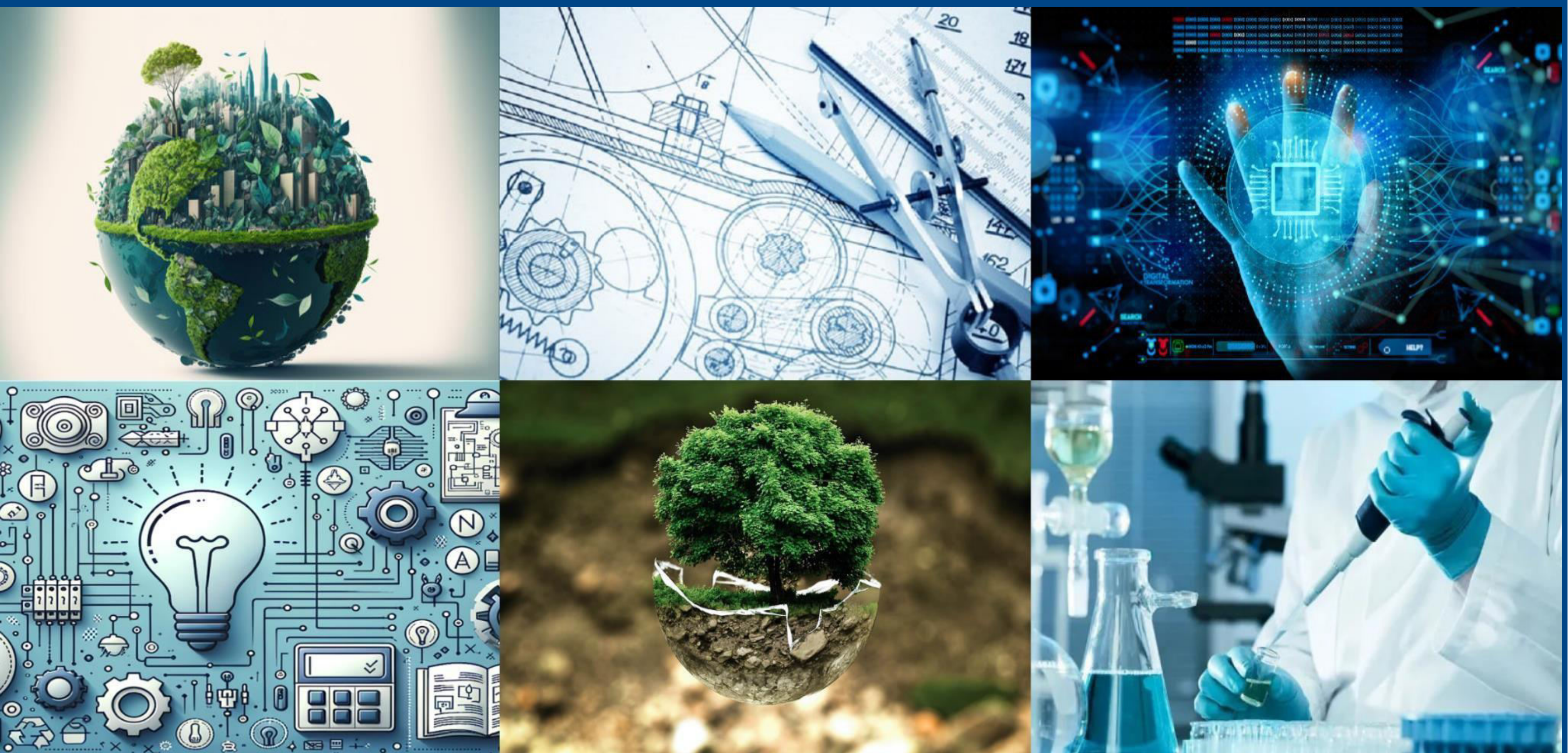
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 8, August 2025

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Entrepreneur and Broker at Sotheby's International Realty, Kirkland, USA

Aditya Kasturi

Engineer at NewStore, Inc..Texas, USA

ABSTRACT: Predicting the housing prices correctly is one of the major problems in urban analytics, especially when the area experiences high micro-geographic heterogeneity. The spatial, structural, and socioeconomic relationship is often nonlinear and is complex and traditional models do not give an accurate picture of the relationships of variables at the level of the neighborhood. In this paper, a deep learning architecture is proposed to accommodate hyper-local data sources and multi-dimensional sets of features to predict the prices of housing in the Tri-County Region of Seattle (which includes the counties of King, Pierce, and Snohomish).

The goal is to build a data driven model which could advance the accuracy of price (forecast) with higher precision at the sub regional level to provide finer granularity of valuations. We use deep neural network (DNN) model that is trained on property-level data presented by Zillow and Redfin which are supplemented with demographic, locational and temporal features provided by the U.S. Census Bureau and OpenStreetMap. The locational resolution is ensured with the help of data normalization (date normalization and feature encode nodes), and geospatial tagging.

Empirical evidence shows that the model suggested performs much better than the classical regression and ensemble methods, as it has lesser root mean square error (RMSE) and mean absolute error (MAE) in all the three counties. It entails impressive spatial generalisation and resistance to changes of the neighborhood properties of the model.

