IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Real estate price forecasting utilizing recurrent neural networks incorporating genetic algorithms

Ting Tin Tin[1*], Cheok Jia Wei[2], Ong Tzi Min[3], Boo Zheng Feng[4], Too Chin Xian[5]

[1]*Faculty of Data Science and Information Technology, INTI International University, 71800 Nilai, Negeri Sembilan, Malaysia.*
[2,3,4,5]*Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology, 53300 Kuala Lumpur, Malaysia.*

Corresponding author: Ting Tin Tin (*Email: tintin.ting@newinti.edu.my*)

Abstract

This study aims to develop and examine the effectiveness of the Recurrent Neural Network (RNN) model incorporating Genetic Algorithm (GA) in forecasting real estate prices. Real estate prices have a significant impact on a country's financial system. Therefore, the ability to accurately forecast its price is valuable. A set of data containing 5.4 million unique records of real estate with their prices is used in the study. The data set, which spans from 2018 to 2021, contains twelve independent variables and one dependent variable. We preprocessed the data set to reduce noise and outliers that could potentially lead to poor model performance. The RNN model was selected because (GA) optimises the hyperparameters in the Recurrent Neural Network (RNN) hidden layer to optimise the performance of the RNN model. Relative Root Mean Squared Error (RRMSE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Absolute Error (MAE) examine the effectiveness of the RNN-GA model.The results show that the RNN-GA outperforms RNN and the traditional statistical models used in the real estate industry. This study provides an understanding of the effectiveness of incorporating GA in optimizing the RNN model for real estate price forecasting, which could benefit stakeholders in the real estate industry and sustain the financial system. This study uses a novel hybrid RNN-GA model for predicting housing prices with a large dataset.

Keywords: Artificial intelligence, Economic development, Genetic algorithm, Machine learning, Real estate forecast, Recurrent neural network.

DOI: 10.53894/ijirss.v7i3.3220
Funding: This work is supported by the INTI International University, Malaysia (Grant number: T&E3877)
History: Received: 23 October 2023/**Revised:** 6 March 2024/**Accepted:** 28 March 2024/**Published:** 30 May 2024

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Institutional Review Board Statement: The Ethical Committee of the INTI International University, Malaysia has granted approval for this study.
Publisher: Innovative Research Publishing

1. Introduction

Real estate prices have been a major concern for people and real estate marketers, especially in emerging countries. Since banks have been using real estate as deposits and lending mortgages to people. This resulted in real estate prices exerting a significant influence on the financial system [1]. This made the ability to forecast real estate prices valuable. Traditional forecasting models have difficulties in accurately forecasting real estate prices due to political, social, and other factors [2]. On the other hand, the rising trend of technology also causes terms and concepts like Artificial Intelligence and Machine Learning to be more widespread and commonly known [3]. Therefore, statistical models like RNN can be used to assist in forecasting real estate prices by using a large real estate dataset. Other than that, real estate is considered the largest asset class. It is important to take note of this because the movement of real estate prices affects the whole financial system because people tend to lend mortgages from the bank and use real estate as collateral. Therefore, it is best to be able to forecast the price of real estate to be able to differentiate between better or worse deals and identify the risk of investing in real estate [4].

This study focuses mainly on developing and evaluating the performance and effectiveness of the RNN model that incorporates GA in forecasting the price of real estate. Furthermore, this study aims to also build a statistical model to estimate the real estate's price employing Recurrent Neural Network incorporating Genetic Algorithm. In addition to that, this study aims to train the statistical model using various methods such as Exploratory Data Analysis, Train Test Split, and Feature Engineering. Lastly, this study will also test the statistical model using RRMSE, RMSE, MAE, and MAPE.

2. Literature Review

Previous studies have tried to use machine learning (ML) algorithms and techniques to forecast real estate prices. For instance, Random Forest(RF), Levenberg-Marquardt (LM), Linear Regression (LR), Decision Tree (DT), Gradient Boosting Machine (GBM), hybrid method of Autoregressive Integrated Moving Average (ARIMA), Ensemble Empirical Mode Decomposition (EEMD), Long Short-Term Memory (LSTM), Long short-term memory with modified Genetic Algorithm (GA-LSTM), Group Method of Data Handling (GMDH), Neural Network (NN), Least Squares Support Vector Regression (LSSVR), Multiple Regression Model (MRA) and Artificial Neural Network (ANN), Light Gradient Boosting Machine regression (Light GBM), Modified Holt's Exponential Smoothing with Whale Optimization Algorithm (WOA-MHES), Autoregressive Integrated Moving Average (ARIMA), and Particle Swarm Optimization Bagging Artificial Neural Network System (PSO-Bagging ANNs) [2, 4-17].

D'Acci [18] states that there are several factors affecting the real estate price, including internal factors and external factors. The internal factors include building specifications such as the exterior appearance of a building, interior aesthetics, physical structures of properties, and so forth, whereas the external factors are the properties of the surrounding areas which consists of the public transportation, distance to Central Business District, as well as the area's quality such as quality of the urban environment, greenery, community surrounding, and shops. The author concludes that the quality of urban areas determines property values. In other words, when there is an improvement in the quality of urban areas, the value of the property will increase accordingly. The Turin study demonstrates that a 1% improvement in site quality leads to a 0.58% increase in house value and a 142% increase in property value, even with minor changes in the neighborhood [18]. Also, a previous study supports the notion that structural characteristics, ease of access, and service amenities impact the real estate market in Shanghai, resulting in a pattern of concentric rings [19].

On the other hand, several variables can have a notable impact on real estate prices in Shenzhen, China [20]. These variables include the distance between real estate and the city center, greens around the apartment, density of population, estate management fee, and level of economics. Additionally, we should amend several traditional factors associated with the educational sector to align with the current state of the real estate market. Subsequently, their research consists of the process of integrating the XGBoost and Hedonic Price Model (HPM) in specific to produce a practical output and better understand the relationships between real estate prices and influential factors [20]. These findings are relatively crucial for policymakers and developers who wish to implement projects consisting of fair housing policies and comfortable neighborhoods.

As previously mentioned, RF is a widely used traditional method for forecasting real estate prices. The RF algorithm can forecast since it uses decision trees as a regression technique. Until the conditions are satisfied, the decision tree divides the input into two or more child nodes. Each of these nodes then forecasts the values, combining and averaging the results [21]. One of the previous studies conducted by Tchuente and Nyawa [4] was able to find that the RF algorithm and the neural network tend to perform better in forecasting the price of the real estate of French cities than other methods that do not take geographical coordinates into account, while algorithms such as RF, Gradient Boosting (GB), and Ada boost (AB) tend to perform better when geographical coordinates are taken into account. As a result, RF was able to obtain a value of 47,992 in RMSE loss and 30,225 in MAE loss. However, there was a limitation in RF from the study in that hyperparameter settings are very confusing, unlike models such as neural networks, where the hyperparameters are more direct and easier to understand [22].

