Received: August 14, 2023. Revised: September 15, 2023.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Real Estate Tax Prediction using Deep Neural Network and Bayesian Optimization

Amal R. Saleh[1]* Motaz A. Elsaban[1] Mohamed M. Saleh[1] Hisham M. AbdelSalam[1]

[1]*Faculty of Computers and Artificial Intelligence, Cairo University, Giza, Egypt*
* Corresponding author's Email: Amal_Shoaeb@yahoo.com

Abstract: Real estate tax base assessment and evaluation systems are crucial for financing public services and digital economy transformation. To automate and unify complex tax procedures, an intelligent taxation model is needed. Deep learning neural network (DLNN) models have been limited for small-scale, high-dimensional data, but this paper creates an automated model using DLNN, offering superiority, interpretability, and dependability. The model offers lower error levels and high accuracy. The intelligent automated taxation model will improve real estate tax inspection offices efficiency by enabling precise tax base assessments and valuations. Deep learning techniques may enhance real estate price projections, resulting in more precise assessments of taxes. The findings reveal that, compared to other benchmark price predictors, the proposed model achieves greater accuracy (95%–99%) in different datasets and it has the potential to be generalized for real estate taxation authorities and tax inspection offices. The approach is helpful in automated real estate price prediction and taxation applications. This ground-breaking research study proposes a revolutionary strategy that employs deep learning neural network DLNN and challenges the conventional approaches to real estate tax base assessment.

Keywords: Real estate tax assessment, Real estate price prediction, Deep neural networks, Hyperparameter optimization, Bayesian optimization method.

1. Introduction

Tax base assessment is a crucial supporting strategy for real estate tax management reform. When tax base assessment is used, real estate developers' tax behavior may be continuously regulated, helping to guarantee that real estate taxes adhere to specific norms and enhancing tax credibility. Furthermore, conducting a number of intricate processes, such as field surveys, market research, and data collecting throughout the assessment, takes a lot of time. The burden of the estimators increases significantly, the inaccuracy increases, and the assessment efficiency decreases, which is not appropriate given the direction of the digital economy's growth. The deep learning neural network (DLNN) model is used in tax base assessment in real estate pricing in order to guarantee the objectivity and consistency of real estate pricing, the government should establish a reputable, scientifically rigorous, fair, and unbiased tax base assessment for the purpose of tax collection and levying. In addition, Real estate tax assessments are fundamentally related to the national economy, and economic professionals since real estate price valuation and prediction are based on current real estate historical data related to each district tax office. Real estate value prediction enables people, businesses, and governments to efficiently create their financial strategies specifically real estate tax valuation and levying which in turn gains more revenues for the government's public projects, facilities, and real estate taxation rules. Real estate taxation valuations are impacted by anticipated real estate prices at the national level. Automated and early estimation of the real estate pricing worth helps real estate owners, buyers, real estate developers, and governments to make decisions and/or legislation and regulation. [1] Primarily discuss real estate price prediction literature in real estate, finance, economy, and business studies. The price-prediction models

that can be applied to small datasets with multiple feature columns have received less attention than real-estate price models, which have large datasets in previously studied literature. A real estate price prediction must take into account the complexity of a wide variety of predictor elements that affect real estate prices to be computationally effective and determine the accurate real estate price value [2-3]. Machine learning (ML) techniques have improved decision-making throughout the years in a variety of applications, including predicting real estate market values [4]. The creation of real estate price prediction models that automatically extract the price information from various real-world datasets was spurred by the success of ML techniques in several sectors. To this goal, a wide range of ML applications, including classification and language translation, employ the artificial neural network (ANN), a popular ML approach. The difficulty of training the small and high-dimensional datasets using ANN, however, is a drawback [5], and a deep learning neural network (DLNN) is a popular and effective extension of an ANN that learns through a hierarchical process. The strength and quantity of the dataset have a significant impact on the DLNN's performance. Even though ML and, later, DLNN performance is enhanced by an abundance of data, there a still instances in real-world situations when the dataset size is relatively small. Despite being widely used for price prediction, DLNN use in real-estate applications is constrained since huge data is typically difficult to acquire and small-size datasets make DLNN training more difficult and reduce the accuracy of its price prediction. In addition, DLNN needs hyperparameter optimization methods, to recognize the optimal set of hyperparameter combinations that enhance DLNN prediction performance and accuracy. This paper conducts exhaustive research and experimentation in hyperparameter optimization techniques and concludes that integrating the Bayesian optimization method BOM with DLNN (BOM-DLNN) is the most efficient method to enhance the prediction accuracy of real-estate prices. This research demonstrates the effectiveness of the BOM-DLNN model on small and high-dimensional real estate datasets in real estate tax district office that is made up of some numerical and categorical features for the tax valuation process. To do this, real datasets from the Kaggle real estate dataset repository, seven cities in different countries; Melbourn- Australia, California, Helsinki-Finland King County -USA, and Kingdom Saudi Arabia KSA, , Bucharest, and Paris, with different feature structures and dataset sizes are used to validate our proposed models. To validate the proposed model, its performance is compared to the benchmark conventional ML models; stepwise & and tuned SVM [6] in Melbourne city, Ensemble Model [7] for California city, multilayer perceptron MLP [8] in Helsinki city, Catboost regression [9] for King County city, and artificial neural network ANN [10] in KSA, these models validated with the same datasets. In addition, the number of trainable parameters associated with the number of layers and neurons is customized for each model separately. Additionally, these models utilized various loss optimization and activation algorithms throughout the iteration. As a result, the best network architecture BOM-DLNN model was set up to allow for comparison and assessment of these models. As a result, the performance of the BOM-DLNN model in terms of price prediction is compared to benchmark models utilizing evaluation metrics for mean square error (MSE), root mean squared error (RMSE) [11], and R^2 evaluation metrics. The BOM-DLNN model created in this work is efficient and effective for applications that anticipate real estate prices when each real estate unit's price is defined by small high-dimensional datasets. The main contributions of this paper are briefly summarized as follows:

- We proposed an optimized model of deep learning neural network to automatically select and retrieve the optimal set of hyperparameters to gain the best performance and accuracy of the real estate price prediction model for accurate real estate tax assessment.
- The Bayesian optimization method with the deep neural network was applied to different small and high dimensional benchmarked datasets available and resulted in high price prediction accuracy in all datasets.
- The performance of the proposed optimized DLNN outperforms the benchmarked traditional techniques stepwise & and tuned SVM [6] in Melbourne city, Ensemble Model [7] for California city, multilayer perceptron MLP [8] in Helsinki city, Catboost regression [9] for King County city, and artificial neural network ANN [10] in KSA, models with the same datasets. Therefore, optimized DLNN outperformed the traditional methods and the Bayesian optimization method for DLNN hyperparameters choice is anticipated to enhance the DLNN model's prediction using real datasets for real estate price prediction problems and real estate tax base assessment.

The rest of this paper is organized as follows: Section 2 reviews the related work of real estate

price prediction models and hyperparameter optimization methods. Section 3 introduces the proposed methodology and architecture of a deep neural network optimized by the Bayesian optimization method for the real estate price prediction approach. Section 4 discusses the experimental design and its outcomes, analysis, and results. Finally, section 5 presents the conclusion.