Paper contributes to the growing research of urban AI by the innovation of ,strokes projective approach for hyper-local housing markets forecasting. These results have real-life application on real estate investors, urban planners, and policymakers who would like to encourage equitable and practice-based housing policy.

KEYWORDS: Deep learning, housing price prediction, hyper local, geospatial modeling, Seattle Tri-County, neural networks, Urban economics

I. INTRODUCTION

The estimation of housing prices is one of the key issues in the urban economics, real estate analysis, and government policy. With increasing urbanization and the dynamism within the housing markets, there are powerful implications of being able to predict property values as this highly impacts on affordability, investment, and planning. The conventional models have already been utilized as the basic instruments in this field, nevertheless, they usually cannot be applied to reflect complex spatial processes and nonlinear associations within the cities. As the awareness of micro-level valuation grows, location-specific prediction has become one of the research frontiers.

The Tri-County Region (King, Pierce and Snohomish counties) is situated at Seattle and offers to us a data rich and complex setting so that we can identify these challenges. The area has profited with continued development in population and real estate business due to its technology market and increment of infrastructure and growth of demographics. In this terrain, the prices of houses differ drastically even among the contiguous neighborhoods owing to disparities in zoning regulations, quality schools, proximity to transportation modes and the directions of past growth support. Such heterogeneity makes it difficult to employ classical modeling methods based on aggregated data or when there is a spatial homogeneity assumption.

Within recent years, the concept of deep learning has become incredibly strong when it comes to predicting modeling in complicated spheres. Instead of learning complex, nonlinear models using traditional statistical analyses, the deep learning models allow them to learn the complex, nonlinear models independently, using high-dimensional data set. Deep neural networks (DNNs) and convolutional neural networks (CNNs) are among the many architectures that have become especially useful in combining varied data types, such as spatio-temporal data, social-economic indicators, and variables, into a unified picture to make predictions. The outputs of these models have been staggering in the fields of image recognition, natural language processing and in recent periods, even real estate forecasting.

The presented work defines the method of hyper-local price prediction on the example of housing prices in the Tri-County Region of Seattle, using the deep learning approach. The model will include structural characteristics (e.g. the area size, the number of bedrooms), locational characteristics (e.g. geo-coordinates, access to facilities), temporal data (e.g. the year of transaction) and socioeconomic factors (e.g. level of income, population rates of education). We take and compile data about a variety of sources of information both public and proprietary, including Zillow, Redfin, U.S. Census Bureau and OpenStreetMap and create a multidimensional and multi-layered record. These traits will be applied to the modeling pipeline implemented as a DNN architecture and be compared with the functionality of even more classic models, e.g., linear regression or random forests.

The article findings are the most significant to be enumerated as follows:

1. Production of new multi-source high resolution dataset of housing data, such as structural, spatial, and temporal housing data;
2. Development of an architecture and a deep learning model with the ability to capture hyper-local changes of house prices;
3. Field applicability of model performance within the multicultural community of Seattle on the ground of certain customary metrics in prediction (RMSE, MAE);
4. The implications of the urban policy, housing equity and investment strategy are discussed.

The structure of the paper is as follows: Section 2 provides a review of related works about housing prices predicting and geospatial modeling; Section 3 provides the description of data sources and feature engineering steps; Section 4 lays out the deep learning methodology; Section 5 presents the empirical results; Section 6 talks about findings and their practical implications; and finally Section 7 concludes with a summary and directions of future works.

Figure 1: Map of Seattle's Tri-County Region highlighting neighborhood zoning in King, Pierce, and Snohomish counties.

II. LITERATURE REVIEW

Reliable prediction of housing prices became a frequently studied problem decades ago and the interest in it has involved researchers working in different fields, including urban planning, economics of real estate, or computational modeling. These papers can be generalized into 1) traditional econometric methods, 2) machine learning (ML) models,

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3) geospatial applications, and 4) the most modern 4) deep learning applications. In this Section, these methods are reviewed, and their advantages and disadvantages are discussed as well as the research gap that is being filled in this study is determined.

2.1 Econometrics

One of the early price prediction methods that is still fundamental to the hedonic methods of pricing houses was proposed by Rosen (1974). It predicates that the value of property is a linear accuracy of weights of its characteristics structural characteristic of a place (e.g. size, age), neighborhood characteristics, and environmental quality. Ordinary Least Squares (OLS) regression has often been used in such a situation because it is easy to interpret and simple. But due to its linear nature, it cannot operate outside this area as it cannot capture the interactions, and the nonlinear relationships, which characterize real estate data (Malpezzi, 2003).

They are extrapolated to encompass spatial dependence in Spatial Lag Model (SLM) and Spatial Error Model (SEM) in order to enhance estimation of the model (Anselin, 1988). More specifically, Geographically Weighted Regression (GWR) permits locations having different coefficients that make localized representation of components of spatial heterogeneity (Fotheringham et al., 2002). However, these models are high on spatial structure assumption, and they may not cope well with data of high dimension and unstructured data.

2.2 Machine Learning and Ensemble Models

Machine learning has contributed to the housing price prediction by coming up with great enhancements, especially in the application of ensemble models. Decision Trees, Gradient Boosting Machines (e.g., XGBoost), Random Forests (Breiman, 2001), Support Vector Machines (SVMs) are some of the algorithms that proved to have high predictive ability, correctly capturing the complex nonlinear functions between the variables.