Furthermore, earlier research showed that ML uses self-evaluating algorithms to try to forecast the future trends or behaviors by learning historical data. A similar theory underlies the prediction of house prices in Pakistan [6]. Pakistanis often have difficulty in computing property prices, and external factors must be considered to refine the prediction's accuracy. To address this, ANN and LR, as ML algorithms, have been successfully applied to housing price prediction. In their study, they have used three ML algorithms, DT, RF, and LR, to predict the price of real estate using the Defence Housing Authority (DHA), Karachi Defence data set in real time. We can evaluate the efficiency and quality of these regression models by calculating the MAE. Comparing the results of the three models, RF outperformed the best results, both in terms of a high accuracy rate of 98.08%and a MAE value of 3799622.3401656877. In short, the developer intends to use the results generated by the house prediction system for two main purposes: firstly, to forecast the appropriate selling price of real estate; and secondly, to assist buyers or sellers in purchasing a house that fits within their budget constraints. Moreover, another study conducted by Xu and Zhang [17] used the LM algorithm with 3 hidden neurons and a training ratio, validation ratio, and testing ratio of 70%, 15%, and 15%, respectively for forecasting real estate prices for one hundred cities in China. The LM algorithm works very similarly to the ANN, where it acts as a supervised learning algorithm that can be implemented in all kinds of situations [5]. They were able to provide accurate forecast results for real estate prices using the model with an average RRMSE score of 0.98, 1.01, and 1.00 for the data set used.

Other than that, Ho, et al. [7] utilize several ML algorithms, namely RF, Support Vector Machine (SVM), and GBM, to forecast house prices. When comparing these three algorithms, GBM and RF perform better than SVM since both algorithms could generate the house price estimate more accurately and with a lower forecast error. To forecast the price of homes in Hong Kong, the study used linear and non-linear ML techniques with 39,554 housing transaction records from June 1996 to August 2014. The size of the property, its geographic location, and the features it offers are among the critical factors that can affect the price [9]. The performance metrics associated with RF are 0.00795 (MSE), 0.08918 (RMSE), 0.32270% (MAPE), and GBM which is 0.00793 (MSE), 0.08903 (RMSE), and 0.32251% (MAPE) also unambiguously outperform those of SVM, 0.01422 (MSE), 0.11925 (RMSE),and0.54467% (MAPE). Although SVM did not perform well in this forecasting, the study showed it is still a valuable algorithm for fitting data since it can generate reasonably precise predictions.

Li, et al. [19] conducted a study that developed a multiscale analysis paradigm using EEMD and compared it with models like SVR, ARIMA, and the polynomial function to examine the effectiveness of the proposed model in forecasting house prices under the effects of extreme events like changes in mortgage policy, adjustments to the down payment ratio, and many more. To generate short-term forecasts, the ARIMA model is highly effective and consistently performs better than complex structural models [23]. The study uses the dataset from 2005 to 2018, which is up to 156 months, to predict housing prices in Beijing. The results showed that the proposed model had the lowest error rate, which is 3174 RMSE and 5,62% MAPE. Despite outperforming other models, the study aims to enhance and broaden the understanding of the impact of specific local regulations on house prices [8].

Aside from that, Liu and Liu [9] proposed GA-LSTM to tackle the traditional models' issues, which are ineffective in tackling nonlinear problems and have severe limitations on input variables. This modified GA allows feature selection and can optimize hyper-parameters of model, such as the number of hidden layers (NHL). The authors of this study incorporated the Shenzhen real estate data set for 2010 to 2017 with the chosen model and found that the GA-LSTM approach performs well in terms of precision in forecasting housing prices. This study also conducted the proposed model using various NHLs and various numbers of units in each hidden layer (NUHL) to evaluate its performance. The proposed model for this study was able to obtain an evaluation score of 41 (RMSE), 40 (MAE), and 0.06% (MAPE) with the hyperparameter setting of 3 hidden layers and with Monika [14]; Nazemi and Rafiean [10] and Ge, et al. [13] neurones in their respective hidden layers in the proposed model. However, according to Liu and Liu [9] this model took a long time to deploy, although it produced good results and may lead to weak performance when there is a small dataset. This can be proven by a previous study, which stated that when NHL is higher, the longer the time taken to train the neural network [24]. In the ANN, model performances are greatly influenced by hidden layers, where accuracy and time complexity are the key limitations in the case of complex issues.

Nazemi and Rafiean [10] proposed the deployment of a GMDH to forecast the price trend in the housing market of Isfahan City in Iran. Data from several boroughs of Isfahan City was collected every six months from 1995 to 2019 to forecast the price of housing. The authors took Isfahan City as an example and obtained a result of 3.26% for MAPE, 4.98 for RMSE, and 24.84 for MAE. Despite the seemingly acceptable results, the authors note that they only selected a limited number of factors due to the absence of reliable data on the influence features, and they believe that incorporating more variables in the future could enhance the model's accuracy.

Apart from that, real estate price prediction has always been inevitable in the modern world, which is why real estate agents often rely on the results given to make informative decisions. The objective of this research is to forecast the values of houses using existing historical data and various ML models. This research has included general regression neural networks, back propagation neural network, LSSVR, and classification and regression trees. Based on the numerical results of MAPE, the LSSVR model performed relatively better than the rest of the ML models [11]. The average MAPE values of LSSVR with selected attributes and the LSSVR models without attribute selection obtained in this research are 0.228% and 1.676%, respectively Xu [11].

Pai and Wang [12] constructed a forecasting model used to forecast real estate auction prices with MRA and ANN in conjunction with each other. In this study, the researchers extract data from Ghana apartment auctions that fall between the time range of 2016 and2020 to initiate the evaluation of the forecasting models. The ANN model was stated to have established a relatively better performance compared to Multiple Regression Analysis and efficient zonal segmentation in terms of the evaluation of auction prices. The MAPE values obtained by both forecasting models are 15.11% and 16.55%, respectively, which indicates the advantage of ANN over MRA. On the other hand, the RMSE values scored by MRA and ANN have been recorded at 0.0042 and 0.0044, respectively [12].