2. Related work

Deep learning neural network DLNN algorithms have recently shown cutting-edge outcomes for classification and regression problems, especially in real estate price prediction on benchmarked problems. Optimized DLNN outperformed other hedonic price and machine learning methods in the presence of complexity, multi-dimensional, and data scarcity in the real estate data in real estate inspection tax offices. Furthermore, traditional approaches such as hedonic and machine learning methods require a domain expert to identify the majority of the applicable features in order to simplify the data and make patterns more apparent to the learning algorithms which is costly and in some cases not available, but the main benefit of deep learning algorithms, is that they attempt to incrementally learn high-level characteristics from data. Hedonic pricing models are one of the most often used strategies in a recent study to operate on actual real estate price data to identify significant price factors. However, it performs poorly because the linear technique HPM for real estate price prediction has difficulty constructing general predictions. Different approaches to deal with the non-linearity problem have been proposed. As a subset of artificial intelligence, machine-learning techniques were used in the real estate dataset to assess the factors that have the greatest impact on real estate prices. Additionally, a variety of machine learning (ML) techniques was used to predict real estate prices, including multiple linear regression [12]; the assumptions and criteria of multiple regression, such as linearity, normality, homoscedasticity, independence, and multicollinearity, must be carefully considered in order to avoid overfitting. In addition, artificial neural network ANN used in [10] to make real estate price predictions in the Kingdom Saudi Arabia used shallow networks; shallow networks suffer from overfitting, computationally expensive, limited Interpretability, and data requirement. Multi-layer perceptron (MLP) used in [8]; despite its capability of applying MLP to complex and nonlinear problems, it is underperformed in small and inaccurate datasets, computationally difficult and time-consuming performance. According to the research, combining the advantages of individual machine learning techniques into a single predictor (an ensemble model) enables the development of a predictor that is more reliable than the individual ML techniques that contributed to it. In [7] used ensemble models to make real estate price assessment in California in the USA, ensemble models offer improved accuracy and performance in complex problems but face computational costs, time complexity, interpretation difficulties, and overfitting and underfitting, based on base models' strength and complexity. In this regard, experiments with real estate price prediction primarily used tree-based ensemble methods such as random forest (RF) in [13-14-15]. RF are not easily interpretable, and computationally intensive for large datasets and it is like a black box algorithm, you have very little control over what the model does. Gradient boosting (GB) [2]; Gradient boosting trees can be more accurate than random forests and capable of capturing complex patterns in the data. However, if the data are noisy, the boosted trees may overfit and start modeling the noise. In addition, [9] used the Catboost regression method for real estate price prediction in KC city in the USA, despite it is powerful in categorical features and offers to process efficiently, but if the variables are not investigated and calibrated, Catboost regression can perform extremely poorly. [6] Used Stepwise regression and tuned SVM in real estate assessment prediction in Melbourne Australia by Stepwise regression, which has the ability to reduce time and effort for the feature selection process, but it might be inconsistent or inaccurate in the feature selection process and suffer from overfitting. In addition, the sample size, the order of the variables, the correlation between the variables, and the degree of significance all affect how stepwise techniques perform, SVM does not perform very well when the data set has more noise and it is more suitable for classification problem when classes are clearly separated from one another. Due to the influence of a wide range of influential factors that affect real estate price valuation, policy adjustment [16], environmental events [17], accessibility to urban service amenities [18], and public transportation services and events [19]. However, Machine learning is less efficient than deep learning in complex and nonlinearity problems. In addition, deep learning is more accurate and scalable than machine learning in prediction problems. Deep learning can handle large and unstructured data and can achieve high accuracy and performance. Deep learning has significantly improved conventional applications in computer vision, speech recognition, and genomics [20, 21]. To

date, various research has investigated the use of artificial intelligence (AI) in a range of construction and real estate prediction applications [1-22]. DLNN uses a backpropagation strategy to learn hierarchically from the preceding levels while employing more hidden layers than ANN [5]. The use of DLNN models for real estate price prediction has recently caught the attention of researchers [23], and applications of various DLNN-based predictors, including convolutional neural network (CNN) [24] and long short-term memory (LSTM) [25], have been explored in the literature. When data size grew up [23] demonstrated that the prediction accuracy of DLNN was increased. They have not looked at DLNN use in the small-sized real-estate dataset; instead, their work is restricted to medium- and large-sized datasets. This study raises concerns about the use of a small dataset. For describing the few price records, the tiny real estate price dataset might include price feature columns. While an increase in price records improves prediction accuracy, an increase in influential feature columns decreases the capacity of the selected DLNN technique to forecast prices. Each pricing information increases the dataset's dimensionality, which makes it harder for the computer to learn (the dimensionality curse). The design of deep neural networks DLNN consists of sequential dense layers starting from the input layer defines the input features followed by the number of sequential hidden layers which consist of a variety of densely connected neurons finally, the output layer [26]. Deep neural network (DLNN) based methods have demonstrated promising results when used in real estate appraisal as an alternative to traditional price prediction approaches [27]. Their key benefit is their capacity to find non-linear correlations between inputs and outputs; as a result, they are well suited to non-linearity prediction for real-estate price evaluation and prediction. [2-28] used a hybrid model to anticipate real estate prices that include a pre-trained CNNs model, and a multilayer perceptron MLP model for tabular dataset/numeric characteristics. Despite the strength of DLNN models employed in prior research, the selection of hyperparameter values has a significant impact on the performance of machine learning models [25]. [29] Introduces studies on house price prediction, which are categorized into those using deep learning methods with the role of joint self-attention mechanisms. Furthermore, deep neural networks can be integrated with dimensionality reduction techniques in real estate price prediction problems and reveal high performance and accuracy such as the PCA-DLNN method [30]. However, few studies paid adequate attention to the hyperparameter optimization methods and more research is necessary to determine the ideal DLNN hyperparameter values for the network structure, optimization function, learning rate, batch size, dropout, regularization, validation split, and activation functions. In addition, various nature-inspired heuristic methods are used for hyperparameter optimization such as monarch butterfly optimization and swarm intelligence [31], [32], and Genetic algorithm [33]. Harmony search algorithm [34] and simulated annealing [35]. Evolutionary optimization [36], multi-threaded training [37]. [38] Used Pareto optimization, also, gradient descent optimization of a directed acyclic graph [39], Bayesian optimization [40], and others. Additional research is required to find the appropriate set of hyperparameters for DLNN. However, most present work on evaluating or predicting real estate prices uses off-the-shelf machine learning or deep learning algorithms without addressing hyperparameter optimization, or any parameter tuning was done. Recently, critical parameters for boosting ensemble regression trees, support vector regression, and Gaussian process regression have been tuned using the Bayesian optimization method (BOM)[41]. There are several commonly used strategies examined for the hyperparameter optimization of the machine and deep learning models these methods are manual search, random search, grid search, and Bayesian model-based optimization is deemed to be the most effective one [42].

3. The proposed method

This section introduces the Bayesian optimization method for optimizing deep neural networks for real estate price prediction problems with small and multidimensional datasets for the assessment of the real estate taxation system. The Bayesian optimization method was utilized to select and optimize the deep neural network hyperparameters for each dataset of the seven benchmarked datasets used in this study. Then build a deep neural network with the optimized architecture and hyperparameters to give the most accurate real estate price prediction models for real estate tax estimation and levying system. The proposed models are evaluated and validated against the conventional benchmarked method by the same real estate datasets.

