**International Journal of Electrical and Computer Engineering (IJECE)**
Vol. 13, No. 4, August 2023, pp. 4516~4525
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i4.pp4516-4525 

# The role of neural network for estimating real estate prices value in post COVID-19: a case of the middle east market

**Laith T. Khrais, Osman Saad Shidwan**
Department of Business Administration, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia

**Article Info**

***Article history:***

Received Nov 9, 2022
Revised Dec 16, 2022
Accepted Dec 17, 2022

***Keywords:***

Artificial neural networks
COVID-19
Digital economy
Python
Real estate prices
Saudi Arabia
Technology

**ABSTRACT**

The main goal of this paper was to explore the use of an artificial neural network (ANN) model in predicting real estate prices in the Middle East market. Although conventional modeling approaches such as regression can be used in prediction, they have a weakness of a predetermined relationship between input and output. In this regard, using the ANN model was expected to reduce the bias and ensure non-linear relationships are also covered in the prediction process for more accurate results. The ANN model was created using Python v.3.10 program. The model exhibited a high correlation between predicted and actual house price data (R=0.658). In this respect, it was realized that the model could be effectively used in appraising real estate by investors. However, a major limitation of the model was realized to be a limited dataset for large and luxurious houses, which were not accurately predicted as data distribution between actual and predicted values became sparse for high house prices. A key recommendation made is that future research should include more variables related to luxurious houses and macroeconomic factors to increase the ANN model accuracy.



***Corresponding Author:***

Laith T. Khrais
Department of Business Administration, College of Applied Studies and Community Service, Imam Abdulrahman Bin Faisal University
P.O. Box 1982, Dammam, Saudi Arabia
Email: Lakhris@iau.edu.sa

## 1. INTRODUCTION

In most economies, real estate is recognized as one of the most lucrative investments due to its ability to sustain long-term high returns on investment. As such, the subject of real estate valuation has drawn significant attention over the decades as experts strive to determine the most accurate methods of calculating the prices of real estate [1]. Generally, the current state of the economy and the value of money for any country are the key predictors of the value of real estate [1]. However, the concept of real estate appraisal has since been complicated due to the introduction of additional factors based on the application of the values identified. Essentially, the increase in population and rapid urbanization, together with numerous transactions involving real estate value in areas such as socialization, taxation, buying property, and loan acquisition, has led to a series of contradictions in accurate valuation [2]. This has contributed to different valuation methods that have different results. Therefore, there is a need for improvement in the manner in which real estate prices are determined. On that note, the main objective of this research is to propose and demonstrate the use of neural network models in estimating real estate prices in the Middle East. Specifically, neural networks in computerization are artificial networks inspired uniquely by the biological nervous system thus mimicking the functionality of the human brain and neurons [3].

For this study, the underlying goal is to use artificial neural networks (ANN) to estimate real estate prices in the Middle East for an accurate valuation. In addition to the social and economic factors that influence real estate prices, the matter has been further complicated by the effect of the coronavirus disease of 2019 (COVID-19) pandemic [4]. While the existing real estate valuation methods and models have shown volatility and inaccuracy in prediction, the aftermath of the pandemic even made matters worse due to the economic downturn. For instance, Saudi Arabian urbanization also faces the challenge of environmental pollution and the influence of oil prices [5]. Besides the pandemic, the models for real estate valuation could not provide viable estimates that represent the economy [6]. This calls for the utilization of neural networks with the capacity to think through the numerous data shifts. The problem of inaccurate appraisal of real estate using conventional regression models needs to be addressed because it can lead to businesses incurring huge losses or customers paying for overvalued properties.

## 2. LITERATURE REVIEW

The estimation of real estate prices is commonly referred to as real estate valuation and is normally conducted for specific reasons such as loan application, taxation, sales, investment, insurance, and listing, among others [7]. In computing the value of real estate, Yeh and Hsu [8] indicated that several factors are considered, such as physical property attributes, socioeconomics, neighborhood features, and government regulations. Additionally, the authors argued that the stated factors were under the influence of utility, demand, scarcity, and transferability [8]. In that regard, the implication is that the valuation of real estate is not a direct process which explains the existence of several computation models. In the same breath, Niu and Niu [9] noted that the failure of most models to give accurate real estate prices was a result of the volatility of the variables stated in [8]. This confirms the performance of models especially following the COVID-19 pandemic, whereby significant volatility of the aforementioned factors was observed globally. According to Abidoye and Chan [10], models that are unable to track the changes in the real estate variables are ineffective and always increase the chances of undervaluing and overvaluing. This may result in financial losses and economic exploitation. In the meantime, Kok *et al.* [11] explained that valuation models are specially designed for the estimation of real estate prices by using many qualitative and quantitative variables. The implication is that any real estate valuation model, whether manual or automatic is always measured to the specified tasks. However, Kok *et al.* [11] argued that the effectiveness of combined manual and automated models was higher than individual models. For instance, a review [3] showed that some of these models that successfully combined artificial neural networks and regression analysis included particle swarm optimization algorithm, support vector machine (SVM), and decision tree. In this respect, neural networks can be integrated with other real estate valuation models for precise estimates. However, Bin *et al.* [12] denoted that regression models presented nonlinearity and multicollinearity among variables which produced inaccuracies in the property valuations. Therefore, the implication is that there is a literature gap in the development of accurate real estate valuation models such as neural networks.