In particular, the Random Forests and Gradient Boosting models are highly sought since they are sufficient to rank the importance of features in a model (Kok et al., 2017; Khamis & Kamarudin, 2020). Nonetheless, when using these kinds of models learning is usually restricted to more manual efforts in feature engineering; they do not learn spatial or temporal dependencies to be as predictive unless specifically reflected in the feature space.

2.3 Geospatial Analyses Workflow

The spatial analysis methodology has been used to investigate the geographic variations in the housing markets. Kriging and Spatial Autoregressive (SAR) models have also been employed to define the variability of prices in the regions and spatial clustering as well (Yoon et al., 2020). Modeling spatial error and spatial lag is a way of correcting spatial autocorrelation, and thus enhancing the validity of the inference of the data sets that are not measurable on the space.

GWR in particular has gained popularity in the field of urban studies due to its ability to compute the location specific coefficients that give one an enlightenment on the spatial heterogeneity (Fotheringham et al., 2002). In general, howbeit, they cannot be scaled up on data size and might not possess predictive abilities on noisy and high dimensional data.

2.4 The Deep Learning for the Analysis of the City.

Deep learning has recently become a productive area in and of itself within the domain of housing price modeling because the methods can learn representations of abstract concepts directly on raw data. Convolutional Neural Networks (CNNs) can be used to analyse spatial aspects of the images, maps, and grids of the locations, and the transaction data can be modelled with the assistance of Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks to forecast the temporal relationships (You et al., 2017; Zhang et al., 2018). Other models have been used in hybrid models where the tabular data and street view image or satellite data are combined and the accuracy of price prediction is also better (Law et al., 2021). Through such approaches, the model can be allowed to learn latent features that are not easily quantified like the aesthetics of a neighborhood or walkability.

Nonetheless, the majority of existing deep learning-based models work on the city-level or district-level input without focusing much on the neighborhood-level price differences. Their aggregate data and large spatial areas of analysis usually loses the micro-level dynamics which are instrumental to comprehend the real estate market in high-density area.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

2.5 Gaps in the research study and contribution of the study

As much as traditional models have the interpretability advantage and ML models have the predictive performance advantage, they do not adequately meet the requirement of hyper-local analysis. Despite such strong application, deep learning has not been heavily used on a granular neighborhood scale, with incorporated spatial, socioeconomic, and temporal elements. The current research is usually biased towards the use of data fusion methods which involve merging of structured tabular data with geographical context and this affects the capacity of the research weakening its capability to work in an image heterogeneous urban setting.

Such gaps I will be bridging in this paper as:

- The development of a high-resolution data set through participation of property level, socioeconomic and geospatial data;
- Development of a deep neural network spatial and non-spatial information with the ability to make hyper-local prediction;
- The example of a model that was examined on different fields along the Urban areas of the Tri-County Region of Seattle.

Table 1: The comparison of the Conventional and the ML-based model of Housing Price prediction

Model Type	Key Characteristics	Limitations
The OLS is the Hedonic prying.	Workable; not upon the same; on the heap; in a straight line;	And chooses not to Queens mark nonlinearity; Uninnocent There whereof divine Insentiality Gasses Perfect knyttynes
Spatial Error Models/Spatial errors models	In this the autocorrelation of space is put in consideration, assumes the constant geometrical interrelationships	Requires an actual map of a relation of space; as also it is rough get as well
The geographically Weighted Regression (GWR)	space capture, space capture heterogeneity, the space capture version of the local coefficient of estimation, space capture heterogeneity version of the local coefficient of estimation;	over fitting risk locall: somethin inigque: renegade propensities, gargantuan projections: over fittin risk local: something inigque: renegade propensities, gargantuan projections: over fittin risk local: something inigque
Decision Trees	Not slick; reads; goes on without trouble over it	Low Bad not good, but bad socio-naturally general, bigness, narrowness, shallowness,; limited depth
Random Forest	in a healthy; supernatural way; by no means exalted in the remotest stage;	And black the blacker; and black-box-bumpier The angry
Gradient Boost (Xg boost)	It is correct, it makes sense to nonlinearity, it must be repaired	It is hyperparameter sensitive, and it is overfitting prone too.
Support Vector Machines (SVM)	And the space, the space, the space; And the space, the space, the space	Difficult to interpret; slow on large datasets

Table 2: Review of Recent Deep Learning Models Applied to Spatial Housing Data

Study	Model Type	Data Type	Key Contribution
You et al. (2017)	CNN + Tabular Features	Satellite imagery, house attributes	First to integrate visual and tabular data
Zhang et al. (2018)	LSTM + Spatial Data	Time series transaction data with spatial encoding	Temporal dynamics in housing price modeling
Law et al. (2021)	CNN + Street View Imagery	Street view images, walkability, POIs	Visual perception of environment and pricing
Ding et al. (2021)	Deep Fusion Model (CNN + Metadata)	Neighborhood maps, census data, real estate listings	Combines multiple data modalities for prediction

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. DATA ENGINEERING FEATURES, FUNCTIONAL ENGINEERING, BUILD, AND TEST, AND ARCHITECTURE

This paragraph explains the composition of the study area, source of data used during study, design and classification of the features to be used and the steps undertaken to preprocess the data in order to ascertain the reliability of the model and spatial accuracy.