3.1 Build BOM model

The first step in the BOM technique is to build the BOM model, i.e., introducing the problem as a hyperparameter selection in which the overall target is the optimization of the network architecture.

Concisely, the Bayesian optimization method's main steps are; to build a Gaussian surrogate probability model of the objective function, find the hyperparameters that perform best on the surrogate and apply these hyperparameters to the true objective function, and then update the surrogate model incorporating the new results, finally, repeat previous steps until max iterations or time is reached. Here, the model's evaluation index R^2, and mean squared error MSE serve as the objective functions, while the expected improvement serves as the acquisition function (EI), and a Gaussian process is used for the prior function. Therefore, we wish to use Bayesian optimization to improve the objective function f(x) in R^2, and mean squared error (MSE) in Eqs. (1) and (2)

$$F(X) = \text{argmax } R^2 (x) \text{ for } x \in X \quad (1)$$

$$F(X) \text{ argmin MSE}(x) \text{ for } x \in X \quad (2)$$

3.2 Build deep neural network DLNN

Initializing the model's hyperparameters of the deep neural network is the first stage of the model, these hyperparameters consist of the learning rate, activation function, number of epochs, number of layers, number of neurons within each hidden layer, and optimization functions, these are the primary hyperparameters of a DLNN model. As a result, the normalized feature matrix X* is used to initialize the network's first layer, and the weight and bias parameters are assigned at random, as
Shown in Eq. (3)

$$Z_1 = W_1 X^* + b_1 \quad (3)$$

Where W_1 and b_1 are the weight and bias matrices respectively, and Z_1 is the input of the activation function or pre-activation parameter, then compute the activation function as follows in Eq. (4)

$$A = \phi(Z_1) \quad (4)$$

where A1 denotes the first layer's activation. Here, φ(.) is the activation function. In this work, ReLU, Sigmoid, and Tanh activation functions were iterated by the Bayesian optimization method to identify the optimum activation of the hidden layers and by experimentation Relu; rectified linear unit is working well in various applications as in Eq. (5)

$$f(x) = max(0,x) \quad (5)$$

A linear activation function is employed for the top layer. As a result, the retrieved information is sent to the following layer utilizing the forward propagation, as stated in Eqs. (6) and (7).

$$Z_L = W_L A_{L-1} + b_L \quad (6)$$

$$A_L = \varphi(Z_L) \quad (7)$$

Where A_L is the activation function from the prior layer, W_L is the weight matrix and b_L is bias. L also represents the corresponding layer by its number. The cost function must be calculated using mean squared error (MSE) in Eq. (8).

$$\text{MSE} \quad \frac{1}{n}\sum_{i=1}^{n}(yi - yi')^2 \quad (8)$$

Where yi the actual price of the i^{th} real estate, yi' is the predicted or forecasted price by the proposed model and n is the number of examples in the dataset. The estimated cost function goes to the starting layers through a process known as backward propagation to enhance the weights. The gradient of the computed loss function for the changed parameters is then determined using backward propagation. Here, the adjusted weight and bias are calculated using the mathematical formula found in Eqs. (9) to (11).

$$\frac{\partial j}{\partial W_L} = \frac{1}{n}\left(\frac{\partial L}{\partial Z_L} A_{L-1}^T\right) \quad (9)$$

$$\frac{\partial j}{\partial b_L} = \frac{1}{n}\left(\sum_{i=1}^{n} \frac{\partial L}{\partial Z_{Li}}\right) \quad (10)$$

$$\frac{\partial L}{\partial A_{L-1}} = W_L^T \frac{\partial_L}{\partial Z_L} \quad (11)$$

Where $\frac{\partial j}{\partial W_L}$ and $\frac{\partial j}{\partial b_L}$ are the derivatives of wights $W_L$and biases b_L metrics respectively, in addition, $\frac{\partial L}{\partial A_{L-1}}$ and $\frac{\partial_L}{\partial Z_L}$ are the derivatives of the activated layer A_{L-1} and the pre-activated Z_L respectively. Furthermore, the weight and bias parameters must be adjusted by the preceding phase with Eqs. (12) and (13) using the Adam optimization function [43] which takes the stochastic gradient descent with momentum algorithm with root mean squared propagation RMSprop together and generally it is shown to work well in many deep neural network architectures specifically in real estate price prediction problems.

$$W_L = W_L - \alpha \ V_{dw}^{corr} / (\sqrt{S_{dw}^{corr}} + \varepsilon) \quad (12)$$

$$b_L = b_L - \alpha V_{db}^{corr} / (\sqrt{S_{db}^{corr}} + \varepsilon) \quad (13)$$

Where α is the learning rate hyperparameter which has to be tuned, W_L, b_L, is the updated weights and biases by the Adam optimization function. also, V_{dw}^{corr} and V_{db}^{corr} are the corrected exponential weighted moving average with the momentum of weights and biases respectively, and S_{dw}^{corr} and S_{dw}^{corr} are the corrected root mean squared propagation for weights and biases respectively. and ε is not an important hyperparameter and according to [43] it will be 1e-8. Following that, the described steps must be repeated for the specified number of iterations. Finally, the real estate prediction price is performed using the training parameters and hyperparameters optimized by the Bayesian optimization method. Repeating the stated DLNN-BOM technique over the parameters and hyperparameter optimized by the Bayesian optimization method for all seven datasets. The proposed framework consists of three main phases: Building the Bayesian optimization method BOM, building a deep neural network optimized by BOM method BOM-DLNN, and finally the process of real state tax assessment as illustrated in Fig. 1.

4. Experimental details and results

This section is organized as follows: Section 4.1 introduces the description of the datasets. Section 4.2 presents the evaluation metrics of the proposed models. Section 4.3 presents the experimental system setting. Section 4.4 highlights the experimental results and findings.

4.1 Datasets

This paper makes use of the real estate price dataset from the Kaggle website. In this context, seven real estate pricing datasets are; Romania-Bucharest, USA-California, Finland-Helsinki, USA-king county, kingdom of Saudi Arabia (KSA), Australia- Melbourne, and France- Paris). Each dataset has a different size and explanatory features including price features. The Bucharest dataset contains seven features for independent and dependent variables (price feature), and 3925 records. California dataset consists of 20,640 examples and 14 features. USA-California, Finland-Helsinki, USA-King County, Kingdom of Saudi Arabia (KSA), Australia- Melbourne, and France- Paris). County, Kingdom of Saudi Arabia (KSA), Australia-Melbourne, and France- Paris). (price feature), and 3925 records. California dataset consists of 20,640 examples and 14 features. USA-California, Finland-Helsinki, USA-king county, kingdom of Saudi Arabia (KSA), Australia- Melbourne, and France-Paris). County, kingdom of Saudi Arabia (KSA), Australia- Melbourne, and France- Paris). Each dataset has a different size and explanatory features including price features. The Bucharest dataset contains seven features for independent and dependent variables (price feature), and 3925 records. California dataset consists of 20,640 examples and 14 features. USA-California, Finland-Helsinki, USA-king county, kingdom of Saudi Arabia (KSA), Australia- Melbourne, and France- Paris). Each dataset has a different size and explanatory features including price features. The Bucharest dataset contains seven features for independent and dependent variables (price feature), and 3925 records. California dataset consists of 20,640 examples and 14 features. The Helsinki dataset contains 4,043 observations and 43 features. King county dataset contains 21,597 records and the number of features is equal to 21. For the kingdom Saudi Arabia dataset, there are 3,718 examples and 23 attributes. For the Melbourne real estate dataset, there are 13,580 records and 23 features. Paris dataset contains 10,000 examples and 17 features. To gain better performance and generalization of artificial neural network ANN the data have to be in a good quality state so datasets were gathered, cleaned, and preprocessed; data imputation, outlier detection, deleting irrelevant and unimportant features, normalized, and finally, standardized, feature scaling, and one-hot encoding for categorical data. The dataset was randomly split into 80%, and 20%, for training (estimation), and testing (validation) samples, respectively as Fig. 2.

