

- **Ames Housing Dataset:** It is one of the most utilized datasets in the academic literature as the predictor of housing price. It has been applied to benchmark algorithms like XGBoost and random forest among the deep regression models.
- **Boston Housing Dataset:** An early dataset which was used classically [20]. Holds 506 records that contain such features like crime rate, rate of property tax, the number of rooms per dwelling and much more. As much as it has its historical value it is now dated as both small and narrow in coverage and has been shelved in certain libraries.
- **Kaggle Housing Price Competition Dataset:** This dataset is an extension (based on the Ames dataset) of a competition on Kaggle. It still acts as a point of reference in the implementation of regression-based ML algorithms [21]. Highly used in ensemble-method trying and feature engineering pipeline experiments.

B. Feature Engineering and Selection Strategies

In housing price prediction, feature antics and feature selection are very important to transform the model to be more accurate, complex, and interpretable. It is highly probable that predictive performance of ML and DL models depend highly on the extent by which input features reflect the underlying valuation dynamics of real estate assets.

1) Feature Selection Techniques

Feature selection implies the selection of the most pertinent subset of features that will be helpful with respect to prediction.

Popular Methods Include:

- Method of filters (e.g. correlation, mutual information),
- Wrapper techniques (e.g. Recursive Feature Elimination)
- Embedded (e.g. Lasso, Ridge, Elastic Net)
- Tree based measures of importance (e.g., feature importance of XGBoost or Random Forest),
- Permutation importance when a model has no built-in interpretability.

Bayesian Optimization played not only a role of hyperparameters fine-tuning in the hybrid model framework proposed [22]. But also feature subset optimization of models such as Cat Boost and ConvLSTM which means that adaptive selection approaches are effective in achieving performance in more complex, highly dimensional datasets.

C. Key Challenges and Future Directions

ML and DL methods have achieved significant progress in terms of housing price forecasting, some data-related problematic issues still require addressing. These difficulties limit the scalability of predictive models, accuracy of predictive models and applicability of predictive models to the real world. This section describes the main limitations identified on different datasets and features and provides recommendations with regard to the further improvement of the data quality and model accuracy.

- **Data Heterogeneity and Fusion Issues:** Real estate data comes in structured form (e.g. size), semi-structured form (e.g. descriptions) and unstructured (e.g. images) making integration complex.
- **Dataset Scarcity and Quality Limitation:** Popular datasets such as Ames and Seattle are not geographically diverse and do not cover social-economic extremes.
- **Feature Redundancy and Interpretability:** Multicollinearity can be applied by the use of redundant features (e.g. area vs. rooms). Although DL models are accurate, they usually become opaque.

Future Research Directions: To address these challenges some steps followed:

- Creating normalized and open multimodal real estate data with geographic, temporal, and socioeconomic properties.
- Using explainable AI (XAI) tools like SHAP and LIME to make extrapolation to deep models.
- Investigating Transfer learning and domain adaptation as one way of not relying on large labels datasets[23].
- Designing and Engaging in causal modelling to draw a line between correlation and reality in terms of housing prices drivers.

V. Literature of Review

This section presents a literature review on housing price forecasting, emphasizing machine learning and deep learning. Table I summarizes key studies, findings, challenges, and future directions, offering a comprehensive overview of recent advancements and research gaps in predictive modeling for real estate valuation.

Huang et al. (2025) aim in conducting this study was to fill this void by analyzing previous studies. To start, they take a look at the history of real estate evaluation and then put out two research topics regarding performance and fusion that could help improve the reliability of appraisal findings. Then, for the first time, they give a thorough categorization and characterization of modalities utilized in real estate evaluation followed by an explanation of multimodal machine learning. Within the context of these two study issues, they investigate works pertaining to data and techniques, as well as the methodologies used to assess them, in order to guarantee clarity. In addition, they have compiled a list of detailed application domains. Lastly, they touch on potential areas for further study, such as the role of technology and modality contributions, multimodal complementarity, and future research prospects [24].

Çetin (2025) predicting home selling prices utilizing a broad dataset including structural, locational, and economic factors was the goal of this study, which aimed to examine the interaction between sophisticated sampling approaches and machine learning models. The research thoroughly examines the effects of five different sampling strategies on a range of machine learning algorithms, with a focus on Stratified

Extreme Ranked Set Sampling (SERSS). These approaches are Cluster, Bootstrap, Systematic, and Random Sampling. By capturing both core and extreme data patterns, SERSS greatly improves the generalizability and resilience of prediction models. It outperforms traditional methods when it comes to retaining dataset variability. The predicted accuracy of ensemble methods such as Cat Boost and Random Forest, as well as similarity algorithms like FIkNN, was consistently higher than that of structured sampling techniques. These methods achieved Mean Absolute Error (MAE) levels ranging from 85 to 650 and high R^3 values [25].