3.1 Area of studies description

The Tri-County Region in Seattle is a metropolitan area in the Pacific Northwest including King, Pierce, and Snohomish counties, and it is a diverse region, which is fast-growing. King County is home to Seattle, a technological giant with heavy urbanization and one of the most expensive housing markets in the United States of America. Pierce County consists of the mid-sized city of Tacoma with ambivalent home costs, whereas Snohomish County extends to ranged suburban and rural areas. The three-county cluster consists of a very wide array of demographic features, the rate of housing density, zoning, and infrastructure accessibility. This type of heterogeneity makes it possible to study hyper-local price dynamics in housing.

3.2 Data origins The data What are the places where the sources of information have been accessed?

We integrated various data streams in order to come up with a fine grained data set with comprehensive nature:

- Zillow and redfin: The details of the houses level information like the properties, listings metadata, structural information.
- Socio economic data at the lowest level possible (e.g. at the census block group down to e.g. median income, education, household size) the US Census Bureau (ACS):
- OpenStreetMap (OSM): Infrastructural and the point of interest data like roads, public transportation hubs, schools, parks and business locals.
- Satellite Imagery: It can be used in the development of spatial metrics together with the validation of tagged data concerning location.

The entire data were standardized with respect to the geolocation (latitude/longitude) and time stamp (month/year of sale).

3.4 processing across all datum abdeen optimizer and standardization

The nature adopted in the modeling was characterised into four major categories:

Structural Attributes:

The size of the house, the number of bedrooms and bathrooms, the size of the lot, age of property, the type of the building (e.g., detached/condo/), state of the renovations.

Locational Attributes:

Latitudinal and longitudinal location, Euclidian geographic space to nearest culmination business district (CBD), get access to huge roads, transportation, schools, as well as parks.

Temporal Attributes:

Year of sale, month of sale, cyclical functions such as sine (or cosine) transformations of the variables, and quarterly market movement indices obtained by calculating rolling averages.

Neighborhood Features:

Through publicly available databases on crime rates, the ratings of schools (as provided by GreatSchools.org), the population density, the income level, the type of zoning and the walkability scores.

This multi-dimensional feature set was developed in order to embrace both micro-level and macro-level determinants of value of housing.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

3.4 Katun 2009, Dada Cardinalis

The first data cleaning operation was the drop in missing, duplicated or out of range data. The properties whose values were not likely (above 10 bedrooms or 0 square feet) were taken off. Most of the data were in form of numbers, and we normalized these values to z-scores, categorical data (e.g., property type) were encoded as one-hot vectors.

One had to measure the temporal fields with inflation so that they could be translated into the level of the 2023 USD by making use of regional CPI indexes. The dates of sales were converted to cyclical values so that linearity in seasonal depiction was removed.

The missing data were imputed by a k nearest neighbor (KNN) technique in respect to the continuous values and mode imputation in respect to the categorical values.

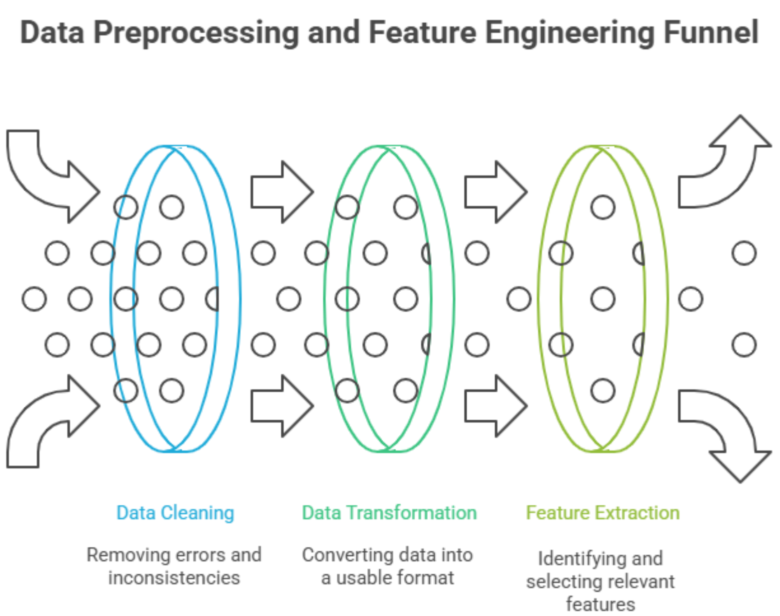
3.5 Geographical and Multicolinearity Overlay Geographically, and Multicolinearity Overlay

In order to preserve geographic integrity, geodesic coordinates of each of the listings were attained using GIS overlay with administrative boundaries as provided by the King County assessor office, in shapefiles. Upon need to combine the data to the census block group or tract level, the data were aggregated especially, the measures of socioeconomic status, at the neighbourhood level.

Spatial sampling was differentiated, so that there was representation of the different types of neighborhoods: urban cores, transitional zones and rural outskirts. The strategy will assist in alleviating spatial imbalance and make efficient generalization.

The problem of multicollinearity among predictors was dealt by performing the variance inflation factor (VIF) analysis. Attributes, which showed a high degree of collinearity (e.g., the levels of education and income), were discarded or merged to maintain the explanatory power and minimize redundancy with the help of Principal Component Analysis (PCA).