In retrospect, the Python programming language is involved in this research for various regression methods of supervised machine learning. In doing so, future house prices are determined based on independent variables solely relying on the calculation of error value. We have used algorithms like Light GBM, RF, SVR, and XGBoost in this study to forecast real estate values. We will use the predicted result to evaluate the different algorithms based on their accuracy. The examination results eventually indicate that the Light GBM method obtains the highest accuracy among the rest according to performance score of 0.906766, 1.100254e+10 MSE, and 119034.779432 RMSE [13].

Other than that, the MLR models and neural networks are the prioritized machine learning models used in this research to forecast housing prices, fed with a dataset of second-hand real estate data in Shenzhen, China. The author obtains a result that indicates that the neural network is relatively better. Additionally, this study claims that consumers and real estate professionals intuitive understand the key factors that significantly influence housing prices through the MLR. The result of this study shows that MLR obtains slightly lower goodness of fit value compared to the neural network, whereas the neural network gains an advantage over MLR when it comes to mean square error [10].

Liu and Wu [15] developed a hybrid model to predict real estate price trends in various Chinese cities. The authors introduced a modified Holt's exponential smoothing (MHES) method that uses historical data to forecast housing prices, which has been proven to produce superior forecast performance [14]. To further enhance the accuracy of their predictions, the authors integrated a whale optimisation algorithm (WOA) into their model. The author used RMSE, MAPE, and SMAPE to evaluate the performance of the model in this study. The WOA-MHES model yielded the best RMSE of 138.8579, while the grey model (GM), ARIMA, and backpropagation neural network (BPNN) resulted in RMSE values of 599.3756, 303.5995, and 1041.0500, respectively [14]. The WOA-MHES model also produced the best MAPE result among the four models at 0.79%.The WOA-MHES models appear to exhibit a lower level of power prediction error and demand less computational time compared to conventional models, making them the most appropriate models for implementation in this study [25]. The WOA-MHES model was effective in analyzing various data characteristics, including the level and trend, which resulted in its superior prediction performance. The study used real estate price data from four cities with different economic statuses and provided a valuable tool to study China's real estate prices and formulate housing policies [14]. In future studies, the authors plan to incorporate additional factors and further optimize the MHES method.

In the next study, Wang, et al. [2] used a deep learning model to predict real estate prices using TensorFlow. The model utilized the Adam optimizer and the Rectified Linear Unit (ReLU) activation function, whereas the ARIMA was used to forecast real estate price trends [2]. ARIMA is a forecasting method that combines three statistical models: AR, MA, and ARIMA. ARIMA model is a useful tool to characterize time series data. The reason is that this model provides an autocorrelation function, which allows it to capture the stochastic behaviour of the data and uncover important information effectively, such as trends, random fluctuations, cyclic patterns, periodic components, and serial correlation [26]. The researchers looked at how well the model worked by using RMSE and MRE metrics and compared the outcomes to those found with a support vector regression (SVR) model [2]. The ARIMA model exhibited superior training and testing performance relative to the SVR model, with RMSE values of 5393 and 7671, and MRE values of 0.19 and 0.22 for the train and test phases, respectively. In contrast, the SVR model produced RMSE values, 9024 and 10216, and MRE values of 0.23 and 0.25 for the training and testing phases, respectively. In general, the results have shown that ARIMA is more effective in predicting house prices than the SVR model and that the predicted trends are consistent with the actual market conditions Wang, et al. [2].

Chou, et al. [16] provide a concise overview of ML methods for house price prediction in their most recent work. The authors decided to use the Taiwan data set that represents the price of the housing transaction in Taipei City [15]. In this research, the effectiveness of four different AI house price prediction techniques, such as ANN, SVR, Classification and Regression Trees (CART), and LR, is evaluated. The authors used metrics such as the Pearson's correlation coefficient, MAPE, RMSE, MAE, and the Synthesis Index (SI) to evaluate the effectiveness of the model [15]. The model achieved an R value of 0.970, an RMSE 3,390,016, MAE of 2,273,866, and an MAPE of 11.59%. Research demonstrates that the PSO-Bagging ANNs model out performed other models [15]. The authors provide users with a range of prediction models to choose from, enabling them to select the one that best fits their needs. This approach allows users to make an informed decision about which model to use based on their specific requirements.

3. Research Methodology

3.1. Data Preprocessing

The dataset that this study has chosen represents a list of Russian Real Estate Attributes with their prices. It contains 5.4 million unique records, and it was taken from Kaggle, a platform where it consistently hosts various types of forecasting competitions. It contains data sets provided by various data scientists. The dataset that this study has chosen is located in Russia which can be seen in Figure 1 which contains all the data plotted on map using the longitude and longitude in each real estate record in the dataset using matplotlib and it contains 12 independent variables which publish data, publish time, latitude, longitude, region, type of building, type of object, story level, apartment level, number of living rooms, estate area, and kitchen area and 1 dependent variable, the real estate price. We will use data and time columns as indicators of time series data Table 1. All these variables will be used as they provide significant details about real estate, which will provide better insight into the real estate. The date and time columns will be used as an indicator of time series data. Furthermore, this data set has the coverage of real estate data from July 31, 2018 until April 30, 2021. The dataset is cleaned and transformed based on Figure 2 workflow.

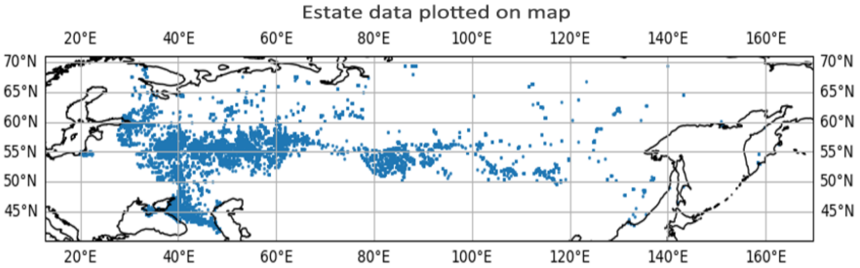

Figure 1.
Estate record in a dataset plotted on a map using matplotlib.

Table 1.
List of variable in data chosen.