4.2 Evaluation metrics

Three evaluation metrics were used to measure the prediction accuracy and to evaluate the proposed deep neural network house price prediction optimized by the Bayesian optimization model. Three loss metrics Mean Squared Error MSE, root mean squared error RMSE, [6], and R^2 can be calculated by evaluating the generated models on our dataset, We experimented with the proposed model with a Bayesian optimization hyperparameter. The mean squared error (MSE) and root mean squared error on the training and testing set were calculated and used to evaluate the models in Eqs. (14) and (15).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(yi - yi')^2 \quad (14)$$

$$RMSE = \text{the square root (MSE)} \quad (15)$$

R^2 measures the degree of variation, quantifies the link between forecast and desired price, and fits

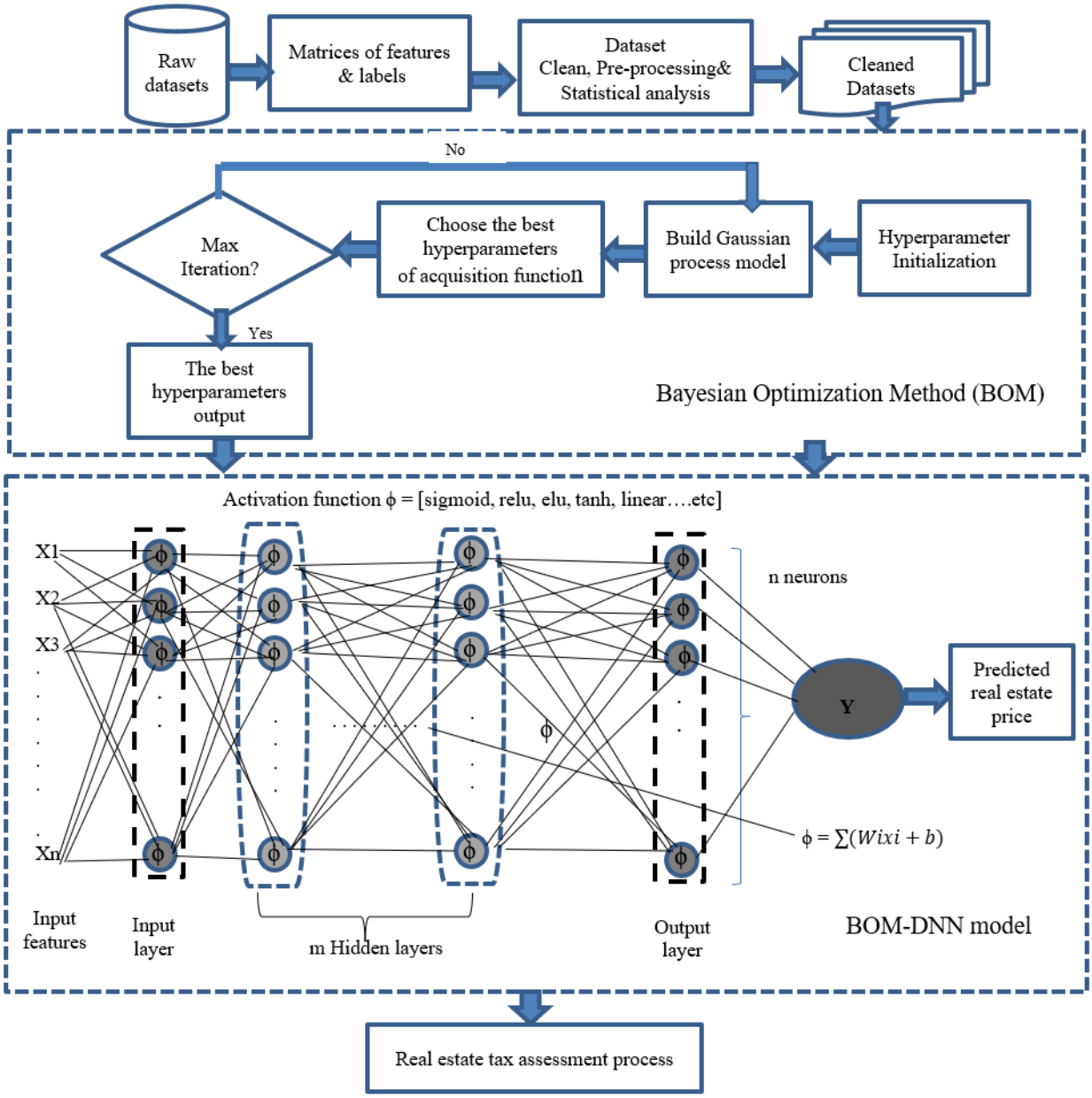

Figure. 1 The proposed deep neural network optimized by Bayesian optimization method BOM –DNN framework for real estate price prediction for real estate tax assessment

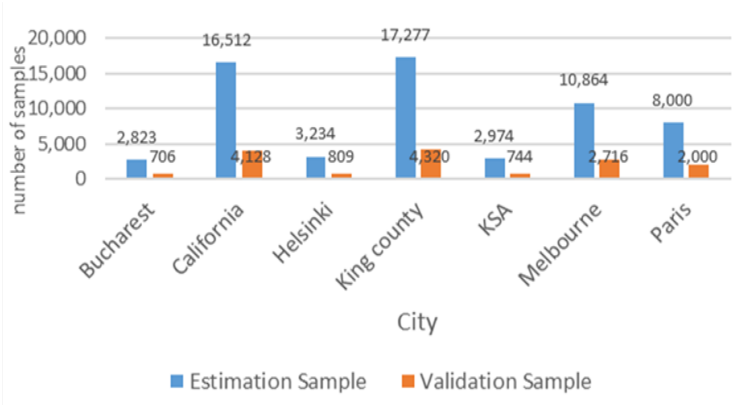

Figure. 2 Estimation and validation sample sizes (80%:20% train /test) for all seven cities' real estate datasets

on a scale of 0–100 percent as in Eq. (16), the higher the R^2 value, the better the performance, and vice versa, and the lower the MSE, and RMSE values, the better the performance, and vice versa.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - yi\prime)2}{\sum_{i=1}^{n}(yi - yi\prime)2} \quad (16)$$

These metrics are often used in real estate appraisal research [44]. Where yi the actual price of the i[th] real estate, yi' is the predicted or forecasted price by the proposed model and n is the number of examples in the dataset. The MSE and RMSE metrics are applied across the normalized projected and actual price labels since they are both sensitive to normalization scaling

4.3 Experimental system setting

The proposed models were built using *Tensorflow* [45], and various third-party Python libraries are used such as NumPy [46] used for scientific computing, Pandas [47] for data structure, and Scikit-learn for evaluation metrics, scaling, normalization, and standardization for feature values and determination and model selection. In addition, the proposed network contains two to ten hidden layers and 32 to 512 neurons. ReLU activation function was used for all the hidden layers and linear for the output layer. In addition, we set the learning rate to 0.001. Adam as the optimization function for 200 to 1000 epochs; the batch size was set to 32. The remaining best hyperparameters are by default values and the network architecture is chosen by the Bayesian optimization method as shown in Table 1. For all seven datasets, 80% of each class is used as a training set, and the rest 20%, as the test set. The seven datasets are fed into the proposed BOM-DLNN model to obtain the best real estate prediction models with minimum MSE, RMSE, and maximum R^2 in Eqs. (14) to (16).

