Estimation of urban land price within holly cities by using integrated GIS-regression models: case study Al-Kufa city- Iraq

HaidarRazzaq.Momad **Khawla kareem Kawther**

10960@uotechnology.edu.iq **40232@uotechnology.edu.iq**

Civil Engineering,Department, University of Technology, Baghdad, Iraq

Abstract

Urban land price is the primary indicator of land development in urban areas. Land prices in holly cities have rapidly increased due to tourism and religious activities. Public agencies are usually facing challenges in managing land prices in religious areas. Therefore, they require developed models or tools to understand land prices within religious cities. Predicting land prices can efficiently retain future management and develop urban lands within religious cities. This study proposed a new methodology to predict urban land prices within holy cities. The methodology is based on two models, Linear Regression (LR) and Support Vector Regression (SVR), and nine variables (land price, land area, distance to a river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, distance to educational locations, and distance to hospital and clinics. Our findings showed that the SVR model had outperformed the LR model, where SVR achieved an accuracy of 82.9%.In contrast, LR has achieved 75.40%. Therefore, the presented models can assess land prices in holly cities like Al-Kufa. Furthermore, this tool can retain land pricing, land management, and urban planning in Iraq.

Keyword: Urban land price, GIS, Support Vector Regression, Linear regression**.**

تقدير أسعار الأراضي الحضرية داخل المدن المقدسة باستعمال نماذج انحدار نظم المعلومات الجغرافية المتكاملة: الحالة الدراسية مدينة الكوفةـ العراق.

م. حيدر رزاق محمد   م.د. خولة كريم كوثر

40232@uotechnology.edu.iq   10960@uotechnology.edu.iq

الجامعة التكنولوجية / الهندسة المدنية

الخلاصة :

إن أسعار الأراضي في المناطق الحضرية هي المؤشر الرئيس لتنمية الأراضي في المناطق الحضرية. ارتفعت أسعار الأراضي في المدن المقدَّسة بسرعة بسبب الأنشطة السياحية والدينية. عادة ما تواجه الهيئات العامة تحديات في إدارة أسعار الأراضي في المناطق الدينية. لذلك ، فهم يحتاجون إلى نماذج أو أدوات مطورة لفهم أسعار الأراضي داخل المدن الدينية. يمكن أن يؤدي التنبؤ بأسعار الأراضي إلى الاحتفاظ بكفاءة بالإدارة المستقبلية وتطوير الأراضي الحضرية داخل المدن الدينية. اقترحت هذه الدراسة منهجية جديدة للتنبؤ بأسعار الأراضي الحضرية داخل المدن المقدسة. تعتمد المنهجية على نموذجين ، الانحدار الخطي (LR) ودعم الانحدار المتجه (SVR) ، وتسعة متغيرات (سعر الأرض ، مساحة الأرض ، المسافة إلى النهر ، المسافة إلى الطرق الرئيسية ، المسافة إلى مواقع التراث ، المسافة إلى المساجد التاريخية والمسافة إلى المواقع التجارية ، والمسافة إلى المواقع التعليمية ، والمسافة إلى المستشفى والعيادات ، وأظهرت النتائج التي توصلنا إليها أن نموذج SVR قد تفوق على نموذج LR حيث حقق SVR دقة مقدارها ٨٢.٩٪ ، بينما حقق LR دقة مقدارها 75.40٪ . لذلك ، يمكن للنماذج المقترحة تقييم أسعار الأراضي في المدن المقدسة مثل الكوفة او غيرها. علاوة على ذلك، يمكن استعمال هذه الأداة بإدارة وتسعير الأراضي والتخطيط الحضري في العراق.

الكلمات المفتاحية:

أسعار الأراضي الحضرية، نظم المعلومات الجغرافية، دعم الانحدار المتجه، الانحدار الخطي.