Mathotaarachchi, Hasan and Mahmood (2024) research filled that need by reviewing the existing literature on ML techniques, specifically looking at neural networks, ensemble methods, and sophisticated regression techniques. They point out important areas where research is lacking, including how little is known about interpretability in ML forecasts and hybrid ML-econometric models. Using a separate dataset, they perform generalization testing to ensure that regression models are robust. The results show that regression models can be used to forecast real estate prices in different types of marketplaces. Their results highlight the need to fill up research gaps in order to move the area forward and make ML approaches more useful for real estate price prediction [23].

Zhan et al. (2023) showed that by combining Hybrid Bayesian Optimization (HBO) models with Stacking (HBOS), Bagging (HBOB), and Transformer (HBOT) techniques, a framework of home price prediction models might improve forecasting performance and solve this problem. The improved prediction accuracy and stability are the result of these hybrid models' use of Bayesian Optimization for hyperparameter adjustment. In addition, the suggested methodology can accurately and statistically evaluate how well house price forecasting models perform in various scenarios. The HBOS-CatBoost model outperformed the HBOB-XGBoost and HBOT-ConvLSTM models in terms of root-mean-squared error (RMSE), reducing RMSE by 5.11% and 25.56%, respectively. An extensive performance evaluation framework, new hybrid models for better home price prediction, and a huge multi-source dataset are the key contributions of this work [26].

Ozalp and Akinci (2023) looked at the housing valuation for Artvin City Centre using four tree-based ML algorithms: XGBoost, AdaBoost, Gradient Boosting Machine (GBM), and Random Forest (RF). Based on the study's findings, the XGBoost and RF algorithms were shown to have the best performance in estimating house worth. The XGBoost strategy achieved an R2 of 0.705, while the RF algorithm had an RMSeq of 0.701. One conclusion is that ML algorithms, especially XGBoost and RF, do a good job of residential real estate evaluation with relatively small datasets, and that their success rate improves with larger datasets [27].

Zaki et al. (2022) study stated goal was to use MLTs to foretell future home prices. Here, they use hedonic regression pricing and the Extreme Gradient (XG) boosting algorithm to forecast the property's value. The potential of machine learning methods (MLTs) to improve economic activity through more precise home price prediction is the focus of this study. The XGBoost algorithm is combined with the outlier sum-statistic (OS) method in this piece. A key indicator of economic growth in the real estate business is the price of property. When predicting future home prices, both the XGBoost and hedonic pricing models take into account thirteen different factors [28]

TABLE I. COMPARATIVE ANALYSIS OF RECENT STUDIES ON PREDICTIVE MODELING TECHNIQUES AND DATA STRATEGIES IN REAL ESTATE VALUATION

Reference	Study On	Approach	Key Findings	Challenges	Future Direction
Huang et al. (2025)	Multimodal ML techniques for real estate appraisal	Survey of fusion and modality in ML	Classified modalities in real estate; proposed research questions for performance and fusion	Lack of standard modality definitions; fusion strategies underexplored	Explore multimodal complementarity and modality contribution to appraisal accuracy
Çetin (2025)	Impact of sampling techniques on ML prediction	Compared SERSS, Cluster, Bootstrap, etc. on CatBoost, RF, FIkNN	SERSS improves generalizability; CatBoost and FIkNN yielded high accuracy	Handling data extremities; sampling bias	Refine sampling for better ML robustness in diverse datasets
Mathotaarachchi et al. (2024)	Evaluation of ML approaches in real estate pricing	Systematic review of neural nets, ensembles, and regression	Highlighted gaps in hybrid models and interpretability; regression models generalized well	Interpretability; limited hybrid ML-econometric models	Develop hybrid interpretable ML frameworks for real-world pricing
Zhan et al. (2023)	Forecasting house prices using hybrid deep learning models	HBOS (Bayesian + Stacking), HBOB, HBOT	HBOS-CatBoost had lowest RMSE; robust across scenarios	Model complexity; tuning instability	Extend hybrid DL frameworks across geographies and datasets
Ozalp and Akinci (2023)	Tree-based ML algorithms for house value in Artvin, Turkey	RF, GBM, AdaBoost, XGBoost	XGBoost and RF performed best; effective even with limited data	Data sparsity; regional generalizability	Test models in broader regions with more features
Zaki et al. (2022)	Integration of hedonic regression with ML	XGBoost + Outlier Sum-Stat + Hedonic pricing	Improved predictive performance using 13 key variables	Need for economic model interpretability	Further integration of econometrics with ML for policy-relevant valuation