To overcome the challenges associated with the computation of real estate prices in Saudi Arabia, the current research proposes the use of a neural network model. The neural network for value determination stems from artificial neural networks, which have received significant applications in different fields, including data mining, speech, machine translation, cybernetics, virtualization, medical diagnostics, and social network filtering. Essentially, neural networks are machine learning innovations that will gather and analyze big and sophisticated data for accurate valuation results. In real estate, the technology is deemed useful for interpreting and understanding the dynamic big data associated with the changes of the pandemic. According to Abidoye and Chan [13], neural network models have been shown to have efficacy in capturing non-linear associations between variables in real estate valuation data. In that regard, the proposed model is capable of addressing the challenges and shortcomings of regression methods of property valuation. Moreover, good predictive prices can be detected using neural networks without pre-processing; hence the proposed model is both time and cost-effective.

Although the application of artificial neural networks has gained popularity among statisticians, engineers, and scientists, its application in the valuation of real estate property is still spreading. As such, the current research presents the introduction of machine learning innovation in the real estate sector of Saudi Arabia. Generally, the research presents an avenue of migration from the traditional means of real estate price estimation to automated artificial intelligence means. Therefore, for Saudi Arabia, this research contributes new value to the economy by improving the accuracy of certain computations linked to real estate value. Also, the true value of real estate means a sincere reflection of the country’s wealth hence aiding the right measures towards development and recovery from the pandemic effects. In the next section, the method used to achieve the research goals is presented.

## 3. METHOD

For this study, a quantitative method was employed. Essentially, the quantitative method entails the collection of numerical data, which are then analyzed using statistical software [14]. A major benefit of the quantitative method is that it is more reliable and accurate in predicting future relationships between variables [15]. For this study, a quantitative method was implemented by developing an ANN model to predict real estate prices in Saudi Arabia. A key advantage of using the ANN model over conventional modeling methods such as linear regression or logistic regression is that it provides a way to understand the non-linear relationship between variables and enables the developed algorithm to predict output from inputs [16]. In this respect, unlike regression, there is no need to assume certain explicit functions or relationships between outputs and input in ANN [17]. Essentially, the ANN model is trained using a large dataset to learn from it and help to predict output based on new test data provided [18]. In summary, ANN often uses three layers, including input, hidden, and output, as shown in Figure 1.

In conducting this study, several steps were undertaken. The first step involved determining suitable variables to use in ANN model training and gathering relevant data. In selecting dataset variables to include in ANN training, previous studies done on the topic in other countries were considered [19]–[25]. The gathered data involved house prices and related variables in Saudi Arabia and were extracted from different databases on property sales [20], [21]. The ANN dataset involved 4,600 records, of which 80% was used for training the ANN model, and 20% was used in testing the model to determine its accuracy. After data collection, an algorithm was developed to train the ANN model on different housing properties and local conditions. The third step involved using the trained ANN model to predict the house prices based on the test dataset and then comparing them with the actual house price values. The coefficient of correlation and the coefficient of determination were computed to understand the accuracy of the ANN model in predicting house prices. The next sections indicate detailed information on data variables and algorithms used in this study. Employing the ANN model in predicting house prices for Saudi Arabia was particularly crucial because the country is sparsely populated with high population density in certain areas, which influences house prices [22]. Population density for Saudi Arabia is shown in Figure 2.

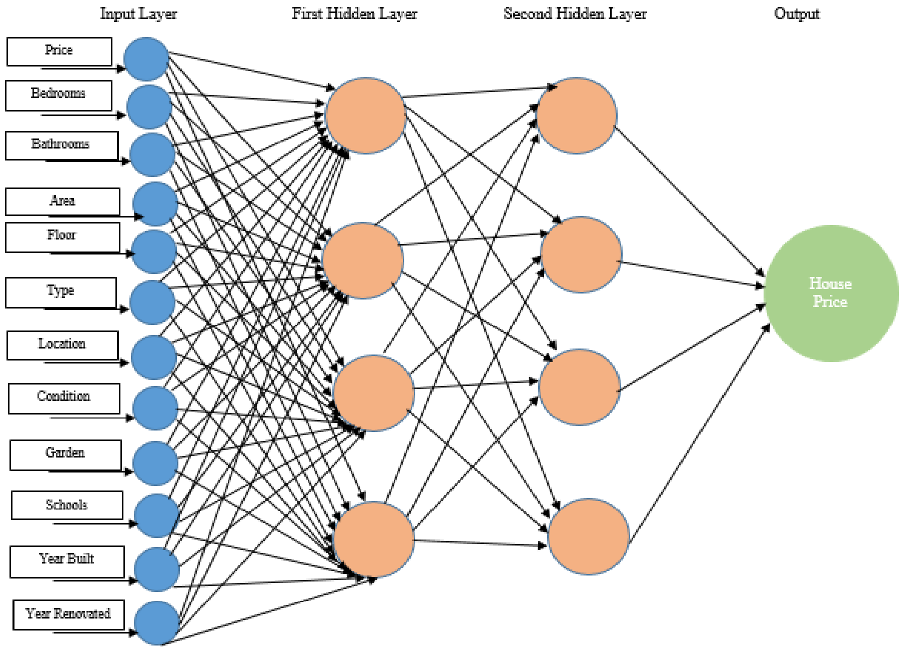


Figure 1. ANN model design

### 3.1. Data variables

The variables which were gathered for each property and used in ANN modeling were as follows:

- Price: Showing the sale price of a property in Saudi Riyal (SAR).
- Bedrooms (BDRM): Showing the number of bedrooms for each property.
- Bathrooms (BATH): Showing the number of bathrooms for each property.
- Area (SQM): Showing the area of each property in square meters.