1- Introduction

Land price is considered the leading indicator of the land market in urban areas, reflecting land resource allocation and the economy (Wu 2007) .In general, the development and the variation of urban land prices play a vital role in the social investment field. Therefore, governments aimed to present scientific bases to formulate land management. With the noticeable development of the economy, urban land prices are rapidly changing in different locations in the world (Wang 2011). Moreover, although lands are located in the same area, many factors such as the unique position, area, and distance to important locations, area significantly affect the land revenue., The land market is closely related to the housing market, describing the land as one of the elements of the production of the housing unit, but instead its most crucial element. The land supply is inelastic, not because of the lack of the offered land area, but because of the lack of a habitable land area that enjoys the availability of services and roads necessary for transportation. It is noticed that the land supply increases and its prices decrease.

The farther we are from the city center - due to the distance from services - and the density of housing units, the scarcity of land, and the difficulty of construction make the supply of demand inelastic in urban centers (Jamal Baqer ٢٠١٦) (Sati Sana, ٢٠١٢). Recently, the estimation of urban land price models has attracted private and governmental agencies due to their feasibility in predicting urban land prices. Thus, selecting and applying an appropriate statistical model is necessary to increase estimation accuracy. In Iraq, Holly cities such as Najaf and Karbala have usually constituted places of residence and worship together. The tourism movement of tourists and visitors from all over the world is active in them (Orbasli 2002). Therefore, land prices are relatively increased in these cities compared to other cities. Private and governmental agencies are faced with challenges with the land pricing process in holly cities due to the complex social and spatial factors (Raeda 2020). The integration between spatial models and advanced statistical models such as machine learning models can present a comprehensive spatial and statistical modeling, presenting an overview of the land prices within holly cities. In this study. The researchers have proposed an integrated model to predict and represent the land price within a holly city in Iraq (Al-Kufa city). The proposed model includes GIS modeling using spatial analysis that considers several factors (Mohammed, 2021). These factors are land area, distance to

the river, distance to main roads, distance to heritage locations, distance to historical mosques,

distance to commercial locations, distance to educational locations, distance to hospitals and clinics, and the recent land price. The estimation analysis is conducted using the Support vector regression (SVR) model and simple linear regression model (LR). The proposed model has shown the spatial and quantitative variation in the land prices within Al-Kufa city, Najaf, Iraq.

2- Related works

Several studies are conducted to evaluate and predict land prices. Spatial models are used for spatial modeling of land prices Azar et al., 1994 the used GIS models to estimate the spatial distribution of land prices in Lebanon (Beirut) to assess the urban condition after the civil war. In another study, Luo et al., 2004 evaluated the spatial variation of land prices and their relationship to the locations of African-American citizens in Wisconsin, United States.

On the other hand, Löchl et al., 2006 created a transportation simulation based on the estimation of land prices in Zurich, Switzerland. Moreover, Tsutsumi et al., 2011 a web-based GIS system to create land price maps based on residential lands in Tokyo, Japan. The traditional GIS models cannot predict because they only depend on the input data. With the development of advanced models such as machine learning models, Lee et al., 2021 estimated the change in land prices based on social and environmental factors using regression analysis and big data in Korea. Choi et al., 2021 developed a model based on a deep learning algorithm and residual kriging to estimate land prices in Seoul, South Korea.

Their results indicated that the spatial locations could optimize the land price estimation. The main drawback of deep learning models is that they require a large amount of data. Lee et al., 2021, developed a model based on integrating machine learning algorithms principal component analysis (PCA) and Artificial Neural Network (ANN) to predict land prices.

Although their model improves the estimation performance through the decreasing of uncertainty, however, their result is difficult to be applied in spatial environments such as GIS environments. Consequently, their results are hard to be presented as maps. In this study, the researchers propose integrating machine learning

algorithms and GIS modeling considering the nature of holly cities in Iraq to predict the land prices in an urban area.