VI. CONCLUSION AND FUTURE WORK

The adoption of ML and DL approaches has brought about a dramatic shift in the housing price forecasting scene. However, when it comes to modelling the complicated nonlinear interactions and high-dimensional feature spaces seen in real estate data, traditional statistical techniques, while useful and interpretable for simple scenarios, are inadequate. With enhanced generalization and feature handling, modern ML algorithms like Support Vector Machines, Random Forest, and XGBoost provide greater prediction performance. In addition, when paired with multimodal data sources like photos, geolocation, and text descriptions, deep learning architectures like LSTM and RNN offer excellent capabilities for capturing non-linear trends and temporal connections. In order to provide reliable predictions, this paper stresses the significance of feature engineering, model interpretability, and high-quality datasets. Commonly used metrics now come from publicly available datasets such as the Ames Housing

Dataset and the Kaggle competition results. However, problems such as data heterogeneity, lack of geographical diversity, and the opaque structure of deep models continue to be obstacles to their practical use. Thus, real estate forecasting success necessitates a delicate balancing act between predictive precision and explanatory power.

To improve the transparency of deep learning models in housing price predictions, future research should concentrate on incorporating explainable AI approaches like SHAP and LIME. Improve the generalizability of the models by creating real estate datasets that are diverse in terms of geography and use several input modalities. Addressing data scarcity could be achieved through the use of transfer learning and domain adaptation approaches. On the other hand, causal modelling can help differentiate between correlation and the real drivers of property value.

REFERENCES

[1] L. el Mouna, S. Hassan, Y. Haynf, M. F. Nann, and C. S. K. Tekouabou, "A Comparative Study of Urban House Price Prediction using Machine Learning Algorithms," *E3S Web Conf.*, vol. 418, pp. 1–6, Aug. 2023, doi: 10.1051/e3sconf/202341803001.

[2] P.-F. Pai and W.-C. Wang, "Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices," *Appl. Sci.*, vol. 10, no. 17, Aug. 2020, doi: 10.3390/app10175832.

[3] S. J. Wawge, "Evaluating Machine Learning and Deep Learning Models for Housing Price Prediction: A Review," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 5, no. 11, pp. 367–377, Apr. 2025, doi: 10.48175/IJARSCT-25857.

[4] A. Yağmur, M. Kayakuş, and M. Terzioğlu, "House price prediction modeling using machine learning techniques: a comparative study," *Aestimum*, vol. 81, pp. 39–51, Feb. 2023, doi: 10.36253/aestim-13703.

[5] S. Chatwani, M. Chotaliya, and V. Sawant, "House Price Prediction Review," *Int. J. Comput. Appl.*, vol. 184, no. 7, pp. 35–39, 2022, doi: 10.5120/ijca2022922043.

[6] W. Chen, S. Farag, U. Butt, and H. Al-Khateeb, "Leveraging Machine Learning for Sophisticated Rental Value Predictions: A Case Study from Munich, Germany," *Appl. Sci.*, vol. 14, no. 20, 2024, doi: 10.3390/app14209528.

[7] C. Li, "Review of Housing Price Forecasting Methods Based on Machine Learning and Deep Learning," *Appl. Comput. Eng.*, vol. 118, no. 1, pp. 146–151, Feb. 2025, doi: 10.54254/2755-2721/2025.20931.

[8] V. Prajapati, "Enhancing Supply Chain Resilience through Machine Learning-Based Predictive Analytics for Demand Forecasting," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 3, pp. 345–354, 2025, doi: 10.32628/CSEIT25112857.

[9] M. Munusamy, C. Muthuveerappan, M. Baba, and M. N. Abdullah @ Mohd Asmoni, "An Overview of the Forecasting Methods Used in Real Estate Housing Price Modelling," *J. Teknol.*, vol. 73, no. 5, pp. 189–193, Mar. 2015, doi: 10.11113/jt.v73.4337.