Feature name	**Description**	**Type of data**
Publish date	Date of publication of the real estate announcement	Object
Publish time	Time of the real estate announcement published	Object
Latitude	Latitude coordinates	Float64 (Numeric)
Longitude	Longitude coordinate	Float64 (Numeric)
Region	Region in Russia (State). Contains 85 subjects.	Int64 (Categorical)
Type of building	Type of fascia. Panel type, monolithic type, brick type, blocky type, wooden type and others	Int64 (Categorical)
Type of object	Type of apartment, secondary real estate market and new building	Int64 (Categorical)
Apartment level	Floor of apartment	Int64 (Categorical)
Storey level	Number of storeys	Int64 (Categorical)
Living room number	Number of living rooms. -1 = Studio apartment	Int64 (Categorical)
Estate area	The total area of the apartment	Float64 (Numerical)
Kitchen area	Kitchen area	Float64 (Numerical)
Estate price	Price of real estate (Rubles)	Float64 (Numerical)

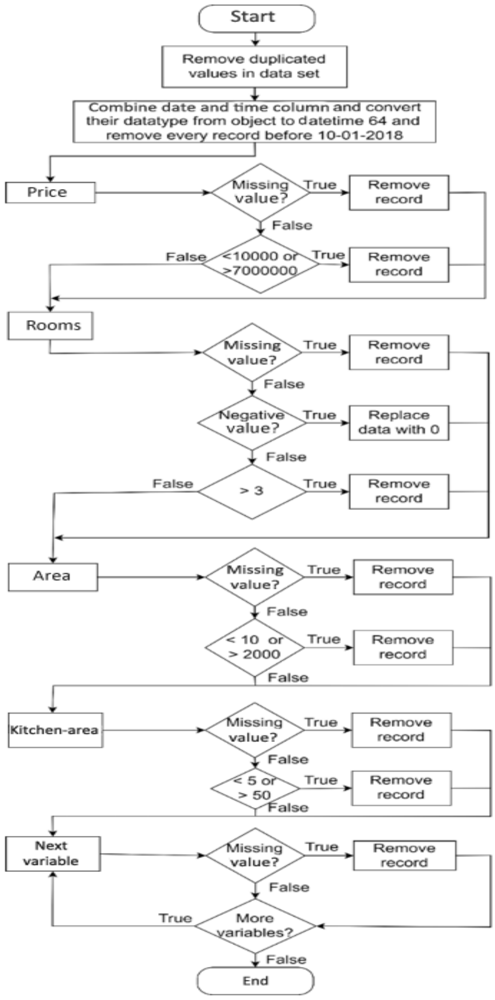

Figure 2.
Flow chart of data cleaning and processing.

3.2. RNN

RNN is one of the ANN models designed to work with sequential data, such as time series, speech, or text data. The RNN contains a recurrent connection that enables them to store information from previous inputs and utilize it to have an impact on the processing of current inputs, unlike those traditional neural networks that only process inputs in a single forward pass. This is clearly because the context of earlier inputs is crucial for comprehending the current input. In addition, they are especially well suited for tasks such as language modelling, speech recognition, and machine translation. An RNN's basic structure consists of a series of recurrent cells; each updating its internal state with each new input. Each time stepfeeds the cell with the input and its previous internal state, while the cell sends its output to the next cell in the sequence. This allows the network to maintain a "memory" of the previous inputs and use it to influence the processing of future inputs. LSTM networks and Gated Recurrent Units (GRUs)are two types of RNNs. These have extra features that help control the flow of information through the model and stop problem like vanishing gradients. Therefore, RNN was chosen because the data set used in this study contains real estate attributes, dates, and prices.

3.3. RNN Mathematical Formula

Recurrent neural networks are frequently trained using the Backpropagation Through Time (BPTT) learning process (RNNs) [27]. By unrolling the network over time, BPTT enables gradients to flow backward from the network's output to its input. Figure 3 shows the steps in the process, which are explained as follows:

1) Forward pass: The input sequence is fed into the network as it is unrolled over time, creating a series of hidden states and output predictions.

2) Error computation: For each iteration, the discrepancy between the projected output and the desired output is calculated.

3) Backward pass: The chain rule of differentiation and backpropagation over time is used to calculate the gradients of the error concerning the network parameters.

4) Parameter update: Using an optimisation approach, such as stochastic gradient descent, the network's parameters are adjusted (SGD).

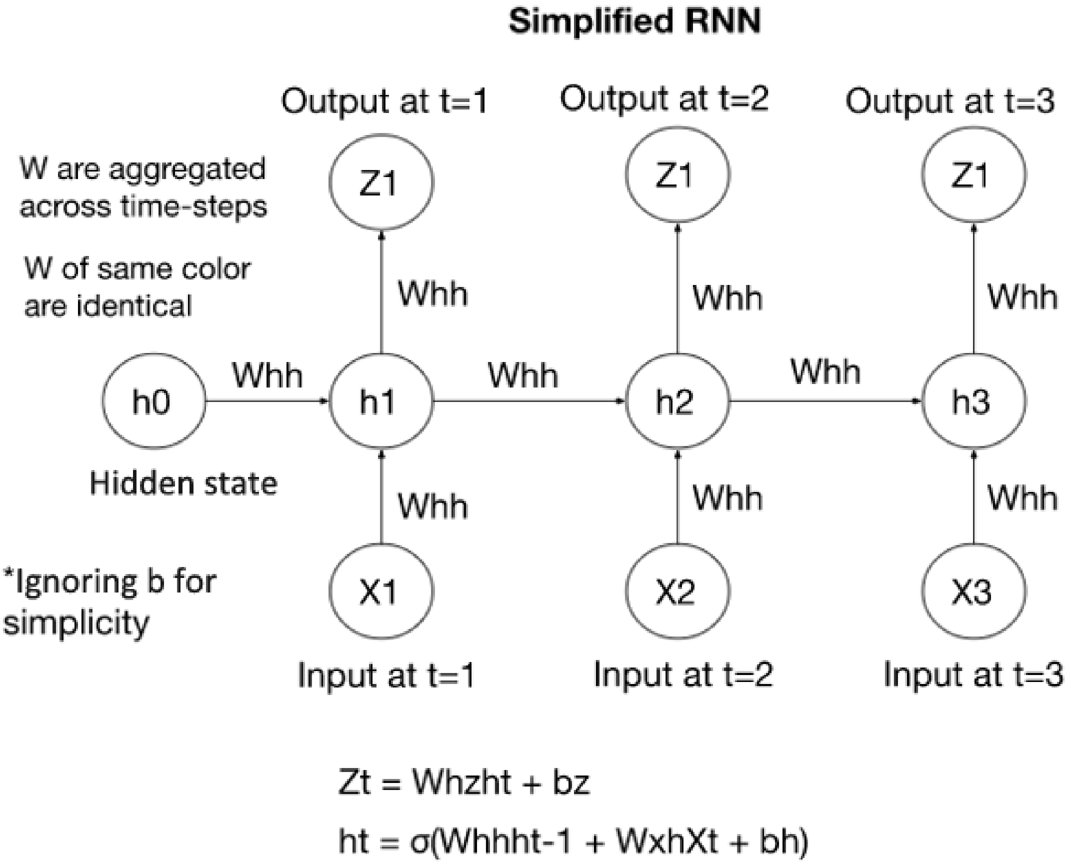

Figure 3.
Workflow and formula of RNN.
Source: Medsker and Jain [27]

The disappearing and exploding gradient problem, which can make it challenging to train RNNs over lengthy sequences, can affect BPTT. Several approaches have been put forward to address this issue, including gradient clipping, and the use of LSTM or GRUs in place of conventional RNN cells, and gradient clipping.