3- Methodology
3-1 Study area

The study area is Al-Kufa city (Figure1), located northeast of Al Najaf province and approximately 170 km south of the Iraqi capital (Baghdad). Al-Kufa city is considered one of the most important historical cities in Islamic history. In addition to its religious significance, which includes sacred sites such as Al-Kufa mosque, it is located at the intersection of Northing (31-59) ◦ and easting (44-19) ◦. Al-Kufa city is an important urban area within al Najaf province. Al-Kufa city became a more popular attraction location for visitors due to its religious and historical significance. Which directly contribute the urban growth and development (Al-Duhaidahawi 2020). Al-Kufa city included several attractive locations such as holy shrines and mosques of (Al-Kufa mosque, Muslim ibn Aqil, Al-Sahla mosque, and Hani Ibn Erwa). Furthermore, it includes historical locations as the house of Imam Ali. In addition, other activities such as commercial, cultural, educational. These factors usually contribute to the variation of land prices within Al-Kufa city.
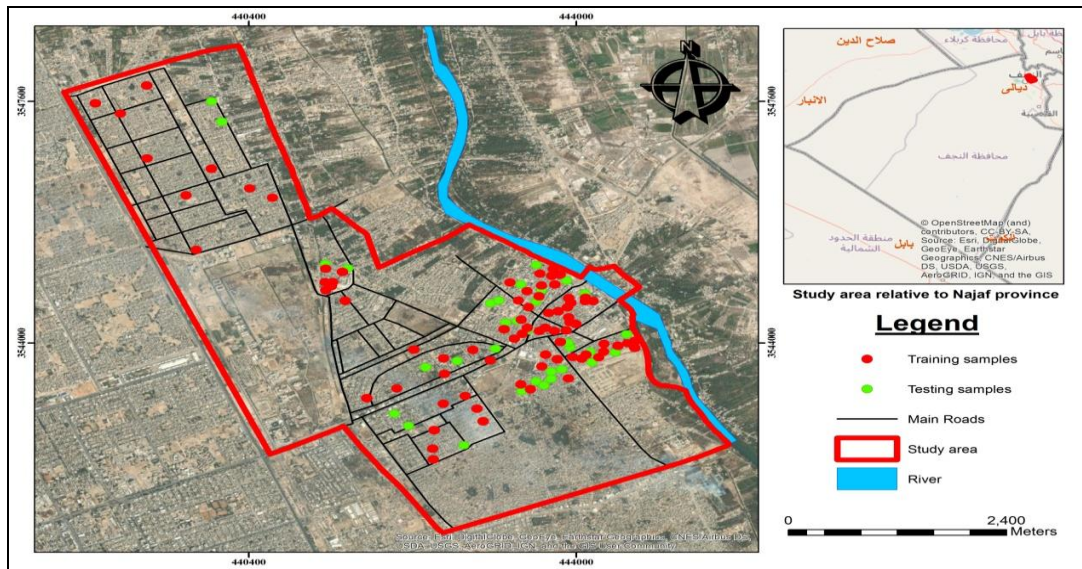

Figure1: The study area.