4.4 Results and analysis

The Analysis of the quantitative measures, namely; MSE, RMSE, and R^2 in the training phase, each dataset from the seven datasets preprocessed, scaled, normalized, and fed into the DLNN model. We compared five of the selected datasets of the proposed BOM-DLNN model with conventional methods; Stepwise & SVM (SSVM) [6], ensemble model (EM) [7]. Multilayer Perceptron (MLP) [8], Catboost regression analysis (CRA) [9], and artificial neural network (ANN) [10]. MSE, RMSE, and R^2 evaluation metrics are used as performance measure metrics in comparison for BOM-DLNN versus the [6-7-8-9-10]. From Table 2, we can observe that the proposed method outperforms the traditional state-of-the-art machine learning methods on five datasets with a significant ratio. This proves that deep neural networks optimized by the Bayesian optimization method are better than other conventional methods. This proves that deep neural networks optimized by the Bayesian optimization method are better than other conventional methods. Additionally, the proposed approach outperformed the Stepwise & tuned SVM [6] and increased the accuracy in terms R^2 from 0.93 to 0.95, so the proposed model achieves higher accuracy by 2% and for MSE the error minimized from 0.05 to 0.002, and RMSE diminished from 0.2 to 0.04. Because Melbourne real estate datasets make real estate valuation by stepwise regression and tuned SVM. Stepwise regression as a dimensionality reduction method may be inconsistent or inaccurate in feature selection; in addition, it is sensitive to sample size, variable order, correlation, and significance, which can affect the performance and overfitting the model. In ensemble model (EM) [7], MSE is decreased from 0.4 to 0.003, RMSE reduced from 0.6 to 0.05, and R^2 accuracy measure improved from 0.91 to 0.96 with a 5% improvement of our proposed DLNN model, which is a significantly good value for R^2.

Table 1. The optimum network architecture of the proposed BOM-DNN for the seven selected datasets.

Criteria/dataset	Bucharest	California	Helsinki	KC	KSA	Melbourne	Paris
No of hidden layers	9	6	7	10	10	6	9
No of neurons	32-512	32-512	32-512	32-512	32-512	32-512	32-512
Total-trainable parameters	747,905	222,081	378,7	588,513	440,513	621,633	1,102,94
No of epochs	200-1000	200-1000	200-1000	200-1000	200-1000	200-1000	200-1000
Activation function	Relu	Relu	Relu	Relu	Relu	Relu	Relu
optimizer	Adam	Adam	Adam	Adam	Adam	Adam	Adam
Loss function	mse	mse	mse	mse	mse	mse	mse
No of Feature	7	14	43	21	23	23	17
Learning rate	0.001	0.001	0.001	0.001	0.001	0.001	0.001

Table 2. Performance of the traditional methods and the proposed method for the selected five real estate datasets in terms of MSE, RMSE and R^2 evaluation metrics

Source	Model	Region	MSE	RMSE	R^2
CrossRef [6]- vs DLNN-BOM	stepwise & and tuned SVM	Melbourne, Australia	0.05 Vs **0.002**	0.2 Vs **0.04**	0.93 Vs **0.95**
CrossRef [7]- vs DLNN-BOM	Ensemble Model	California-USA	0.4 Vs **0.003**	0.6 Vs **0.05**	0.91 Vs **0.96**
CrossRef [8] - vs DLNN-BOM	MLP	Helsinki-Finland	0.1 Vs **0.0002**	0.3 Vs **0.01**	0.95 Vs **0.99**
CrossRef [9] vs DLNN-BOM	Catboost regression	King County- United States	0.8 Vs **0.0005**	0.9 Vs **0.02**	0.91 Vs **0.98**
CrossRef [10] vs DLNN-BOM	ANN	Kingdom Saudi- Arabia	0.0006 Vs **0.00009**	0.02 Vs **0.009**	0.93 **Vs 0.99**

Table 3. Performance of four optimization techniques in terms of MSE, MAE, and R^2 for all seven datasets

Optimizer	Adam			RMSprop			SGD			Adadelta		
Dataset	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2	MSE	MAE	R2
Bucharest	0.001	0.02	**0.97**	0.001	0.02	**0.97**	0.007	0.06	0.84	0.005	0.05	0.88
California	0.003	0.04	0.93	0.002	0.03	**0.96**	0.009	0.06	0.84	0.01	0.07	0.80
KSA	9.8e-05	0.006	**0.99**	0.0009	0.01	0.98	0.00005	0.004	0.99	0.001	0.01	0.97
Melbourn	0.002	0.03	0.95	0.001	0.02	**0.96**	0.005	0.05	0.90	0.007	0.06	0.86
Helsinki	0.0002	0.01	**0.99**	0.0002	0.01	**0.99**	0.0009	0.02	0.97	0.001	0.03	0.95
King County	0.0005	0.01	**0.98**	0.0007	0.01	**0.98**	0.004	0.04	0.92	0.005	0.05	0.89
Paris	3.5e-05	0.003	**0.99**	0.001	0.02	**0.99**	0.001	0.02	0.95	0.02	0.09	0.86

This demonstrates superiority of DLNN compared with ensemble model in real estate price assessment in California dataset furthermore ensemble model come with computational costs, temporal complexity, interpretation challenges, overfitting and underfitting dependent on the power and complexity of the base models. For real estate price prediction in Helsinki Finland [8], our proposed optimized DLNN model outperformed the multilayer perceptron MLP model in all evaluation metrics. For MSE minimized by significant ratio from 0.1 to 0.0002 and for RMSE the error reduced from 0.3 to 0.01 in our proposed model and R^2 enhanced from 0.95 to 0.99 by 4% accuracy improvement. Furthermore, Helsinki dataset is small and contains duplicated data MLP underperformed in tiny and incorrect datasets. For [9] our proposed DLNN revealed superiority performance when compared with Catboost regression which might perform very poorly if the variables are not examined and calibrated. DLNN outperformed (CR) in all evaluation metrics for real estate price valuation in king county in USA, in terms of MSE was minimized with extremely ratio from 0.8 to 0.0005 and RMSE minimized from 0.9 to 0.02 in our proposed model. In R^2 accuracy measure enhanced by 7% percent from 0.91 to 0.98 in our proposed model. Finally, In [10] it can be noticed that with the shallow artificial neural network ANN model used in real estate price prediction in the kingdom of Saudi Arabia KSA our proposed model outperformed ANN by 6% from 0.93 to 0.99 in R^2. In MSE, the value reduced from 0.0006 to 0.00009, and in RMSE, the error minimized from 0.02 to 0.009 in our DLNN proposed model.