- Floors (FLR): Showing floor where the property is located in a building.
- Type (TYP): Dummy variable showing whether a building is residential (0) or commercial (1).
- Location (LOC): Dummy variable showing the zone where the property is located from the capital city (0).
- Condition (CON): Shows condition of property on continuum of 1-5 (good condition=5, poor condition 1).
- Garden (GRD): Dummy variable where the presence of garden is 1.
- NSN: Number of schools nearby.
- DMR: Distance to a major road (meters).
- BLT: Year the property was built.
- REN: Year of renovation.

The descriptive statistics of the variables used in this study were summarized down in Table 1. It is realized that there is a large disparity in house prices in Saudi Arabia because the standard deviation (SAR 563,834.7) is greater than the mean house price value of SAR 551,963. In this regard, it is realized that it can be difficult to predict house prices because prices can significantly change based on a particular variable in a non-linear manner. In the next section, the complete results of the ANN modeling are presented.

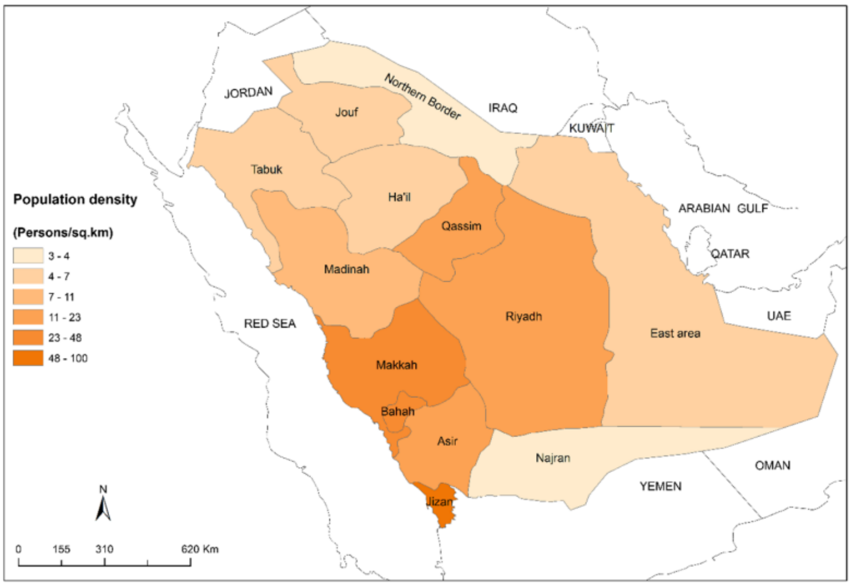


Figure 2. Population density heat map of Saudi Arabia showing regions in the Southwest, such as Makkah and Bahah, to have high population density [21]

Table 1. Descriptive statistics for data variables

| | Price | BDRM | BATH | SQM | FLR | TYP | LOC | CON | GRD | NSN | DMR | BLT | REN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 551,963 | 3.401 | 2.16 | 213.93 | 1.51 | 0.007 | 0.241 | 3.452 | 0.459 | 2.02 | 64.83 | 1,970 | 1,994 |
| Standard Error | 8,313.28 | 0.013 | 0.01 | 1.420 | 0.007 | 0.001 | 0.0115 | 0.00998 | 0.00734 | 0.01 | 0.13 | 0 | 0 |
| Median | 460,943.5 | 3 | 2.25 | 198 | 1.5 | 0 | 0 | 3 | 0 | 2 | 65 | 1,976 | 2,001 |
| Mode | 0 | 3 | 2.5 | 194 | 1 | 0 | 0 | 3 | 0 | 3 | 50 | 2,006 | 2,000 |
| SD | 563,834.7 | 0.90 | 0.78 | 96.32 | 0.53 | 0.084 | 0.778 | 0.677 | 0.498 | 0.82 | 8.93 | 29 | 21 |
| Variance | 3.18E+11 | 0.82 | 0.61 | 9,277.6 | 0.28 | 0.007 | 0.606 | 0.458 | 0.248 | 0.67 | 79.79 | 883 | 456 |
| Kurtosis | 1,044.35 | 1.23 | 1.86 | 8.291 | -0.53 | 134.5 | 10.464 | 0.197 | -1.974 | -1.51 | -1.18 | 0 | 5 |
| Skewness | 24.79 | 0.456 | 0.61 | 1.723 | 0.551 | 11.68 | 3.342 | 0.959 | 0.161 | -0.03 | 0.00 | 0 | -2 |
| Range | 26,590,000 | 9 | 8 | 1317 | 2.5 | 1 | 4 | 4 | 1 | 2 | 30 | 114 | 102 |
| Minimum | 0 | 0 | 0 | 37 | 1 | 0 | 0 | 1 | 0 | 1 | 50 | 1,900 | 1,912 |
| Maximum | 26,590,000 | 9 | 8 | 1354 | 3.5 | 1 | 4 | 5 | 1 | 3 | 80 | 2,014 | 2,014 |
| Count | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 | 4,600 |