3-2 The Overall methodology

The proposed methodology comprises several steps; the first step is the data collection included (land price, land area). The next step is the GIS modeling through spatial analysis tools to extract information (distance to river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, and distance to educational locations, and distance to hospital and clinics). Then the extracted data are used in the regression analysis by using Support Vector regression and linear regression models to estimate a land price per million in Al-Kufa city. Finally, the models are validated using testing data, and the root mean square function is used to calculate the accuracy of models. Figure 2 shows a flow chart for the overall methodology
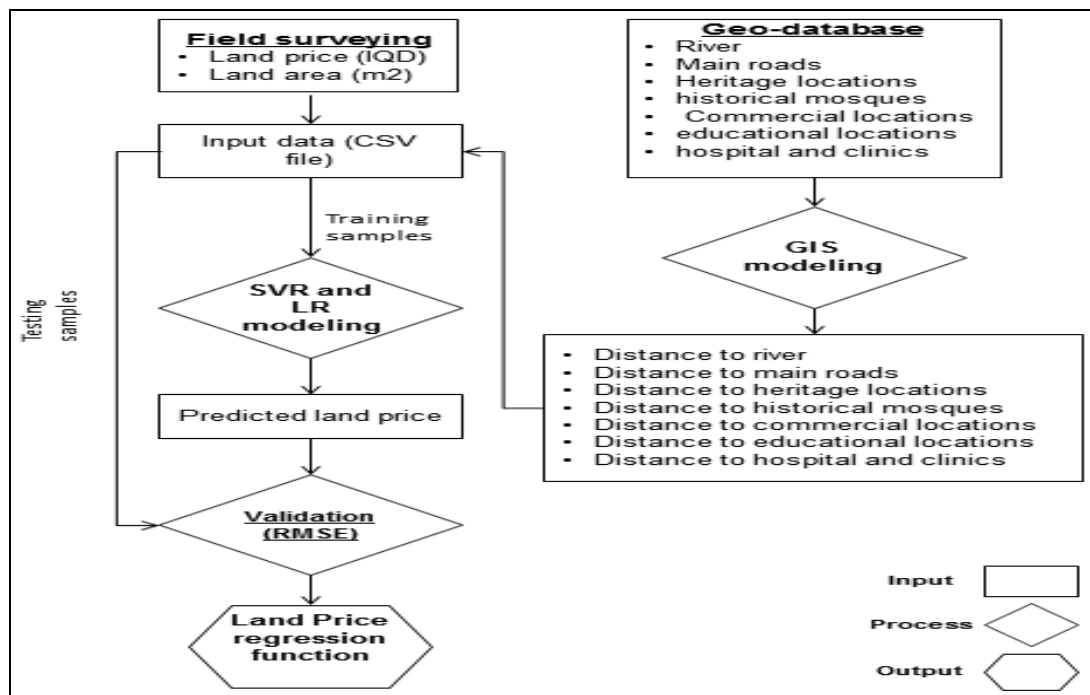

Figure2: Overall methodology

3-3 The Data collection and preparing

A survey was conducted during June 2021 using a questionnaire form that included the land area (square meters) and land price (Iraqi dinar).In addition, a Global Position System (GPS) is used to observe the position of 105 samples from Al-Kufa city. A GPS device model Garmin (eTrex10) collects coordinates based on the projected coordinate system, Universal Transverse Mercator (UTM) projection, Datum WGS84, and Zone 38N. The collected data is converted from hard copy to soft copy by using Microsoft excel program.

3-4 GIS works

A GIS data belonging to Al-Kufa city have been collected from the Governmental agency (Najaf province). The collected data included shape files of (rivers, main roads, heritage locations, historical mosques, commercial locations, educational locations, and hospitals and clinics). First, the data within the GIS environment of ArcGIS software were prepared. Then the spatial analysis was used to extract distance to locations information. The Euclidean distance algorithm (ED) is used for the calculations, where ED refers to the straight distance between two points within a GIS environment.

The ED algorithm refers to the distance from each cell to the closest source (Curriero 2006). The resulting information was extracted as raster format data representing distances to many factors calculated in meters. These factors included distance to a river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, distance to educational locations, and distance to hospitals and clinics).

Then, this information was converted to tabular data based on the extraction of data from raster's to the points of samples by using the extract multi-values to point's tool. Then samples were divided into training sample (68) and testing sample (29) by using Geostatistical analyst tools. Finally, these data were converted to a particular type of information data called comma-separated values (CSV) to be easily used in the estimation analysis

3-5 Regression modeling

Regression algorithms are important models among machine learning models. They are considered supervised models that work based on training data. They are usually used to predict dependent variables based on independent

variables obtained from training data. Regression is utilized to predict values at a continual range like land price values rather than classifying them into

categories like (land and house). Generally, regression technique is categorized into two main types. The first type is the simple regression model, while the second is the multivariate regression model (Fahrmeir 2007). The simple regression models consider only the highest linear correlated independent variable with the dependent variable.