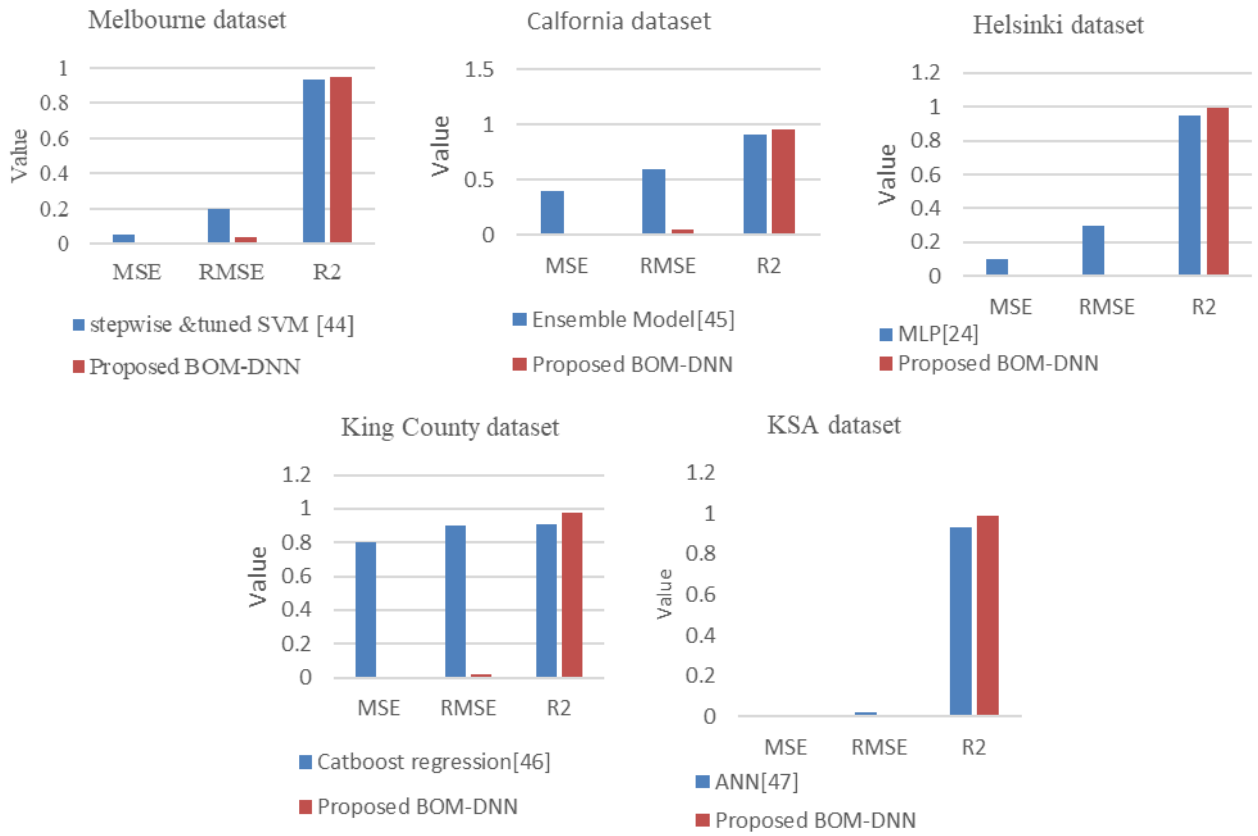

Figure. 3 Performance measures (MSE, RMSE, and R^2) comparison between the proposed approach BOM-DNN and other approaches using real datasets

Furthermore, overfitting, computationally expensive, limited interpretability, and data requirement are problems with shallow ANN networks. The performance of proposed model versus traditional models in terms of all evaluation metrics; R2, MSE, and RMSE can be shown in Fig. 3. The proposed model achieves comparable results in the real estate price prediction model when compared to benchmarked models in all datasets, which reveals the efficiency and effectiveness of the proposed framework. It could also be observed that the suggested framework produces the best outcomes when compared to the other approaches on the five datasets and generalized to Bucharest and Paris datasets and achieves 0.97 and 0.99 in R^2 respectively. The proposed framework performs much better than other techniques that were examined. Fig. 4 indicates the difference between the original real estate price and the prediction price, with the solid line and marks designating the original price and the prediction, respectively. The majority of all datasets fitted very well and approximate 100% predicted prices.

In addition, we conducted several experiments on optimization techniques to analyze the effect on performance and accuracy of each optimizer used in the deep neural networks training phase such as Stochastic Gradient descent (SGD) [48], but in the real estate dataset in our problem, SGD generalized poorly. Adaptive, optimization methods such as Adam [43] and RMSprop perform very well and are generalized fast and accurately in real estate price prediction problems, Finally, Ada-Delta [49] an adaptive learning rate optimizer is the worst optimization technique in all seven real estate datasets. We conducted several experiments to approve this claim as shown in Table 3. To attain real-time performance, however, the suggested approach still has to be made faster. It is still quite difficult to strike a balance between computational complexity and performance but in real estate, price prediction problem accuracy is the prime concern. In the future, we want to look at alternative strategies to reduce system complexity and computing strain.

5. Conclusion

This study presents an innovative approach to real estate tax base assessment that makes use of the unrealized potential of deep learning neural network, breaking new ground and opening up new possibilities. In this paper, we proposed a deep neural network optimized by the Bayesian optimization method for real estate price prediction of small and high dimensional real estate datasets for estimating