### 3.2. Algorithm steps

For this study, the ANN model was created by using algorithms developed in Python Software v. 3.10 on the Jupyter Notebook online platform [26]. For this study, the ANN model was created by using algorithms developed in Python Software v. 3.10 on the Jupyter Notebook online platform [26]. The simulation process steps followed were similar to those of other researchers who have investigated ANN usage in prediction [27], [28]. The simulation first step involves identifying input and output data to understand the complexity of the model, thereby influencing the decision on the number of hidden layers for ANN [27]. For few input and output variables, a maximum of two hidden layers is recommended [28]. The next simulation step on splitting and training data is crucial because it helps the algorithm to determine the best way of connecting input and output data [29], [30]. During ANN training, the program assigns weights to different input variables based on their impact on output prediction. The process is repeated in several iterations until the predicted output matches the actual output based on the training data provided. After training, a test data is fed into the ANN model and used to predict the output [30]. The output of the prediction from test data helps to achieve the fourth step of determining accuracy of the ANN model by comparing the predicted values and the actual output. In developing the proposed algorithm, there were several key steps undertaken. In summary, the simulation process was followed in developing the proposed algorithm involved:

Step 1: Initialize ANN model

The ANN model was initialized by first importing relevant files such as NumPy, Pandas, Matplotlib, and Keras that were intended to help in the analysis.

```
Import numpy as np
Import pandas as pd
Import math
Import keras #initialize ann
From keras. layers import dense #inform number of layers, neurons per layer and activation function
From sklearn. metrics import r2_score #to import r square function
Import matplotlib. pyplot as plt
From sklearn. model_selection import train_test_split #to split the data into training and testing set
```

In the step

*ANN*: Artificial neural network,

*NumPy*: Python library that enables working with arrays of columns and rows

*Pandas*: Python library that enables cleaning, exploring, and manipulating panel data

*Keras*: Python library that enables solving of complex machine learning problems

*Math*: Python library that offers standard mathematical functions and constants

*Sklearn*: Python library that enables splitting of data into training and testing data sets

*Matplotlib*: Python library that enables creation of visualization plots from dataset.

Step 2: Data import

The gathered data in a .csv file was imported to Python software and then split into 80% for training and 20% for testing using the available functions. The allocation of a large percentage of data for training was aimed at ensuring the model has adequate data points to learn patterns and developed accurate prediction patterns.

*#Split the data into 80% training and 20% (testing (10%) and validating (10%))*

$$\mathrm{X\_train, X\_val\_and\_test, Y\_train, Y\_val\_and\_test = train\_test\_split(X, Y, test\_size\ = 0} \quad (1)$$

$$\mathrm{X\_val, X\_test, Y\_val, Y\_test = train\_test\_split(X\_val\_and\_test, Y\_val\_and\_test, test\_size\ = 0.5)} \quad (2)$$

*#Training set has 3680 data points while the validation and test set have 460 data points each*

*#The x variables have 12 input features while y variable only has one feature*

$$\mathrm{print(X\_train.\,shape, X\_val.\,shape, X\_test.\,shape, Y\_train.\,shape, Y\_val.\,shape, Y\_test.\,shape)}$$

In the step

$X\_train$: X values used in training of algorithm to detect patterns in dataset

$Y\_train$: Y values used in training of algorithm to detect patterns in dataset

$X\_val$: X values used in fine tuning the final ANN algorithm to improve prediction accuracy

$Y\_val$: Y values used in fine tuning the final ANN algorithm to improve prediction accuracy

$X\_test$: X values used in testing of algorithm to determine accuracy of prediction

$Y\_test$: Y values used in testing of algorithm to determine accuracy of prediction

Step 3: Model architecture

The architecture of the ANN model was developed where it was assumed that it consisted of an input layer, two hidden layers, and an output layer. Additionally, it was assumed that each hidden layer and the output layer would consist of 32 neurons. Due to runtime challenges of the Python program, more hidden layers could not be included in the ANN model. A loss function was also used to measure how well the ANN model did on training and try to improve on it using the optimizer.