In contrast, multivariate regression models consider all correlated independent variables. In the last decades, many expert models have been developed to optimize the regression analysis in terms of the non-linear relationships that cannot be considered in traditional regression models. Support Vector regression is a machine learning algorithm used to estimate values by considering the linear and non-linear relationships between independent and dependent variables.SVR model is mainly extracted from the Support Vector Machine model used for classification issues by creating the best line to separate data into classes. In contrast, SVR models try to find a line of best fit that decreases a cost function's error. Equation1 illustrates the basic regression calculation based on multiple variables:

$$f(x, y, z) = w1x + w2y + w3z + b \qquad (1)$$

In this function, the coefficients or weights that will be trained by our model are represented by w. While attributes or information of each observation are defined by (x, y, z). And the bias value is represented by b. In general, bias mean the value added to the error of the difference between predicted and observed values. In this study, the researchers will estimate land prices as a dependent variable based on several independent variables (land area, distance to river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, distance to educational locations, and distance to hospital and clinics). WEKA software was used to implement the modeling process (Azeez 2019; 2018).

On the other hand, the proposed model was validated based on the testing data, and root mean square error model (RMSE), which is widely used to calculate the difference between predicted and the reliable values of the modeled subset data. Obviously, this difference refers to the error and impacts the model's accuracy. It is also named as residuals (Willmott, 2005). Equation 2 shows the mathematical calculation of the
RMSE model:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \qquad (2)$$

$Xobs$ = observed values
$Xmodel$ = modeled values
n = samples

4- Results and Discussion
1.1   Result of data collection and GIS modeling

There are two types of data collected. The first type is the data obtained from the field surveying, implemented by questionnaire form, which included land price and land area. On the other hand, the second type is the data extracted from GIS modeling included Euclidean distance to other variables.Table1 shows the summary of data collected. The highest value of the land price was detected at 320 million (IQD), while the minimum value was 35 million (IQD)

On the other hand, the largest land area recorded 400 m^2, while the smallest was 50 m2. According to the data resulting from the spatial analysis, it was found that the longest distance from our samples to the river was 3754.94 m, and the shortest distance was 92.2. The distance to main roads showed that the longest distance to main roads was 552.27 m, and the shortest distance was 10 m. concerning the distance to heritage locations that involved 13 locations, the longest distance was 4762.66 m, and the shortest distance was 20m. Another essential variable is the historical mosque included (Al-Kufa mosque and Al-Sahla mosque). Our analysis also showed that the farthest location of the mosque relative to our samples was 3818.48 m, while the closest mosque's location was 110.4. Furthermore, the farthest distance from our samples to commercial locations was 3841.94 m, the closest was 30m. The distance to educational locations such as schools, universities, and faculties ranged from 3696.23 m to 36.06. Finally, the longest distance to hospitals and clinics is 4314.19 m, and the shortest distance is 125.30 m.

Factors	Min.	Max.	Mean	StdDev
Land price (million)	35.00	320.00	126.75	67.87
Area of land (m2)	50.00	400.00	166.66	78.56
Distance to river (m)	92.20	3754.94	1257.06	889.14
Distance to main roads (m)	10.00	552.27	151.43	114.92
Distance to heritage locations (m)	20.00	4762.66	859.70	963.80
Distance to historical mosques (m)	110.45	3818.48	972.24	707.51
Distance to commercial locations (m)	30.00	3841.94	573.63	828.33
Distance to educational locations (m)	36.06	3696.23	660.81	798.31
Distance to hospital and clinics (m)	125.30	4314.19	835.82	887.50

Table 1: The summary of the data collected

4-2 Result of Regression modeling

In the current study, the researchers trained the SVR and LR models to predict prices of lands within Al-Kufa city based on 9 variables. The land price represents the dependent variable. The other variables are land's area, distance to the river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, distance to educational locations, distance to hospitals and clinics. Both SVR and LR models were operated based on 68 samples. A regression equation has resulted from the regression analysis.