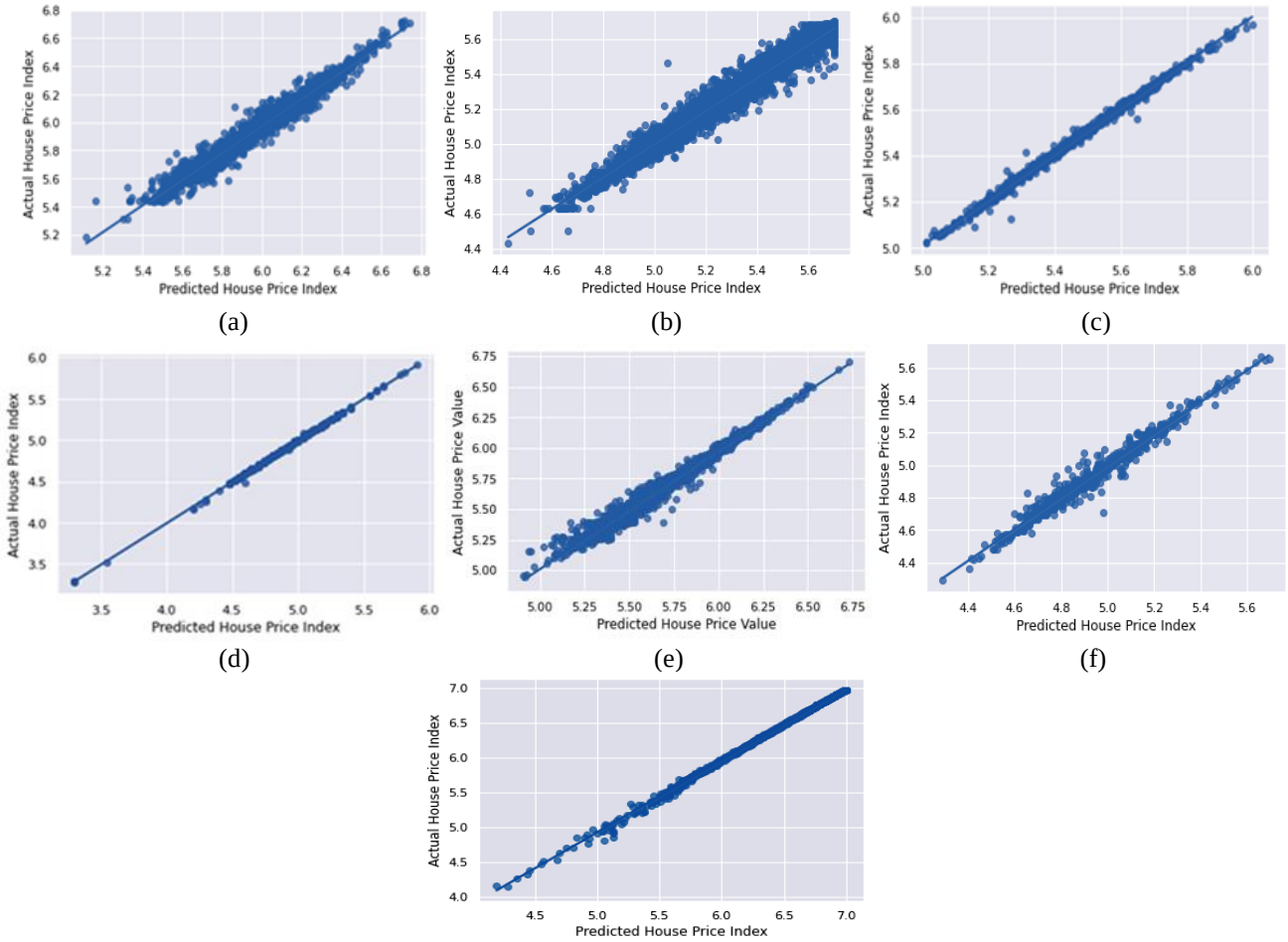

Figure. 4 Performance of the optimized model in a scatter plot of Actual (solid line) and predicted (dotted line) for all seven real estate datasets: (a) Melbourne city, (b) California city, (c) Helsinki, (d) KSA city, (e) King County city, (f) Bucharest city, and (g) Paris city

and levying real estate taxes. The proposed framework achieves a high-accuracy real estate price prediction model. Extensive experimentations were conducted on five benchmark real datasets: Melbourne, California, Helsinki, king county, and the kingdom of Saudi Arabia to validate and verify the efficiency and effectiveness of our proposed framework. The results obtained revealed that the proposed framework outperforms the compared techniques mentioned in section 4.4 and can be generalized to other real estate datasets such as Bucharest and Paris datasets. Our proposed framework achieved higher accuracy from 2% to 7% in the R^2 evaluation metric when applied to the five datasets mentioned in section 4.1; it also achieved minimum mean squared error MSE when it was applied for all five datasets.

The proposed method has significant advantages: 1) BOM automated and obtained the best hyperparameters for deep neural networks than other hyperparameter optimization techniques. 2) The proposed model achieves high performance in real estate price prediction problems when compared with other traditional approaches with feasible running time, making it ideal for a variety of applications. 3) The high-dimensional real estate pricing dataset is managed effectively by the suggested BOM-DLNN model with small and high-dimensional real estate datasets. In the future, additional studies may develop a more reliable and accurate real estate price appraisal model by using additional and different feature datasets. Also, expanding the hyperparameter optimization procedure to other kinds of deep neural networks for real estate price prediction problems and using ensemble models for hyperparameter optimizations. In addition, used hybrid dimensionality reduction models with optimized hyperparameters deep neural network. Finally, investigate time series analysis with deep neural networks for different real estate price prediction datasets for real estate tax assessment in different periods.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Amal R. Saleh, Motaz A. Elsaban; methodology, Amal R. Saleh, Motaz A. Elsaban, Mohamed M. Saleh; software, Amal R. Saleh; validation, Motaz A. Elsaban, Mohamed M. Saleh, Hisham M. AbdelSalam, formal analysis, Amal R. Saleh, Motaz A. Elsaban, Mohamed M. Saleh, data curation, Amal R. Saleh; writing-original draft preparation, Amal R. Saleh, Motaz A. Elsaban; writing— review and editing, Amal R. Saleh, Motaz A. Elsaban; Mohamed M. Saleh; visualization, Amal R. Saleh; supervision, Mohamed M. Saleh, Hisham M. AbdelSalam.

References

[1] M. H. Rafiei and H. Adeli, "A novel machine learning model for estimation of sale prices of real estate units", *Journal of Construction Engineering and Management*, Vol. 142, No. 2, p. 04015066, 2016.

[2] W. K. Ho, B. S. Tang, and S. W. Wong, "Predicting property prices with machine learning algorithms", *Journal of Property Research*, Vol. 38 No. 1, pp. 48-70, 2021.

[3] Z. Jiang and G. Shen, "Prediction of house price based on the back propagation neural network in the Keras deep learning framework", In: *Proc. of the IEEE on 6th International Conf. on Systems and Informatics (ICSAI)*, pp. 1408-1412, 2019.

[4] F. Wang, Y. Zou, H. Zhang, and H. Shi, "House price prediction approach based on deep learning and ARIMA model", In: *Proc. of the IEEE on 7th International Conf. on Computer Science and Network Technology (ICCSNT)*, pp. 303-307, 2019.

[5] M. Awad and R. Khanna, *Efficient learning machines: theories, concepts, and applications for engineers and system designers*, Apress, Berkeley, CA, p. 268, 2015.

[6] T. D. Phan, "Housing price prediction using machine learning algorithms: The case of Melbourne city, Australia", In: *Proc. of the IEEE of International Conf. on machine learning and data engineering (iCMLDE)*, pp. 35-42, 2018.

[7] Y. Huang, "Predicting home value in California, United States via machine learning modeling. Statistics", *Optimization & Information Computing*, Vol. 7, No. 1, pp. 66-74, 2019.

[8] J. Kalliola, J. K. Dzikienė, and R. Damaševičius, "Neural network hyperparameter optimization for prediction of real estate prices in Helsinki", *PeerJ Computer Science*, Vol. 7, p. e444, 2021.

[9] Y. Wang, and Q. Zhao, "House Price Prediction Based on Machine Learning: A Case of King County", In: *Proc. of 7th International Conf. on Financial Innovation and Economic Development (ICFIED)*, Atlantis Press, pp. 1547-1555, 2022.

[10] E. Alzain, A. S. Alshebami, T. H. H. Aldhyani, and S. N. Alsubari, "Application of artificial intelligence for predicting real estate prices: The case of Saudi Arabia", *Electronics*, Vol. 11, No. 21, p. 3448, 2022.

[11] R. Pal, "Chapter 4-validation methodologies", *Predictive Modeling of Drug Sensitivity*, pp. 83-107, 2017.

[12] Q. Zhang, "Housing price prediction based on multiple linear regression", *Scientific Programming*, pp. 1-9. 2021.

[13] L Hu, S He, Z Han, H. Xiao, S. Su, M. Weng, and Z. Cai, "Monitoring housing rental prices based on social media: An integrated approach of machine-learning algorithms and hedonic modeling to inform equitable housing policies", *Land Use Policy*, Vol. 82, pp. 657-673, 2019.