*#Build the model architecture of the artificial neural network* $model = keras.models.Sequential()$
*#Initializes the ANN*

$$\mathrm{model.add(keras.layers.Dense(32, activation} = \mathrm{'relu', input\_shape} = (12,)))\qquad(3)$$

*#Assume 32 neurons for the input layer and 12 input variables.*

$$\mathrm{model.add(keras.layers.Dense(32, activation} = \mathrm{'relu'))}\textit{\#First hidden layer}\qquad(4)$$

$$\mathrm{model.add(keras.layers.Dense(32, activation} = \mathrm{'relu'))}\textit{\#Second hidden layer}\qquad(5)$$

$model.add(keras.layers.Dense(1))$ *#Output layer #Loss function* measures how well the model did on training and try to improve on it using the optimizer.

$$\mathrm{model.compile(optimizer} = \mathrm{'adam', loss} = \mathrm{'mean\_squared\_error', metrics} = [\mathrm{'Accuracy'}])\ (6)$$

In the step
*relu*: activation function that transforms a specific weighted input into output.
*adam*: optimization function that converges inputs to optimal solution.
Step 4: Model training

The ANN model was trained using 30 iterations, and the dataset consisting of 3,680 training data points and 12 input variables. The training was to allow the ANN model to learn from the data set and identify the most accurate patterns linking the inputs and output variables.
*#Train the ANN model*

$$\mathrm{hist} = \mathrm{model.fit(X\_train, Y\_train,}\qquad(7)$$

$$\mathrm{batch\_size} = 32,\qquad(8)$$

$$\mathrm{epochs} = 30\qquad(9)$$

$$\mathrm{validation\_data} = \mathrm{(X\_val, Y\_val))}\qquad(10)$$

In the step
*batch_size*: Number of samples passed through the ANN in one iteration during machine learning
*epoch*: Number of iterations during training of dataset.
Step 5: Testing

The ANN model was tested using one dataset array to gauge its accuracy and debug any issues. After debugging the algorithm, the ANN model was used to predict all values of house prices based on provided independent variables.
*#Test the prediction ability of the model using data for row no. 7 (P=335000)*

```
test_data = np.array([2, 2, 135, 1, 0, 0, 3, 0, 3, 72, 1976, 0])
        print(model.predict(test_data.reshape(1,12), batch_size = 1))
        1/1 [==============================] - 0s 464ms/step
                        [[365066.34]]                                (11)
```

As shown in the screenshot, the actual house price was SAR 335,000 while the predicted value was SAR 365,066.34. This revealed that the ANN model did not have debugging issues and was fairly accurate in prediction. In this regard, the whole testing dataset which was split at the start of the ANN model was predicted based on X values.
*#Make a prediction*

```
prediction = model.predict(X_test)
print(prediction)
print(Y_test)                                                        (12)
```

In the step
*Test data:* Dataset used to make tentative testing to check accuracy of the ANN model.

Step 6: Compare predicted, and actual house prices

The predicted house prices from the ANN model were compared against the actual house prices from the split test data. The comparison was crucial in revealing not only the extent of correlation between the predicted and actual house price values but also the dispersion of data trends in the plot generated.

$$\text{R\_square} = \text{r2\_score}(\text{Y\_test}, \text{prediction}) \tag{13}$$

$$\begin{aligned}&\text{print}(\text{'coefficient of determination'}, \text{R\_square})\\&\text{print}(\text{"correlation R = "}, \text{math.sqrt}(\text{R\_square}))\end{aligned} \tag{14}$$

In the step

*R_square*: Coefficient of determination that shows how well the original data fits the model.

*R*: Correlation coefficient that shows the relationship between dependent and independent variables

Step 7: Graph generation

Relevant graphs were generated to show the comparison and accuracy of the model. Additionally, an R-square value was generated.

*#Visualize the training predicted values and the Actual Y-test values*

$$\begin{aligned}&\text{plt.plot}(\text{Y\_test}, \text{prediction}, \text{"o"})\\&\text{plt.title}(\text{'Prediction vs Actual House Prices'})\\&\text{plt.ylabel}(\text{'Predicted House Prices (SAR)'})\\&\text{plt.xlabel}(\text{'Actual House Prices (SAR)'})\end{aligned}$$

In the step

*Y test*: Y values in testing data set that was split at the start of the ANN model

*Prediction*: Predicted Y values using X values of testing data that was split at the start of ANN mode

## 4. RESULTS AND DISCUSSION

A key result noted from this study involved comparing the actual house prices with the estimated output from the ANN model when all the independent variables from test data were employed. Essentially, it was realized that there was a linear relationship between the ANN predicted and actual values of house prices. The result for plotted ANN model values against actual house prices was summarized in Figure 3.

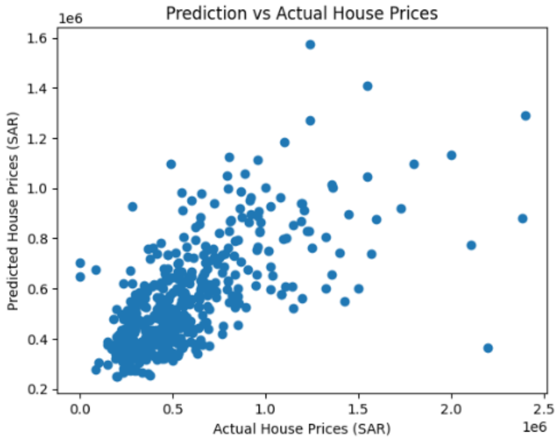


Figure 3. ANN predicted values of house prices vs actual house prices from test data