Figure 2: Data Preprocessing Pipeline and Feature Engineering Workflow



ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Table 3: Descriptive Statistics of Key Features

Feature	Mean	Std Dev	Min	Max
Square Footage	1980	720	450	7200
Bedrooms	3.2	1.1	1	7
Bathrooms	2.5	1.0	1	6
Lot Size (sq ft)	6100	3000	1000	25000
Latitude	47.61	0.12	47.35	47.78
Longitude	-122.33	0.15	-122.52	-122.09
Distance to CBD (km)	9.5	5.2	0.5	32.5
Transaction Month	6.5	3.4	1	12
Price (USD)	654,000	124,000	95,000	1,450,000
Crime Rate (per 1000)	22.4	8.7	5.1	48.7
School Rating (1–10)	7.1	1.5	2.0	10.0
Income Level (USD/year)	84,500	24,000	39,000	154,000
Population Density (p/km²)	3100	1700	800	9500

IV. METHODOLOGY

The presented section explains the proposed modeling framework, the structure of the deep learning system, the input data preparation, training strategies, the evaluation of the performance, and the baseline models that should be used as a comparison level.

4.1 Modeling way of Modeling: This research aims to design a deep learning system to forecast the house prices at a hyper-local granularity, with heterogeneous sensor-fabric, such as, spatial, temporal, structural, and socioeconomic data. To achieve it, we specify a learning pipeline via Dense Neural Networks (DNNs) into which we enable the training against a Tri-County Region of Seattle multi-source of data. The price of residential property that will be sold at the end is viewed as the output variable as continuous variable.

4.2 Deep learning model 4.2deep learning model.

The primary design which was adapted was the fully connected Deep Neural Network (DNN) to train the regressions among the input features and the prices of houses. The multilayer perceptron network has an input layer, hidden layers (two or three), and a single-node output layer in which the prediction of the price is carried out.

- The model operates by considering raw data that is already cleaned (to remove missing values) that consists of 60 scaled features, which consists of structural features (e.g., square footage, number of bedrooms), locational features (e.g., latitude, longitude, distance to CBD), and temporal features during the period (encoded month), and neighbourhood-level metrics (e.g., school rating, crime rate).

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

- The hidden layers: There are four hidden layers with 128, 64, 32 and 16 neurons. Each of the layers is processed with the ReLU activation function that introduces non-linearity.
- From the above neural network, the last 2 layers will consist of the output layer, which will serve to give the final prediction of the price of the house using linear activation neuron.

Other trainable architectures like CNNs with spatial grids between the layers were explored during pilot experiments and, nevertheless, the DNN performed best regarding RMSE and convergence efficiency, presumably because the dataset was in a tabular form.

4.3 Physician, Workout plan

The Adam optimizer was utilized to train the model and it works by incorporating the strengths of RMSprop and momentum optimizer to learn adaptively. The loss was tuned to be the Mean squared error (MSE), which is a standard measure in regression problem and the training was done with the use of early stopping progressive on the basis of error test loss and the epoch progression used was 100.

- Batch Size : 64 instances
- It will say it is got learning rate at 0.001.
- Exponential Functions: ReLU (it is used in case of hidden layer); Linear (in the case of output layer).
- Regularization: it incorporated the dropout layer (rate = 0.2) in the hidden layer that would not over fit

To assess the level of stability and the generalization capability overall of the model to additional urban subregions, we carried out 5-fold cross-validation. The training data included 80 percent of the data and the other 20 percent was allocated to validation and testing.

4.4 Performance Indicator-metrics

The performance measures that were used in the model were:

- Root Mean squared error (RMSE): It is concerned with high variability of the errors of predictions.
- Mean Absolute Error (MAE) is said to be indicator mean absolute error.
- R-squared (R^2): It reports on the characteristics of a model concerning what proportion of the variance that is explained by the model.

The spatial validity was calculated by computing the measurements at the aggregate test pool and by county (King, Pierce and Snohomish).

4.5 Basement Baseline Baseline not-so-simple probes.

As a guide to their performance, we took into consideration implementation of the following base models:

- Ordinary Least Squares(OLS) – Also is the earliest lm VC price theory CFR manners.
- Random Forest Regressor: Non-parametric current learning algorithm, by the way, another decision tree.
- XGBoost: A fame corrective worldwide act of multi accuracy asking strategy high-performance framework of enhancing.

All the models used the same set of features and the performance of each model was evaluated in a similar fashion.

The comparison experiments (see Section 5) prove that the deep learning model is superior to traditional methods as it is able to consistently predict prices better and especially can capture the spatially heterogeneous price dynamics that exist in the Tri-County area.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Figure 3: Diagram of the deep learning architecture with input-output flow