The first equation in Figure 3, represents the output of iteration t as Zt in the first equation. We calculate it as a bias term bz plus the multiplication of the weight matrix Whz and the value of the hidden state in iteration ht-1.

Following the second equation, the value of the hidden state at iteration t is represented as ht, and is calculated as the elementwise multiplication of the output of the activation function σ and the sum of two terms: the product of the dots between the weight matrix Wxh and the input at the current time step xt, and the product of the dots between the weight matrix Whh and the hidden state at the previous time step ht-1.

The second equation updates the hidden state at a given time step based on the current input and the previous hidden state in the second equation, while the first equation essentially calculates the RNN's output at a given time step based on the hidden state at the previous time step.Combined, these equations give the RNN the ability to handle sequential input by identifying temporal dependencies.

This study adopts RNN with improvements based on Figure 4.An input layer, typically a vector representation of the sequence, first processes the input sequence.Each time step combines the current input vector (house prices) with the previous hidden state. The result is passed through an activation function to produce the new hidden state. The hidden state is then used to produce an output value, which is passed through an activation function to produce the predicted price for the next time step. Additionally, we train the network using a regression loss function like mean squared error, which gauges the discrepancy between the predicted and actual prices. The weights in the network are updated using backpropagation through time (BPTT), where the gradients are computed for each time step and accumulated over the entire sequence.

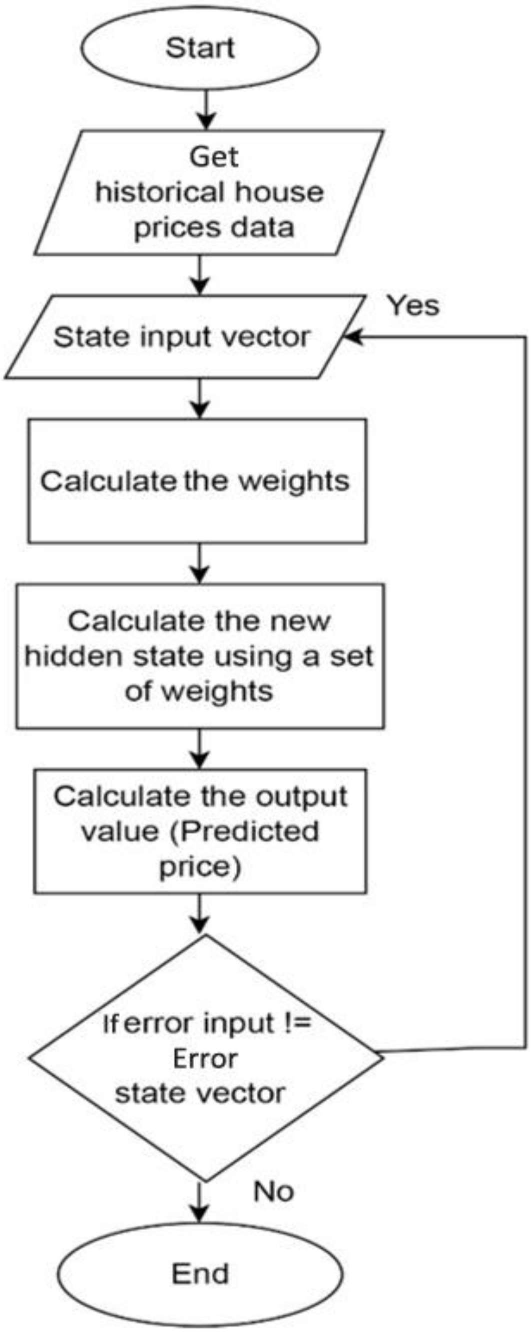
Start
Get historical house prices data
Yes
State input vector
Calculate the weights
Calculate the new hidden state using a set of weights
Calculate the output value (Predicted price)
If error input != Error state vector
No
End

Figure 4.
Flowchart of RNN used in this study.

3.4. Genetic Algorithm

Genetic Algorithm (GA) is a search heuristic [28]. It tends to mimic the process of natural selection, where the concept of elitism comes to life. Only the best of the best remains and that is the final solution. It is made possible by using population-based evolution. Usually we use GA to refine solutions with improved parameter settings. For example, as a metaheuristic function,GA can potentially optimise NUHL, but this is not a given [29, 30].

GA optimizes problems using an iterative approach, as shown in Figure 5, where the few main steps can be classified into: initializing the generation of new formulas, Calculation of Fitness of each formula, Selection process, Crossover process, and Mutation process. The GA process repeats itself until it meets the optimal solution criteria. Therefore, we can utilize GA to optimize NUHL in our RNN. This will provide the near-best possible solution for RNN for forecasting the real estate price.