Figure 3 suggests that the ANN model accurately predicts house prices to a higher degree when the values are small, but for very high house values, there is large dispersion of the plot and inaccuracy in the model. The result suggests a weakness in the ANN prediction model and needs improvement. Essentially, the result shows that incorporating diverse house prices, including pricey high-end houses, can help improve the accuracy of the ANN prediction model. The coefficient of correlation for the relationship between the predicted and actual house prices in Saudi Arabia was determined as shown in (15).

$$\text{print("correlation R = ", math.sqrt(R\_square))}$$
$$\text{Correlation R} = 0.6582646543270136 \tag{15}$$

The coefficient of correlation shows the strength of the relationship between two variables. For this case, the result supports the argument made that ANN model house prices have a positive relationship with actual house prices, but the strength of the relationship reduces with high house prices. The result is attributed to the omission of certain key variables related to luxurious houses in the ANN model, which may have limited its accuracy of prediction. At the same time, certain macroeconomic factors which affect forces of supply and demand, such as real income, population, and interest rates, were not included in the ANN model, thereby limiting its capacity for accurate prediction. The result implied that the ANN model could be used by real estate businesses and customers in Saudi Arabia to predict house prices in preparation for negotiations and making deals. The finding resonated with those of previous authors, such as Kok *et al.* [11], who highlighted that valuation models are more accurate when both qualitative and quantitative models are considered. The developed ANN model included three qualitative values and ten quantitative variables, which may have influenced the positive result noted. Based on Kok *et al.* [11] views, the prediction accuracy of the created ANN model can be improved by including more qualitative variables related to house features, such as the presence of garages, swimming pools, or home gyms.

At the same time, the high prediction accuracy of the ANN model can be explained by Bin [12], highlighting that neural networks are unlike regression models that are often undermined by issues of multicollinearity and nonlinearity, which lead to property valuation inaccuracies. On the other hand, the R-square value relating the predicted and the actual house prices was also computed using (16), and the output was obtained as shown. The R-square value is crucial in showing the variability of y-values that can be explained by x-values.

$$\text{R\_square} = \text{r2\_score(Y\_test, prediction)}$$
$$\text{print('coefficient of determination', R\_square)}$$
$$\text{Coefficient of determination} = 0.43331235513626276 \tag{16}$$

Essentially, the result shows that the coefficient of determination was 43.33%, thereby suggesting the ANN model output could be moderately accurate in predicting the actual house prices in Saudi Arabia. However, the result suggests that the ANN model still has many unexplained patterns in predicting house prices. As such, the result reveals the need to increase the number of independent variables which can be added to the input layer to improve the accuracy of the ANN model prediction. The issue could not be associated with small data points for training because 80% of the collected data was allocated for the purpose. Instead, the result was attributed to a lack of crucial macroeconomic variables such as interest rates, real income, population, and fuel prices which influence Saudi Arabia’s economy. The result is consistent with the views of Yeh [8], who underlined that in improving the accuracy of neural networks for real estate valuation, government regulations and socioeconomic factors should be considered besides the property’s physical properties.

In the context of system construct, key mistake noted when creating the ANN model was the omission of several crucial socioeconomic variables, such as income, employment rates, education level, and interest rates, that influence real estate prices. Therefore, future research can improve the ANN model by including diverse socioeconomic variables. A non-obvious mistake that can be made in developing the ANN involves choosing a data sample from only one region, which can affect the generalization of the obtained model to the whole country. Therefore, stratified random sampling of real estate prices should be used to select data from each region. Basically, the algorithm was run in Python, which has a non-obvious challenge of many runtime errors, implying that a large part of the dataset is needed for training. Therefore, future researchers can explore the ANN model using alternative programming languages such as C#, Java, and GO to determine whether there is a significant difference in prediction accuracy. A key weakness of the algorithm developed was that it could not predict prices based on location in Saudi Arabia because of missing data regarding property locations from the selected database. As such, future researchers can improve the model's accuracy by including location variables. The other mistake noted when developing the ANN algorithm was a failure to import necessary packages in the Python notebook in the first code cell. The mistake led to many errors being generated when running the code, which could have been avoided by including all the important functions in the first cell.

Nonetheless, the developed ANN model also considered the views of Niu and Niu [9], showing that the failure of prediction models to estimate accurate real estate prices was linked to the volatility of selected variables. Although the created ANN model showed that it had many unexplained patterns in predicting house prices, the failure was not linked to the volatility of variables since all the input variables used were relatively stable and related to physical attributes of the property and neighborhood features. The low R-square value of the model can also be explained by Abidoye and Chan [10], who warned that models that indicate high variability in prediction can cause overvaluing or undervaluing of property which poses a high risk of loss for a business. In this respect, improvement of the ANN model to lower prediction variability should aim at not only adding more input variables but also integrating decision trees and support vector machines to ensure algorithm optimization, as noted by Peter *et al.* [3]. Based on the discussion, it is realized that although the algorithm of the ANN model developed was effective in predicting house prices in Saudi Arabia, there are still many strategies that can be employed to enhance its prediction accuracy.