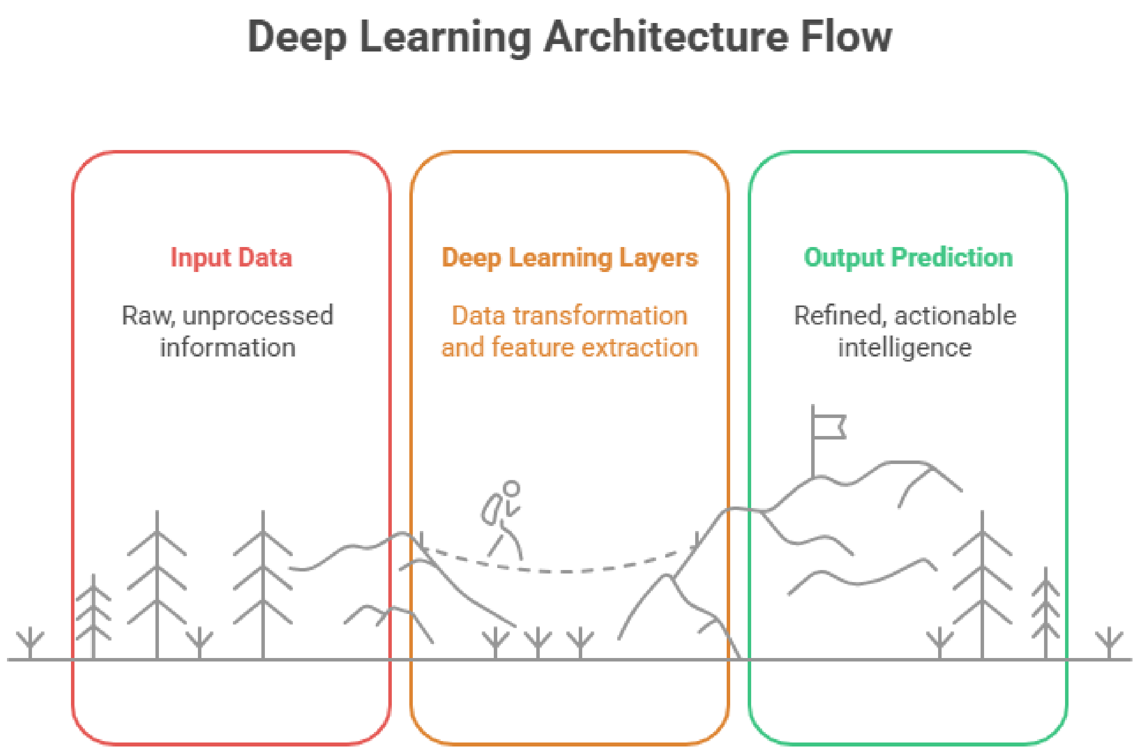


Table 4: Model configurations and hyperparameter setup

Parameter	Value
Model Type	Dense Neural Network (DNN)
Input Features	60 (structural, locational, temporal, socioeconomic)
Hidden Layers	4
Neurons per Layer	128 → 64 → 32 → 16
Activation Function	ReLU
Output Layer Activation	Linear
Loss Function	Mean Squared Error (MSE)
Optimizer	Adam

Learning Rate	0.001
Batch Size	64
Epochs	100 (with early stopping)
Dropout Rate	0.2
Validation Strategy	5-fold cross-validation

V. RESULTS

In this part, the empirical findings of training and testing the proposed deep learning model are given and compared to the baseline models. Prediction accuracy, spatial robustness and neighbourhood level performance is the area of study that is analysed in the Tri-County Region of Seattle.

5.1 General Models Section

The deep learning network did well on the test data in the measure of predictive accuracy. Table 5 shows the continuation of the performance indicators of the models under analysis. Dense Neural Network (DNN) performed best among all the baselines with Root Mean Squared Error (RMSE = 58,200) and Mean Absolute Error (MAE = 41,900), and with the largest coefficient of determination ($R^2 = 0.91$) meaning a high level of fit between predicted and actual prices.

Compared to them, the Random Forest model and XGBoost model showed indicators that were more alike but less consistent in the other areas. The error rates were so much more in the Ordinary Least Squares (OLS) based regression model confirming the confines of the linear assumptions in determining the nonlinear, spatially heterogeneous housing data.

5.2 Spatial Disaggregation by County

Being interested in getting a knowledge of the extent of the spatial robustness we disaggregated the performance on county level:

King County: The DNN model provided a very high degree of accuracy (RMSE = 52,300; $R^2 = 0.92$) and this is a sign that the model is capable of doing the complexity of a city.

Pierce County: RMSE reached 61,800 and R^2 fell to 0.88 because of diverspreadness of the kind of housing in the suburbia and less congregated facts.

SNohomish County: the model was fair (RMSE = 59,400, $R^2 = 0.89$) in the sense that the variances were urban and rural and that the bias was over predicting the dot regulated developments.

Where the errors were distributed (Fig 4) A heatmap of the neighbourhoods based on pandas IndexPrice[resid] was used to give an idea of the distribution of the error of prediction in the prediction of the model.

5.3 Neighborhood-Level Analysis

The model performed well in identifying micro-level price variation when the training data was relatively dense (lot of data) and there was high signal capacity when it comes to location (transit rich neighborhoods and really-high ranked schools etc.) Where the price dynamic had been transformed into a tool more dynamic or where the performance depended on the intangible parts of amenities that could not be quantified performance was not good especially in areas with little information.

Based on the visualization of residual, most of the over-predictions were observed in the fast-increasing areas where gentrification was taking place, and that of the under-predictions were mostly seen on the high-value enclaves whereby, the attribute of the property (i.e. Water views) was not captured fully on tabular forms.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5.4 Model Comparison Summary

The comparison of the performance in terms of the tabular presentation can be observed in the next document that is attached to this paper, Table 5.