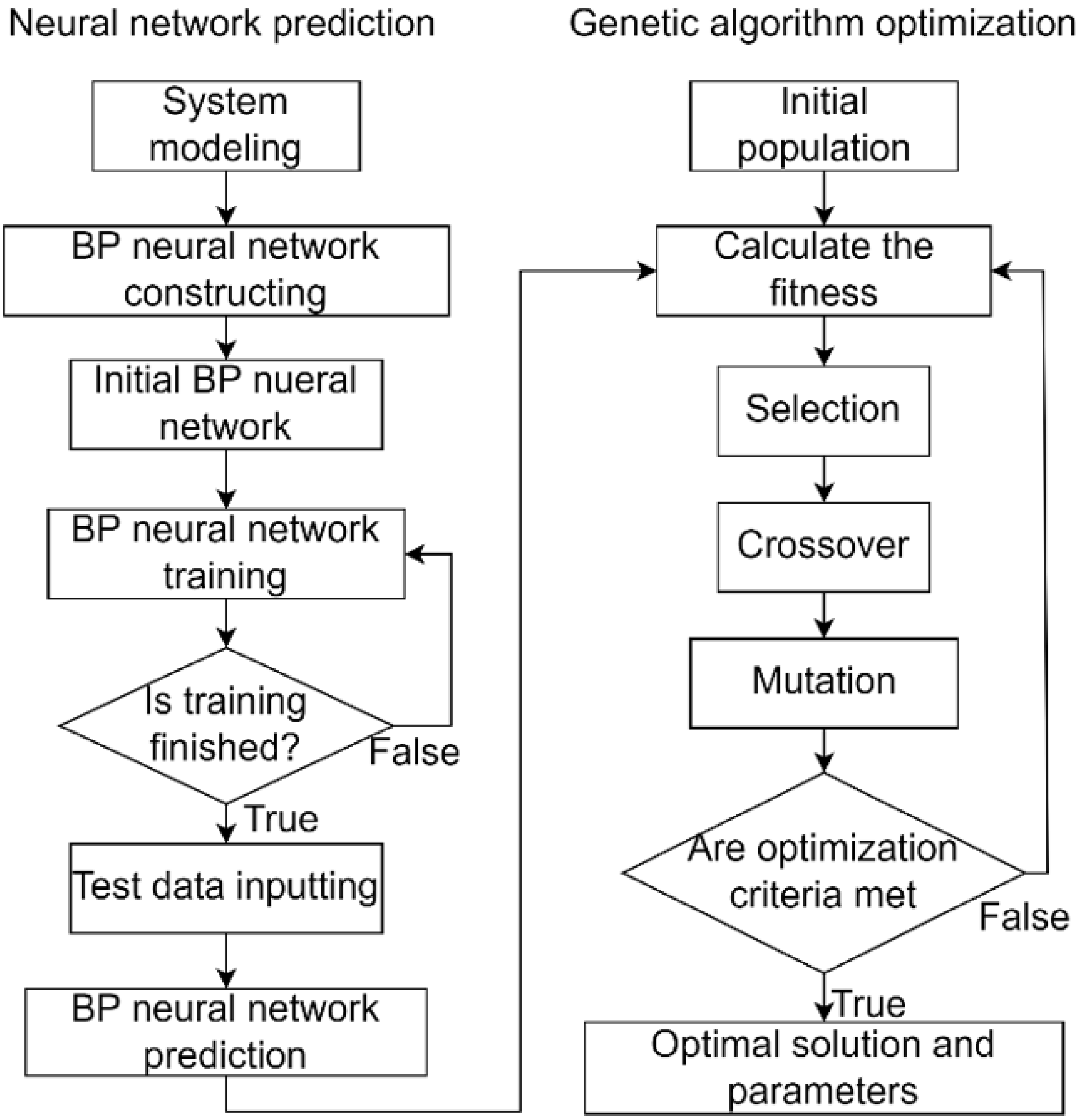
Neural network prediction
Genetic algorithm optimization
System modeling
BP neural network constructing
Initial BP nueral network
BP neural network training
Is training finished?
False
True
Test data inputting
BP neural network prediction
Initial population
Calculate the fitness
Selection
Crossover
Mutation
Are optimization criteria met
False
True
Optimal solution and parameters

Figure 5.
Flowchart of genetic algorithms.
Source: Ding, et al. [28].

4. Result and Discussion

The proposed model was conducted using Google Colab, and 1,288,430 samples were taken from the Kaggle platform. We used 1,030,744 samples as training data and the remaining 257,686 samples as testing data. The test size was set to 0.2, andthe dataset was split into 20% of the testing set and 80% of the training set to avoid overfitting or underfitting. Also, the appropriate NUHL has been determined using Genetic Algorithm as this algorithm optimises the hyperparameters.

The proposed model implemented various numbers of hidden layers (NHL) with NUHL and obtained the results in Table 2 and Table 3. It is obvious that the result of the training data, which consists of MAE, MAPE, RMSE, and RRMSE, strikes the lowest figure when there are three hidden layers and 9 neurons in each hidden layer in Table 2. The result of the validation data, which can be seen in Figure 6, retrieved a lower value than the training data, proving that there is no overfitting, and that the model performs well when it can identify unseen data. In this case, underfitting will not occur because there is a large amount of data to train the model. Table 2 also displaysthe comparison of models used by various studies.

Table 2.
GA-RNN performance of GA-RNN in training data.

NHL	NUHL	MAE	MAPE (%)	RMSE	RRMSE
1	9	0.652	29.844	0.896	0.128
2	10, 10	0.578	27.109	0.828	0.119
3	9,9,9	0.600	26.307	0.786	0.113
4	2,2,2,2	0.731	34.251	1.018	0.146
5	2,2,2,2,2	1.176	50.551	1.514	0.217
10	12,1,1,1,1,1,1,1,1,1	1.176	50.562	1.514	0.217

Table 3.
Performance of GA-RNN in validation (VAL) data.

NHL	NUHL	VAL MAE	VAL MAPE (%)	VAL RMSE	VAL RRMSE
1	9	0.649	29.660	0.893	0.128
2	10, 10	0.584	26.043	0.834	0.120
3	9,9,9	0.558	26.094	0.781	0.112
4	2,2,2,2	0.730	34.218	1.013	0.145
5	2,2,2,2,2	1.176	50.546	1.414	0.202
10	12,1,1,1,1,1,1,1,1,1,1	1.176	50.459	1.516	0.217