The first regression function considers only the highest correlated independent variable (land area) with the land prices and neglects other independent variables. Because the traditional linear regression only considers linear relationships (see equation 3). Meanwhile, the SVR equation considers all independent variables (see equation 4)

$$\text{Predicted land price} = 0.8 * Area - 4.75 \quad (3)$$

$$
\begin{aligned}
\text{Predicted land price} \\
= +0.7207 * Area - 0.0154 * River - 0.036 \\
* Roads - 0.0135 * heritage - 0.0147 * mosques \\
+ 0.0113 * commercial + 0.0291 * Schools \\
- 0.03 * Hospital + 12.1539 \quad (4)
\end{aligned}
$$

Table 2 shows the result of regression modeling. The results showed that the LR model is achieved an overall accuracy of 66.30, a Correlation coefficient of 0.90, root mean squared error of 13.70. As well as the R-squared value is 0.90, where R-squared shows the correlation strength between observed and predicted values (land prices).On the other hand, SVR achieves an overall accuracy of 75.50, a Correlation coefficient of 0.94, a root mean squared error of 8.90 and an R-squared value of 0.97. Therefore, the modeling results showed that the SVR is outperformed the LR model. The Overall accuracy increased from 66.30 to 75.50. Furthermore, the Correlation coefficient increased from 0.90 to 0.94.

On the other hand, the root means the squared error is decreased from 13.70 million IQD to 8.90 million IQD. Regarding the processing time, results indicated that LR is faster than SVR in terms of processing time. The processing time with LR and SVR is 0.04 seconds and 25.68 seconds, respectively.

Table 2: Results of regression models (Training data)

	Simple Linear Regression (LR)	Support Vector Regression (SVR)
Correlation coefficient	0.90	0.94
Root mean squared error (million)	13.70	8.90
Overall accuracy (%)	66.30	75.50
R-squared value	0.90	0.97
Time of processing / seconds	0.04	25.68
Total Number of Instances	68	68

According to the LR and SVR models assessment, the researchers used (29) testing samples, the resulting equations, and the root mean square model. The overall validation accuracy of LR and SVR are 75.40 and 82.90. While the Correlation coefficient resulting based on the LR model is 0.97, and 0.98 resulted from the SVR model. The root mean square obtained from the LR model is 10.80 million IQD and 7.90 million IQD resulting from the SVR model (see table 3). The R-squared value of LR and SVR are 0.9712 and 0.976, respectively (Figure 3).