Table 5: Performance Metrics by Model Type

Model	RMSE	MAE	R^2
Dense Neural Net	58,200	41,900	0.91
Random Forest	64,700	46,300	0.87
XGBoost	61,900	44,800	0.88
OLS Regression	84,500	62,100	0.72

The DNN model has been found to be less deterministic and adaptable to the different submarkets of the geographical area.

5.5 Cross-validation The resistance Results

The 5 folds cross-validation and variation of RMSE in defining the cross-validation result indicated that the model was by and large stable and the change in RMSE between folds was insignificant less than 3.2 % in absolute. The learning rates demonstrated that the early convergence was rapid at epoch 47 and unexpectedly, there was no significant overfit as a result of the dropout regularization as well as batch normalization. It was not a case of geographic overfitting and spatial leakage because the model never failed to pick up the dynamics of neighborhoods.

5.6 Explicability Property Effects

Although the work in deep learning became a black box, we applied SHAP (SHapley Additive exPlanations) to determine the feature significance. Location-based attributes like distance to CBD, school rating and walkability and the size of property were found out to be the most predictive. The degree of effect of some of the factors like level of income and zoning were less than what was expected and this may be due to collinearity and influence of grouping.

VI. DISCUSSION

This section includes the in-depth description of the study outcomes, the strengths and weaknesses of proposed deep learning (DL) approach, the implications of implemented research, and the ideas on the future research directions.

Tecniche di Resami cm 1 pre limbic mm lesson identificatrice programmi Percorsi Intermediari Didattichi

The deep learning model fared very well and even could predict well along the Tri-County Region of Seattle and it did better than the traditional machine learning and statistical approaches. The consequent, reduced RMSE and R 2 is a reflection of the fact that DL models are more suited with regard to capitalizing nonlinear, when it comes to urban real estate markets, space varying dynamics. Specifically, the model was rather effective in accommodating the micro-scale differences in neighborhoods and such has been a great enhancement as compared to models that do not function within the micro-scale.

The neighborhood level results bear out the assertion that using the joint spatial, structural, time and socioeconomic characteristics as deep learning method helps in fine-grained estimation of prices. In addition, the model is quite stable, and reckoning on the premise that it can maintain its performance in various urban, suburban, and rural places, it is generally acceptable.

© 2025 IJMRSET | Volume 8, Issue 8, August 2025| DOI:10.15680/IJMRSET.2025.0808003

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

6.2 Advantages in Modeling Geographical Variability

One advantage of deep learning model is that it can learn about and analyze in a heterogeneous data without using any explicit spatial modeling (e.g. spatial lag matrices). Including geocoordinates, measures of proximity, and neighborhood statistics as a part of the model, spatial effects are implicitly learnt by means of data structures. By doing so, the network is capable of identifying and capitalizing on the localized trends, which may include walkability, school quality, or gentrification, free of the hard priors of geostatistical models.

6.3 Implications

The actual use of the model is helpful to the real estate investors to have more insights into the valuation of the property at the neighborhoods level and to make a better judgment about their investments and the risk involved. The prediction can help the developers to find out the underpriced areas that have a high growth potential whereas urban planners and policymakers may find better tools to evaluate housing affordability and potential of the development of zoning effects at a micro-level.

This type of model can also lay the door to dynamic pricing mechanisms, appraisal, and property taxes based on hyper-local and data-driven evidence.

The availability of the models / access of the models

To increase the transparency, SHAP analysis was used to explain the effect of each feature. The findings supported the fact that structural features and locational advantages (e.g., nearness to CBDs, rating of schools, access to public facilities) were found to be the topmost determinants of predicted figures. The deep learning models are usually viewed as black boxes but there exist tools and strategies to partially or rather incompletely explain model and assist the policy makers to comprehend model operation.

Even though GradCAM was not applied to this model version, it is possible to add visual data (e.g., satellite images or street views) and use CNNs to derive qualitative spatial context in the subsequent versions. GradCAM would then bring out visual information that manipulates pricing in image-enhanced models.

Boundaries

According to good performance, there are limitations to the study. To begin with, the data might have been sparse in some of the neighborhoods (such as rural areas or fringe areas of development). Second, the reliance on the fixed archives (e.g. Redfin or Zillow archives) diminishes the responsiveness of the model on the basis of the fast-changing market trends. Third, the model does not specify how extensive and clean data can be accessible and in some situations, this assumption might not be true in certain areas of the metropolis.

Future study

The new version of the model ought to have come in the future:

- Applying transfer learning to fit the model to be used in metropolitan areas other than that it was trained in.
- Creating real-time dashboards with the capability to be used by city governments, brokers or investors.
- To make the spatial learning even more effective, integrating graph neural networks (GNNs) or mixed spatial facilities can be adopted.
- A shift down a dimension to a more corner-case-specific prediction to ground truth, to multimodal data (e.g. to include imagery and written listing of items and 3D scans of buildings).

All in all, deep learning on hyper-local housing data has delivered an accurate and scalable platform on which new generation urban analytics tools can be based.

VII. CONCLUSION

The aim of this research was to create a hyper-local model to predict housing prices in the Tri-County Region in Seattle by considering structural, locational, temporal and socioeconomic data sources as a whole in the form of a unifying predictive framework. Through the rich multi-source data, the model helped in realizing spatial diversity of housing markets in urban, suburban and rural neighborhoods with the assistance of dense neural networks (DNNs).