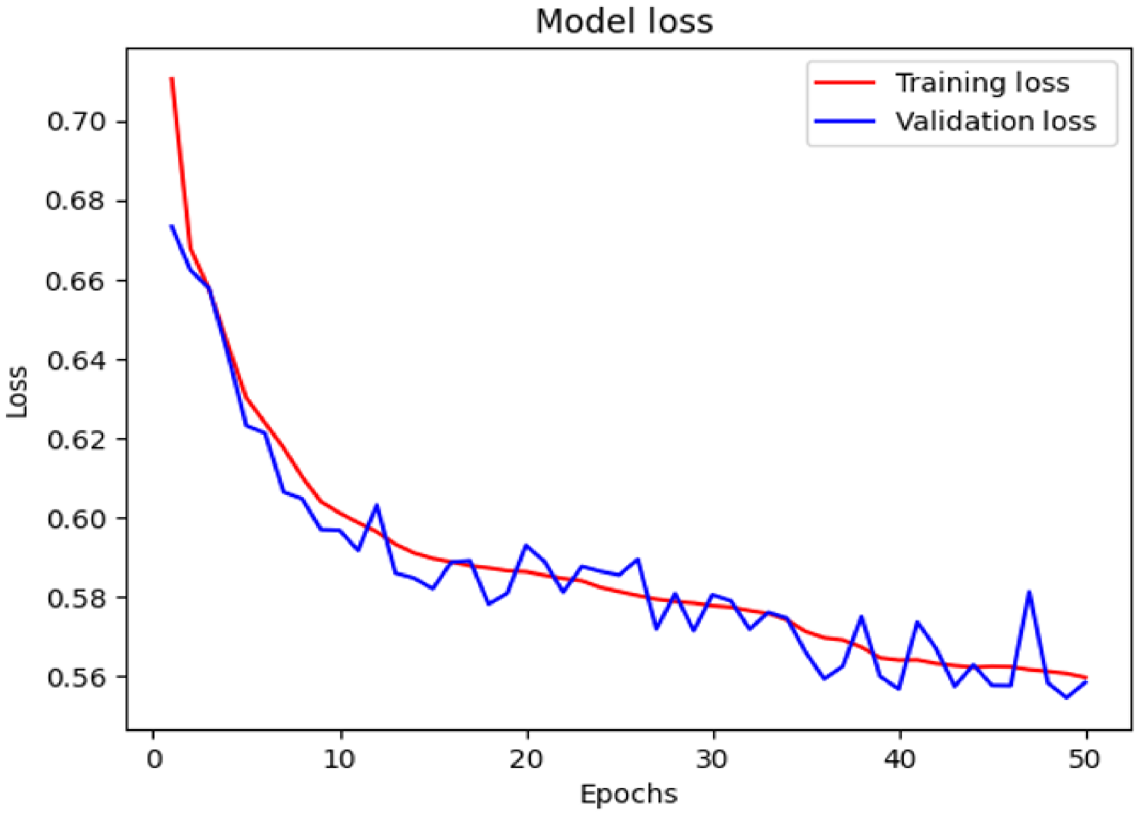

Figure 6.
Loss curve of RNN-GA modelwith 3 NHL.

Table 4.
Comparison of RMSE, MAE, MAPE, RRMSE if each model proposed by different studies.

Study	Model used	RMSE	MAE	MAPE (%)	RRMSE
Tchuente and Nyawa [4]	RF	47992	30225	-	-
Gokalani, et al. [6]	RF	-	379622.34	-	-
Xu and Zhang [17]	LM	-	-	-	0.98
Ho, et al. [7]	RF	0.089	-	0.323	-
Li, et al. [8]	EEMD	3174	-	5.62	-
Liu and Liu [9]	GA-LSTM	41	40	0.06	-
Nazemi and Rafiean [10]	GMDH	4.98	24.84	3.26	-
Pai and Wang [12]	LSSVR	-	-	0.228	-
Ge, et al. [13]	ANN	0.004	-	15.11	-
Monika [14]	Light GBM	104893.009	-	-	-
Xu [11]	NN	2.041	-	-	-
Liu and Wu [15]	WOA-MHES	138.858	-	0.79	-
Wang, et al. [2]	ARIMA	7671	-	-	-
Chou, et al. [16]	POS-Bagging ANN	3390016	2273866	11.59	-
Li, et al. [20]	XGBoost	0.057	0.039	0.825	-

Note: POS = Part-of-speech.

Due to the proposed model's normalization to the range from 0 to 1 to reduce noise, the MAE in Table 2 was lower than the results in Table 4.Therefore, the real value of MAE was $558,400 when NHL is three and NUHL is 9. In other words, the difference between actual housing prices and predicted housing prices was estimated to be approximately $558,400. It can be see that there is an enormous difference between the MAPE values in Table 2 and Table 4 because the combination of characteristics in the data set and the model chosen for each study is different. In this study, the dataset consists of continuous data, but it is not a time series dataset. Hence, the result may not seem ideal in this case and lead to an enormous difference between actual and forecasted housing prices.

Furthermore, due to the incorporation of a genetic algorithm with a recurrent neural network, it will automatically optimise the best numbers of neurons in each hidden layer for the proposed model. As a result, less time is required for the implementation of the proposed model. However, this research displayed a bad outcome as the dataset chosen was not a time series dataset but was implemented using RNN, which is a time series model. Also, the features used, such as the geographic latitude and longitude, should be processed before being adopted into our proposed model; otherwise, it may lead to inaccurate results. Other than that, there is another limitation for this study where the feature chosen for this dataset only included building specifications, but other studies included factors surrounding real estate such as public transportation, quality of the area, population density, distance to the city, green view index, and so on, and this proves that there are other important variables that would affect the housing price.

5. Conclusion

In this study, GA was used to optimize the NUHL in RNN to forecast the real estate price in Russia. The publish date, publish time, latitude, longitude, region, type of building, type of object, apartment level, story levels, number of living rooms, estate area, and kitchen area were used as input parameters, and real estate price was used as an output parameter in the models. Based on the historical data used to forecast real estate prices using the proposed model, the best results obtained from our model are MAE = 0.5597%, MAPE = 26.3066%, RMSE = 0.8925, RRMSE = 0.1127, VAL MAE = 0.5584, VAL MAPE = 26.0943%, VAL RMSE = 0.7810%, VAL RRMSE = 0.1119%. In this study, the performance of GA-RNN with three hidden layers and nine neurons was the best for predicting the price of real estate compared to other ranges, as this range provided the lowest MAE, MAPE, and RMSE.

In future work, the feature selection process can consider not only building specifications but also more significant features related to housing prices. Different types of data can be part of the measuring parameters, such as reviews and comments from various social media platforms, amenities, crime rate, and external building characteristics, to improve the performance of the model's output results. When the chosen data set contains geographic information, geocoding could enhance the performance of the stated model. Instead of using a time series model, which is not suitable in this case, we should replace the proposed model with an appropriate one for the modeling part.

References

[1] J. Cruz-Cárdenas, E. Zabelina, O. Deyneka, J. Guadalupe-Lanas, and M. Velín-Fárez, "Role of demographic factors, attitudes toward technology, and cultural values in the prediction of technology-based consumer behaviors: A study in developing and emerging countries," *Technological Forecasting and Social Change,* vol. 149, p. 119768, 2019. https://doi.org/10.1016/j.techfore.2019.119768

[2] F. Wang, Y. Zou, H. Zhang, and H. Shi, "House price prediction approach based on deep learning and ARIMA model," presented at the 2019 IEEE 7th International Conference on Computer Science and Network Technology (ICCSNT) IEEE, 2019.