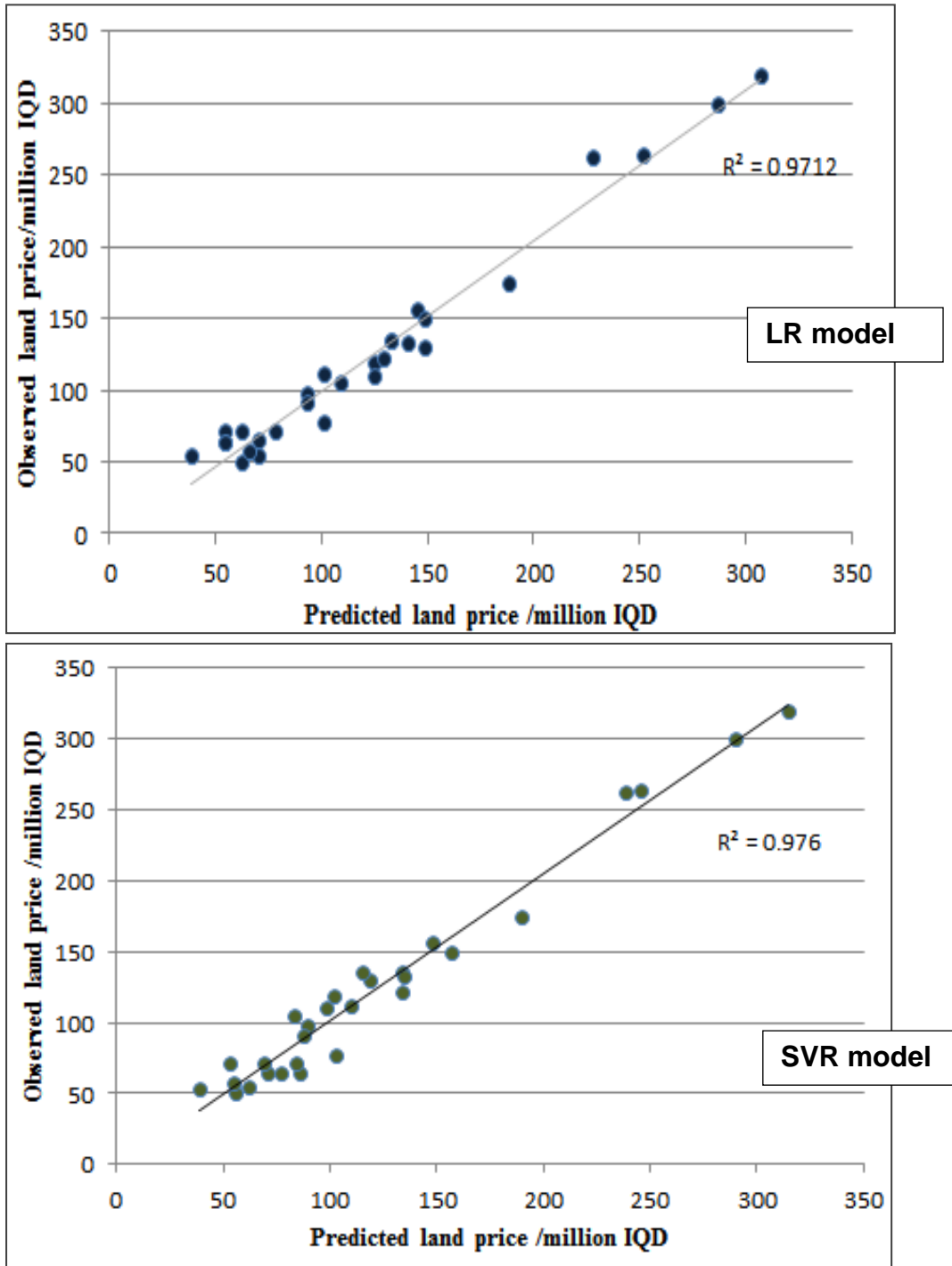|	Simple Linear Regression (LR)	Support Vector Regression (SVR)
Correlation coefficient	0.97	0.98
Root mean squared error (million)	10.80	7.90
Overall accuracy (%)	75.40	82.90
R-squared value	0.971	0.976
Time of processing / seconds	0.01	7.53
Total Number of Instances	31.00	31.00

Table 3: The result of validation

Figure 3: The correlation between observed and predicted land price based
on LR and SVR models

Based on the LR model, the highest difference between predicted and observed land price is 34 million IQD, while the lowest is 1.2 million IQD. On the other hand. Based on the SVR model, the highest difference between observed and predicted land prices is 25.5 million IQD, while the lowest value is 1 million IQD (figure 4). Results of comparison between the LR and SVR models have shown that the SVR model has achieved an accuracy higher than the LR model based on training and testing subsets due to the SVR model's significant ability to consider the non-linear as linear relationships. On the contrary, LR considers only independent variables having a linear correlation with the dependent variables.