[10] R. Patel and P. Patel, "Machine Learning-Driven Predictive Maintenance for Early Fault Prediction and Detection in Smart Manufacturing Systems," *ESP J. Eng. Technol. Adv.*, vol. 4, no. 1, pp. 141–149, 2024, doi: 10.56472/25832646/JETA-V4I1P120.

[11] R. Q. Majumder, "Machine Learning for Predictive Analytics: Trends and Future Directions," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 4, pp. 3557–3564, 2025.

[12] S. S. S. Neeli, "Key Challenges and Strategies in Managing Databases for Data Science and Machine Learning," *Int. J. Lead. Res. Publ.*, vol. 2, no. 3, p. 9, 2021.

[13] N. Malali, "Model Validation and Governance for AI / ML in Actuarial Applications," *Int. Res. Journals*, vol. 12, no. 4, 2025.

[14] D. S. R. Kumar, S. Bhatt, and H. Phudinawala, "Predicting House Price with Deep Learning: A Comparative Study of Machine Learning Models," *Int. J. Multidiscip. Res.*, vol. 5, no. 2, pp. 1–10, 2023, doi: 10.36948/ijfmr.2023.v05i02.1849.

[15] V. Singh and P. Gupta, "AI-Driven Cash Flow Forecasting in ERP Systems: Integrating Economic Indicators and Real-Time Transaction Data Using LSTM-Based Time-Series Models," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 42s, 2025.

[16] M. H. Hasan, M. A. Jahan, M. E. Ali, Y.-F. Li, and T. Sellis, "A Multi-Modal Deep Learning Based Approach for House Price Prediction." 2024. doi: 10.48550/arXiv.2409.05335.

[17] S. Garg, "Next-Gen Smart City Operations with AIOps & IoT : A Comprehensive look at Optimizing Urban Infrastructure," *J. Adv. Dev. Res.*, vol. 12, no. 1, 2021.

[18] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.-H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism," *IEEE Access*, vol. 9, pp. 55244–55259, 2021, doi: 10.1109/ACCESS.2021.3071306.

[19] Y. Han, "Price Prediction of Ames Housing Through Advanced Regression Techniques," *BCP Bus. Manag.*, vol. 38, pp. 1965–1974, Mar. 2023, doi: 10.54691/bcpbm.v38i.4014.

[20] Y. Zhang, "Application of Machine Learning in Boston House Price Prediction," *Adv. Econ. Manag. Polit. Sci.*, vol. 43, no. 1, pp. 176–184, 2023, doi: 10.54254/2754-1169/43/20232154.

[21] J. Yu, "Prediction on Housing Price Based on the Data on Kaggle," in *Proceedings of the 2022 3rd International Conference on E-commerce and Internet Technology (ECIT 2022)*, Dordrecht: Atlantis Press International BV, 2023, pp. 627–634. doi: 10.2991/978-94-6463-005-3_64.

[22] J. Guo, "Feature Selection in House Price Prediction," *Highlights Business, Econ. Manag.*, vol. 21, pp. 746–752, Dec. 2023, doi: 10.54097/hbem.v21i.14755.

[23] K. V. Mathotaarachchi, R. Hasan, and S. Mahmood, "Advanced Machine Learning Techniques for Predictive Modeling of Property Prices," *Information*, vol. 15, no. 6, May 2024, doi: 10.3390/info15060295.

[24] C. Huang, Z. Li, F. Chen, and B. Liang, "Multimodal Machine Learning for Real Estate Appraisal: A Comprehensive Survey," 2025.

[25] A. İ. Çetin, "Evaluating Machine Learning Models for House Price Prediction with Different Sampling Techniques," *Int. J. Comput. Exp. Sci. Eng.*, vol. 11, no. 2, Jun. 2025, doi: 10.22399/ijcesen.2870.

[26] C. Zhan, Y. Liu, Z. Wu, M. Zhao, and T. W. S. Chow, "A hybrid machine learning framework for forecasting house price," *Expert Syst. Appl.*, vol. 233, Dec. 2023, doi: 10.1016/j.eswa.2023.120981.

[27] A. Y. Ozalp and H. Akinci, "Comparison of tree-based machine learning algorithms in price prediction of residential real estate," *Gümüşhane Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 14, no. 1, pp. 116–130, Oct. 2023, doi: 10.17714/gumusfenbil.1363531.

[28] J. Zaki, A. Nayyar, S. Dalal, and Z. H. Ali, "House price prediction using hedonic pricing model and machine learning techniques," *Concurr. Comput. Pract. Exp.*, vol. 34, no. 27, 2022, doi: 10.1002/cpe.7342.