[3] V. Venkatesh, "Adoption and use of AI tools: A research agenda grounded in UTAUT," *Annals of Operations Research*, vol. 308, no. 1, pp. 641-652, 2022. https://doi.org/10.1007/s10479-020-03918-9

[4] D. Tchuente and S. Nyawa, "Real estate price estimation in French cities using geocoding and machine learning," *Annals of Operations Research*, vol. 308, no. 1, pp. 571-608, 2022. https://doi.org/10.1007/s10479-021-03932-5

[5] J. Bilski, B. Kowalczyk, A. Marchlewska, and J. M. Zurada, "Local levenberg-marquardt algorithm for learning feedforwad neural networks," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 10, no. 4, pp. 299-316, 2020. https://doi.org/10.2478/jaiscr-2020-0020

[6] L. B. Gokalani, B. Das, D. K. Ramnani, M. Kumar, and M. A. Shah, "House price prediction of real time data (DHA Defence) Karachi using machine learning," *Sir Syed University Research Journal of Engineering & Technology*, vol. 12, no. 2, pp. 75-80, 2022. https://doi.org/10.33317/ssurj.504

[7] 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. https://doi.org/10.1080/09599916.2020.1832558

[8] Y. Li, Z. Xiang, and T. Xiong, "The behavioral mechanism and forecasting of Beijing housing prices from a multiscale perspective," *Discrete Dynamics in Nature and Society*, vol. 2020, pp. 1-13, 2020. https://doi.org/10.1155/2020/5375206

[9] R. Liu and L. Liu, "Predicting housing price in China based on long short-term memory incorporating modified genetic algorithm," *Soft Computing*, vol. 23, no. 22, pp. 11829-11838, 2019. https://doi.org/10.1007/s00500-018-03739-w

[10] B. Nazemi and M. Rafiean, "Forecasting house prices in Iran using GMDH," *International Journal of Housing Markets and Analysis*, vol. 14, no. 3, pp. 555-568, 2020. https://doi.org/10.1108/ijhma-05-2020-0067

[11] R. Xu, "Research on housing price forecasting model based on multiple linear regression model and neural network model," in *Proceedings of the 4th International Conference on Economic Management and Model Engineering, ICEMME 2022, November 18-20, 2022, Nanjing, China. https://doi.org/10.4108/eai.18-11-2022.2327165*, 2023.

[12] P.-F. Pai and W.-C. Wang, "Using machine learning models and actual transaction data for predicting real estate prices," *Applied Sciences*, vol. 10, no. 17, p. 5832, 2020. https://doi.org/10.3390/app10175832

[13] B. Ge, M. M. Ishaku, and H. I. Lewu, "Research on the effect of artificial intelligence real estate forecasting using multiple regression analysis and artificial neural network: A case study of Ghana," *Journal of Computer and Communications*, vol. 9, no. 10, pp. 1-14, 2021. https://doi.org/10.4236/jcc.2021.910001

[14] R. Monika, "House price forecasting using machine learning methods," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 11, pp. 3624-3632, 2021.

[15] L. Liu and L. Wu, "Predicting housing prices in China based on modified Holt's exponential smoothing incorporating whale optimization algorithm," *Socio-Economic Planning Sciences*, vol. 72, p. 100916, 2020. https://doi.org/10.1016/j.seps.2020.100916

[16] J.-S. Chou, D.-B. Fleshman, and D.-N. Truong, "Comparison of machine learning models to provide preliminary forecasts of real estate prices," *Journal of Housing and the Built Environment*, vol. 37, no. 4, pp. 2079-2114, 2022. https://doi.org/10.1007/s10901-022-09937-1

[17] X. Xu and Y. Zhang, "House price forecasting with neural networks," *Intelligent Systems with Applications*, vol. 12, p. 200052, 2021. https://doi.org/10.1016/j.iswa.2021.200052

[18] L. D'Acci, "Quality of urban area, distance from city centre, and housing value. Case study on real estate values in Turin," *Cities*, vol. 91, pp. 71-92, 2019. https://doi.org/10.1016/j.cities.2018.11.008

[19] H. Li, Y. D. Wei, Y. Wu, and G. Tian, "Analyzing housing prices in Shanghai with open data: Amenity, accessibility and urban structure," *Cities*, vol. 91, pp. 165-179, 2019. https://doi.org/10.1016/j.cities.2018.11.016

[20] S. Li, Y. Jiang, S. Ke, K. Nie, and C. Wu, "Understanding the effects of influential factors on housing prices by combining extreme gradient boosting and a hedonic price model (XGBoost-HPM)," *Land*, vol. 10, no. 5, p. 533, 2021. https://doi.org/10.3390/land10050533

[21] W. Yuchi *et al.*, "Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city," *Environmental Pollution*, vol. 245, pp. 746-753, 2019. https://doi.org/10.1016/j.envpol.2018.11.034

[22] Y. Ao, H. Li, L. Zhu, S. Ali, and Z. Yang, "The linear random forest algorithm and its advantages in machine learning assisted logging regression modeling," *Journal of Petroleum Science and Engineering*, vol. 174, pp. 776-789, 2019. https://doi.org/10.1016/j.petrol.2018.11.067

[23] Q. Ma, "Comparison of ARIMA, ANN and LSTM for stock price prediction," presented at the E3S Web of Conferences. EDP Sciences, 2020.

[24] M. Uzair and N. Jamil, "Effects of hidden layers on the efficiency of neural networks," presented at the 2020 IEEE 23rd International Multitopic Conference (INMIC). IEEE, 2020.

[25] M. Tekin and I. U. Sari, "Real estate market price prediction model of Istanbul," *Real Estate Management and Valuation*, vol. 30, no. 4, pp. 1-16, 2022. https://doi.org/10.2478/remav-2022-0025

[26] L. Zhao, J. Mbachu, and Z. Liu, "Exploring the trend of New Zealand housing prices to support sustainable development," *Sustainability*, vol. 11, no. 9, p. 2482, 2019. https://doi.org/10.3390/su11092482

[27] L. Medsker and L. C. Jain, *Recurrent neural networks: Design and applications*, 1st ed. CRC Press. https://doi.org/10.1201/9781003040620, 1999.

[28] S. Ding, C. Su, and J. Yu, "An optimizing BP neural network algorithm based on genetic algorithm," *Artificial Intelligence Review*, vol. 36, no. 2, pp. 153-162, 2011. https://doi.org/10.1007/s10462-011-9208-z

[29] J. D. Schaffer, D. Whitley, and L. J. Eshelman, "Combinations of genetic algorithms and neural networks: A survey of the state of the art," in *Proceedings COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks. IEEE*, 1992, pp. 1-37.

[30] S. Bandaru and K. Deb, *Metaheuristic techniques. In Decision Sciences; Sengupta, R., Gupta, A., Dutta, J., Eds*. Boca Raton, FL, USA: CRC Press, 2016.