[14] T. C Peng and C. C Wang, "The application of machine learning approaches on real-time apartment prices in the Tokyo metropolitan area", *Social Science Japan Journal*, Vol. 25, No. 1, pp. 3-28, 2022.

[15] M. Chen, Y. Liu, D.A. Bel, and A. Singleton, "Assessing the value of user-generated images of urban surroundings for house price estimation", *Landscape and Urban Planning*, Vol. 266, p. 104486, 2022.

[16] S. Zheng and L. Yan, "Influence of policy adjustment on housing prices: An empirical analysis based on Chinese data since 2008", In: *Proc. of International Conf. on Construction and Real Estate Management, American Society of Civil Engineers Reston*, VA, pp. 1093-1106, 2016.

[17] W. Yue, C. Ni, C Tian, H. Wen, and L. Fang, "Impacts of an urban environmental event on housing prices: Evidence from the Hangzhou pesticide plant incident", *Journal of Urban Planning and Development*, Vol. 146, No. 2, p. 04020015, 2020.

[18] C. Zhang, M. Xiong, and X. Wei, "Influence of accessibility to urban service amenities on housing prices: Evidence from Beijing", *Journal of Urban Planning and Development*, Vol. 148, No. 1, p. 05021063, 2022.

[19] H. Wen, Z. Gui, C. Tian, Y. Song, and G. Zhou, "Expressway proximity effects on property prices in Hangzhou, China: Multidimensional housing submarket approach", *Journal of Urban Planning and Development*, Vol. 148, No. 1, p. 04021070, 2022.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, Vol. 521, No. 7553, pp. 436-444, 2015.

[21] L. Koumakis, "Deep learning models in genomics; are we there yet?", *Computational and Structural Biotechnology Journal*, Vol. 18, pp. 1466-1473, 2020.

[22] B. Cao and B. Yang, "Research on ensemble learning-based housing price prediction model", *Big Geospatial Data and Data Science*, Vol. 1, No. 1, pp. 1-8, 2018.

[23] H. Seya and D. Shiroi, "A comparison of residential apartment rent price predictions using a large data set: Kriging versus deep neural network", *Geographical Analysis*, Vol. 54, No. 2, pp. 239-260, 2022.

[24] Y. Piao, A. Chen, and Z. Shang, "Housing price prediction based on CNN", In: *Proc. of the IEEE on 9th International Conf. on Information Science and Technology (ICIST)*, pp. 491-495, 2019.

[25] H. Kim, Y. Kwon, and Y. Choi, "Assessing the impact of public rental housing on the housing prices in proximity: based on the regional and local level of price-prediction models using long short-term memory (LSTM)", *Sustainability*, Vol. 12, No. 18, p. 7520, 2020.

[26] J. Xu, "A novel deep neural network-based method for house price prediction", In: *Proc. of the IEEE On International Conf. of Social Computing and Digital Economy (ICSCDE)*, pp. 12-16, 2021.

[27] A. Varma, A. Sarma, S. Doshi, and R. Nair, "House price prediction using machine learning and neural networks", In: *Proc. of the IEEE On International Conf. of inventive communication and computational technologies (ICICCT)*, pp. 1936-1939, 2018.

[28] Y. Zhao, G. Chetty, and D. Tran, "Deep learning with XGBoost for real estate appraisal", In: *Proc. of the IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1396-1401, 2019.

[29] 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.

[30] F. Mostofi, V. Toğan, and H. B. Başağa, "Real-estate price prediction with deep neural network and principal component analysis", *International Journal of Organization, Technology, and Management in Construction*, Vol. 14, No. 1, pp. 2741-2759, 2022.

[31] N. Bacanin, T. Bezdan, and E. Tuba, I. M. Strumberger, "Monarch butterfly optimization based convolutional neural network design", *Mathematics*, Vol. 8, No. 6, p. 936, 2020.

[32] N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, and M. Tuba, "Optimizing convolutional neural network hyperparameters by enhanced swarm intelligence metaheuristics", *Algorithms*, Vol. 13, No. 3, p. 67, 2020.

[33] J. H. Han, D. J. Choi, S. U. Park, and S. K. Hong, "Hyperparameter optimization using a genetic algorithm considering verification time in a convolutional neural network", *Journal of Electrical Engineering & Technology*, Vol. 15, pp. 721-726, 2020.

[34] S. H. Kim, Z. W. Geem, and G. T. Han, "Hyperparameter optimization method based on harmony search algorithm to improve performance of 1D CNN human respiration pattern recognition system", *Sensors*, Vol. 20, No. 13, p. 3697, 2020.

[35] L. L. Lima, J. R. F. Junior, and M. C. Oliveira, "Toward classifying small lung nodules with hyperparameter optimization of convolutional neural networks", *Computational Intelligence*, Vol. 37, No. 4, pp. 1599-1618, 2021.

[36] H. Cui and J. Bai, "A new hyperparameters optimization method for convolutional neural networks", *Pattern Recognition Letters*, Vol. 125, pp. 828-834, 2019.

[37] D. Połap, M. Woźniak, W. Wei, and R. Damaševičius, "Multi-threaded learning control mechanism for neural networks", *Future Generation Computer Systems*, Vol. 87, pp. 16-34, 2018.

[38] D. Plonis, A. Katkevičius, A. Gurskas, V. Urbanavičius, R. Maskeliūnas, and R. Damaševičius, "Prediction of meander delay system parameters for internet-of-things devices using Pareto- optimal artificial neural network and multiple linear regression", *IEEE Access*, Vol. 8, pp. 39525-39535, 2020.

[39] M. Zhang, W. Jing, J. Lin, N. Fang, W. Wei, M. Woźniak, and R. Damaševičius, "NAS-HRIS: Automatic design and architecture search of neural network for semantic segmentation in remote sensing images", *Sensors*, Vol. 20, No. 18, p. 5292. 2020.

[40] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, "Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks", *IEEE Access*, Vol. 8, pp. 52588-52608, 2020.

[41] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints", *arXiv Preprint arXiv:1403.5607*, 2014.

[42] W. Koehrsen, "A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning", *Towards Data Science*, 2020.

[43] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014.

[44] C. Xue, Y. Ju, S. Li, Q. Zhou, and Q. Liu, "Research on accurate house price analysis by using GIS technology and transport accessibility: A case study of Xi'an, China", *Symmetry*, Vol. 12, No. 8, p. 1329, 2020.

[45] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, and J. Dean, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems", *arXiv Preprint arXiv:1603.04467*, 2016.

[46] C. R. Harris, K. J. Millman, S. J. V. D. Walt, R. Gommers, P. Virtanen, D. Cournapeau, and E. Wieser, "Array programming with NumPy", *Nature*, Vol. 585, No. 7825, pp. 357-362, 2020.

[47] W. McKinney, "Data structures for statistical computing in Python", In: *Proc. of International Conf. of the 9th Python in Science Conference*, Austin, TX, Vol. 445, No. 1, pp. 51-56, 2010.

[48] H. Robbins and S. Monro, "A stochastic approximation method", *The Annals of Mathematical Statistics*, pp. 400-407, 1951.

[49] M. D. Zeiler, "Adadelta: an adaptive learning rate method", *arXiv Preprint arXiv:1212.5701*, 2012.