## 5. CONCLUSION

The main objective of this paper was to create an ANN model to estimate real estate prices in Middle East. The use of ANN is considered more appealing because of its ability to make non-linear relationships between variables, thereby ensuring better decisions based on predicted values. For the case of Saudi Arabia's real estate sector, the developed ANN model revealed a high correlation between predicted and actual house price values (R=0.658). The result implied that the developed ANN model could be used by buyers and sellers in the Saudi Arabia market to estimate property prices with the presented parameters. In this way, the set research objective was achieved, showing that using the ANN model in estimating Saudi Arabia's real estate price is a viable option because it depicts a high correlation between predicted and actual house price values. Nonetheless, the low value of R-square obtained (R-square=0.433) suggested that there are still many unexplained aspects of the actual house prices using the ANN model, thereby indicating the need to improve the model accuracy. The improvement of the ANN model was realized to involve the addition of input variables as well as the integration of support vector machine analysis and decision trees.

A major strength of the ANN model developed is that it considered a wide array of factors in predicting house prices which ensured it can be used by most real estate agents when seeking to appraise the property. However, a key limitation of the neural network created is that it does not accurately predict the prices of luxurious houses, which have significantly higher prices compared to ordinary houses. The limitation was influenced by the small dataset available for luxurious houses in the Saudi Arabia database, which limited the training of the ANN model. At the same time, the trend may have been influenced by limited independent variables related to luxurious houses, such as swimming pools, home gyms, or gourmet kitchens which can help to improve the accuracy of the ANN model. A key recommendation for practice made from the study is that businesses engaged in selling property in Saudi Arabia can use the developed ANN model to set reasonable property prices that are aligned with market rates. At the same time, buyers can use the ANN model to appraise property in preparation for negotiation with sellers. Meanwhile, a key recommendation for theory is that future researchers should include more independent variables involving macroeconomic factors to ensure the dynamics of supply and demand are considered to improve the accuracy of the model regarding house prices. The variables used in this study were limited due to the small scope of the work. Additionally, future researchers should redesign the ANN model to focus on diverse groups of properties, including industrial, since this study only considered residential and commercial properties.

## REFERENCES

[1] O. S. Oshodi, W. D. Thwala, T. B. Odubiyi, R. B. Abidoye, and C. O. Aigbavboa, "Using neural network model to estimate the rental price of residential properties," *Journal of Financial Management of Property and Construction*, vol. 24, no. 2, pp. 217–230, Aug. 2019, doi: 10.1108/JFMPC-06-2019-0047.

[2] J. Ćetković *et al.*, "Assessment of the real estate market value in the European market by artificial neural networks application," *Complexity*, vol. 2018, pp. 1–10, 2018, doi: 10.1155/2018/1472957.

[3] N. J. Peter, H. I. Okagbue, E. C. M. Obasi, and A. O. Akinola, "Review on the application of artificial neural networks in real estate valuation," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 2918–2925, Jun. 2020, doi: 10.30534/ijatcse/2020/66932020.

[4] H. Mahmood, T. T. Y. Alkhateeb, M. M. Z. Al-Qahtani, Z. Allam, N. Ahmad, and M. Furqa, "Urbanization, oil price, and pollution in Saudi Arabia," *International Journal of Energy Economics and Policy*, vol. 10, no. 2, pp. 477–482, Jan. 2020, doi: 10.32479/ijeep.8914.

[5] A. A. Alola, "Evidence of speculative bubbles and regime switch in real estate market and crude oil price: Insight from Saudi Arabia," *International Journal of Finance & Economics*, vol. 26, no. 3, pp. 3473–3483, Jul. 2021, doi: 10.1002/ijfe.1971.

[6] R. Alharbi, "An appraisal of the early impact of COVID-19 on affordable housing finance in Saudi Arabia's vision 2030," *International Journal of Building Pathology and Adaptation*, Apr. 2022, doi: 10.1108/IJBPA-12-2021-0169.

[7] V. Del Giudice, P. De Paola, and G. Cantisani, "Valuation of real estate investments through fuzzy logic," *Buildings*, vol. 7, no. 4, pp. 26–48, Mar. 2017, doi: 10.3390/buildings7010026.

[8] I.-C. Yeh and T.-K. Hsu, "Building real estate valuation models with comparative approach through case-based reasoning," *Applied Soft Computing*, vol. 65, pp. 260–271, Apr. 2018, doi: 10.1016/j.asoc.2018.01.029.

[9] J. Niu and P. Niu, "An intelligent automatic valuation system for real estate based on machine learning," in *Proceedings of the International Conference on Artificial Intelligence, Information Processing and Cloud Computing*, Dec. 2019, pp. 1–6. doi: 10.1145/3371425.3371454.