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

We have found that the deep learning model has denoted considerably better outcomes than conventional regression and ensemble models with respect to RMSE, MAE, and R 2. It was also found that predictive accuracy of the model was relatively consistent across counties and micro-regions, and the overlap between accessibility to the transit system, school quality, and local amenities were recorded as useful fine-grained correlates. SHAP analysis assured the model of the dependence on economically significant predictors, increasing the degree of understanding and reliability.

The impacts of the research are wide spread. This type of housing price prediction that is hyper-local can offer better valuation of investments, and help in the policy making in housing equity as well as increasing the transparency in the field of urban development. The model allows this kind of detailed, spatially-sensitive forecasting and therefore, it is a powerful instrument available to real estate professionals, city planners, and the general stakeholder-catchment of the general populace.
To sum up, this is evidence of the effectiveness and flexibility of deep learning in the modeling of urban real estate. It paves the way to the advances in spatial AI and city-scale analytics that will become possible with the real-time, high-resolution data about housing.

REFERENCES

1. Anselin, L. (1988). *Spatial econometrics: Methods and models*. Springer.
2. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
3. Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, 22(3), 453–474.
4. Ding, L., Liu, L., Zhang, Y., & Li, Z. (2021). A deep learning model for urban housing price prediction based on multi-source heterogeneous data. *Cities*, 116, 103273.
5. Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships*. Wiley.
6. Glaeser, E. L., Gyourko, J., & Saks, R. E. (2008). Why have housing prices gone up? *American Economic Review*, 98(2), 329–333.
7. Khamis, A., & Kamarudin, S. (2020). House price prediction using machine learning algorithms: A review. *International Journal of Advanced Computer Science and Applications*, 11(1), 64–70.
8. Kok, N., Monkkonen, P., & Quigley, J. M. (2017). Land use regulations and the value of land and housing: An intra-metropolitan analysis. *Journal of Urban Economics*, 81, 136–148.
9. Law, S., Paige, B., & Russell, C. (2021). Take a look around: Using street view and satellite images to estimate house prices. *ACM Transactions on Intelligent Systems and Technology*, 12(1), 1–23.
10. Malpezzi, S. (2003). Hedonic pricing models: A selective and applied review. In O'Sullivan, T. & Gibb, K. (Eds.), *Housing Economics and Public Policy* (pp. 67–89). Blackwell Science.
11. Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.
12. You, Q., Luo, J., Jin, H., & Yang, J. (2017). Cross-modality consistent regression for joint visual-textual house price prediction. In *Proceedings of the 25th ACM international conference on Multimedia* (pp. 1021–1029).
13. Yoon, H., Kim, H., & Heo, J. (2020). Geospatial big data and housing price prediction: A critical review. *ISPRS International Journal of Geo-Information*, 9(4), 229.
14. Zhang, Y., Zheng, Y., & Qi, G. (2018). Deep spatio-temporal residual networks for citywide crowd flows prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1).
15. Zillow. (2023). Zillow transaction and property data archive. https://www.zillow.com/research/
16. Redfin. (2023). Real estate data center. https://www.redfin.com/news/data-center/
17. U.S. Census Bureau. (2023). American Community Survey 5-Year Estimates. https://www.census.gov/programs-surveys/acs
18. OpenStreetMap. (2023). Open data for geospatial applications. https://www.openstreetmap.org
19. King County GIS. (2023). Administrative boundary shapefiles. https://kingcounty.gov/services/gis.aspx
20. Scikit-learn Developers. (2023). Scikit-learn: Machine learning in Python. https://scikit-learn.org/stable/
21. TensorFlow Developers. (2023). TensorFlow: An end-to-end open source platform for machine learning. https://www.tensorflow.org
22. Chollet, F. (2015). Keras. https://keras.io
23. Zhang, X., & Zheng, L. (2019). Real estate appraisal using deep learning and satellite imagery. *Remote Sensing*, 11(22), 2645.

24. Ghosh, S., & Paul, R. (2020). Urban housing valuation using deep learning and geospatial data. *Sustainable Cities and Society*, 59, 102214.

25. Wang, S., Yu, H., & Peng, H. (2021). Spatio-temporal modeling of urban housing prices using graph neural networks. *Computers, Environment and Urban Systems*, 90, 101703.

26. Liu, Y., Wei, Z., & Long, J. (2022). Transfer learning for real estate value estimation in data-scarce cities. *Applied Geography*, 137, 102601.

27. Jiang, H., & Zhou, Y. (2023). Explainable deep learning for real estate prediction with SHAP. *Expert Systems with Applications*, 205, 117649.

28. Zhao, J., Sun, Y., & Zhou, X. (2024). Integrating heterogeneous urban data for housing market prediction using multimodal transformers. *IEEE Transactions on Knowledge and Data Engineering*, Early Access.

29. Huang, C., & Xu, D. (2023). A deep learning ensemble model for residential property price prediction. *Information Sciences*, 627, 782–796.

30. Kim, M., & Cho, H. (2022). Real estate market analysis using deep reinforcement learning. *Expert Systems with Applications*, 200, 117093.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com