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

[11] N. Kok, E.-L. Koponen, and C. A. Martínez-Barbosa, "Big data in real estate? From manual appraisal to automated valuation," *The*

*Journal of Portfolio Management*, vol. 43, no. 6, pp. 202–211, Sep. 2017, doi: 10.3905/jpm.2017.43.6.202.
[12] J. Bin *et al.*, "Regression model for appraisal of real estate using recurrent neural network and boosting tree," in *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*, Sep. 2017, pp. 209–213. doi: 10.1109/CIAPP.2017.8167209.
[13] R. B. Abidoye and A. P. C. Chan, "Modelling property values in Nigeria using artificial neural network," *Journal of Property Research*, vol. 34, no. 1, pp. 36–53, Jan. 2017, doi: 10.1080/09599916.2017.1286366.
[14] D. Stockemer, *Quantitative methods for the social sciences*. Cham: Springer International Publishing, 2019, doi: 10.1007/978-3-319-99118-4.
[15] S. Eti, "The use of quantitative methods in investment decisions," in *Research Anthology on Personal Finance and Improving Financial Literacy*, IGI Global, 2021, pp. 1–20. doi: 10.4018/978-1-7998-8049-3.ch001.
[16] A. A. Shahmansouri, M. Yazdani, S. Ghanbari, H. Akbarzadeh Bengar, A. Jafari, and H. Farrokh Ghatte, "Artificial neural network model to predict the compressive strength of eco-friendly geopolymer concrete incorporating silica fume and natural zeolite," *Journal of Cleaner Production*, vol. 279, Jan. 2021, doi: 10.1016/j.jclepro.2020.123697.
[17] T. R. Gadekallu *et al.*, "A deep neural networks based model for uninterrupted marine environment monitoring," *Computer Communications*, vol. 157, pp. 64–75, May 2020, doi: 10.1016/j.comcom.2020.04.004.
[18] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," *Future Computing and Informatics Journal*, vol. 3, no. 2, pp. 334–340, Dec. 2018, doi: 10.1016/j.fcij.2018.10.003.
[19] IBM, "The neural networks model," *IBM Corporation*, 2022. https://www.ibm.com/docs/en/spss-modeler/18.0.0?topic=networks-neural-model (accessed Oct. 10, 2022).
[20] Bayut, "Properties for sale in Saudi Arabia," *Bayut*, 2022. https://www.bayut.sa/en/ksa/properties-for-sale/ (accessed Oct. 20, 2022).
[21] Property Finder, "Properties for sale in Saudi Arabia," *Property Finder*, 2022. https://www.propertyfinder.sa/en/buy/ar-riyadh/properties-for-sale.html (accessed Oct. 20, 2022).
[22] M. Algahtany, L. Kumar, E. Barclay, and H. M. Khormi, "The spatial distribution of crime and population density in Saudi Arabia," *Crime Prevention and Community Safety*, vol. 20, no. 1, pp. 30–46, Feb. 2018, doi: 10.1057/s41300-017-0034-3.
[23] A. P. Wahyu, R. M. Auramuwahid, A. S. Noviantina, D. S. Silitonga, and J. J. Lasmana, "Prediction of the accuracy of student boarding house prices around Widyatama University using neural network backpropagation algorithm," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 11, pp. 1403–1410, 2021.
[24] Y. Feng and K. Jones, "Comparing multilevel modelling and artificial neural networks in house price prediction," in *2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM)*, Jul. 2015, pp. 108–114. doi: 10.1109/ICSDM.2015.7298035.
[25] V. Chiarazzo, L. Caggiani, M. Marinelli, and M. Ottomanelli, "A neural network based model for real estate price estimation considering environmental quality of property location," *Transportation Research Procedia*, vol. 3, pp. 810–817, 2014, doi: 10.1016/j.trpro.2014.10.067.
[26] Jupyter, "Jupyter notebook," Jupyter, https://jupyter.org/ (accessed Oct. 20, 2021).
[27] P. G. Asteris and V. G. Mokos, "Concrete compressive strength using artificial neural networks," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11807–11826, Aug. 2020, doi: 10.1007/s00521-019-04663-2.
[28] M. R. Maarif, R. F. Listyanda, Y.-S. Kang, and M. Syafrudin, "Artificial neural network training using structural learning with forgetting for parameter analysis of injection molding quality prediction," *Information*, vol. 13, no. 10, Oct. 2022, doi: 10.3390/info13100488.
[29] M. O. Al-Shawwa, A. A.-R. Al-Absi, S. A. Hassanein, K. Baraka, and S. S. Abu-Naser, "Predicting temperature or humidity in the surrounding environment using artificial neural network," *International Journal for Academic Development*, vol. 2, no. 9, pp. 1–6, 2018.
[30] N. S. El-Jerjawi and S. S. Abu-Naser, "Diabetes prediction using artificial neural network," *International Journal of Advanced Science and Technology*, vol. 121, pp. 55–64, 2018.

## BIOGRAPHIES OF AUTHORS

**Laith T. Khrais** is currently working as an assistant professor at Imam Abdulrahman Bin Faisal University, Saudi Arabia. He received his Master degree from Malaysia and his Ph.D. from Poland. He has published many scientific papers in many international ranked journals, and also works as a reviewer for many international ranked journals. He can be contacted at email: Lakhris@iau.edu.sa.

**Osman Saad Shidwan** is currently working as a head of the department of business administration at Imam Abdulrahman Bin Faisal University, Saudi Arabia. He received his Master degree and Ph.D. from the University of Khartoum. He also received training on case-based education at Brunel University in the United Kingdom and training in the use of the COMFAR III program at the United Nations in Vienna - Austria. He can be contacted at email: Osahmed@iau.edu.sa.