On the other hand, SVR can minimize errors during the estimation analysis, making it a preferable model in estimation and forecasting studies. According to the estimation results, several variables can affect the land price within religious cities. The presented models can be optimized to be more generalized for other study areas and increase the number of variables by adding social and environmental indicators such as population density, locations of landfills. Therefore, the presented models can be utilized to assess land prices in holly cities like Al- Kufa. Furthermore, this tool can retain land pricing, land management, and urban planning in Iraq.
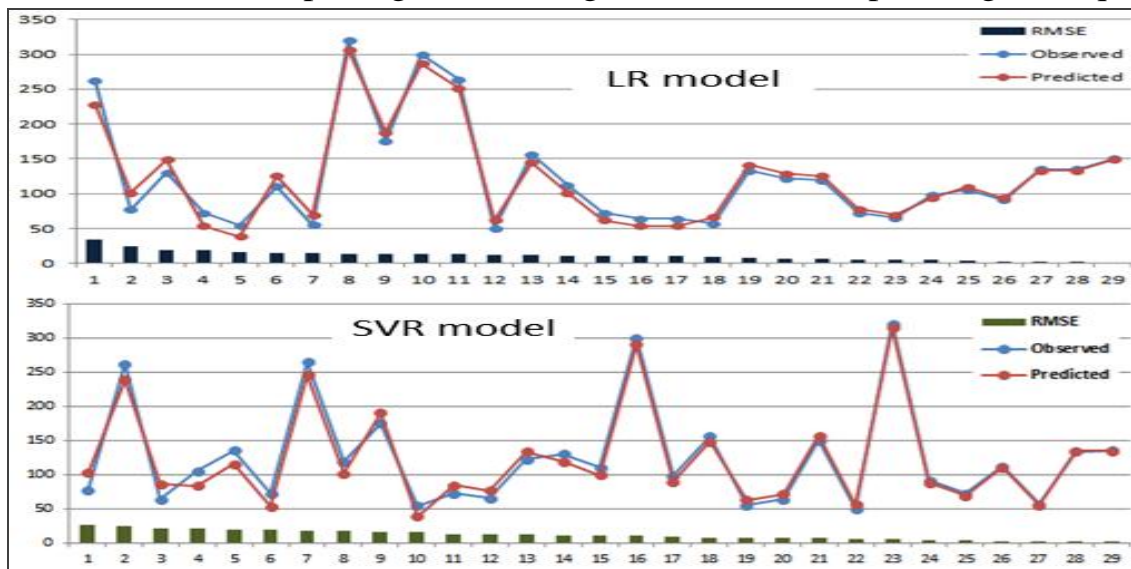
Figure3: The differences between observed and predicted land price per million IQD based on LR and SVR models

5- Conclusion

Land price is considered as the primary indicator of the land market in urban areas. Land prices in religious cities such as Al-Kufa city have rapidly increased due to religious and tourism activities. Governments are usually facing challenges in assessing the land prices in these cities. Therefore, they need expert models or tools to obtain an overview of land prices within religious cities. Predicting land prices can efficiently retain future management and develop urban lands within religious cities. In this study, the researchers proposed a methodology based on two models, Linear Regression (LR) and Support Vector Regression (SVR), and nine variables (land price, land area, distance to a river, distance to main roads, distance to heritage locations, distance to historical mosques, distance to commercial locations, distance to educational locations, and distance to hospital and clinics. Our findings showed that the SVR model had outperformed the LR model, where SVR has achieved an accuracy of 82.9%. In comparison, LR has achieved 75.40%. Therefore, the presented models can assess land prices in holly cities like Al-Kufa. Furthermore, this tool can retain land pricing, land management, and urban planning in Iraq.

Acknowledgments

References

Wu, D. J., Liu, Y. L., & Huang, Q. Y. (2007). An Artificial Neural Network Appraisal Model for Urban Base Land Price. Journal of Zhengzhou Institute of Surveying and Mapping, 24(4), 255-258.

Wang, Z., Fu, M., Yu, Y., & Zhao, J. (2011). Prediction of urban land price based on Grey-Markov model. In Proceedings of 2011 International Conference on Computer Science and Network Technology (Vol. 2, pp. 708-712). IEEE.

Jamal Motlak, (2016), "Analysis and evaluation of factors affecting housing unit supply", Journal of the planner and development (vol.21, pp.292).

Sati Sana (2012). "Dynamics of Urban Growth in Iraq", Journal of the planner and development (vol.26, pp.90).

Orbasli, A. (2002). Tourists in historic towns: Urban conservation and heritage management. Taylor & Francis.

Raeda, a. D., & Asadi, s. (2020). Strategic urban planning of religious cities: the case study of karbala city in iraq. Вестник мгсу, 15(8).

Mohammed H. R., A Alkanaani, H. & Alowaid, H.k," Using GIS technology for land suitability analysis to select drainage project location: Nasiriya city south of Iraq as a case study", Journal of Physics: Conference Series, 2021, 1895(1), 012001.

Azar, K. T., Ferreira Jr, J., Abed, J., & Kazzaz, T. (1994). Using GIS for estimating the spatial distribution of land value in metropolitan Beirut. URISA. http//www. sgi. ursus. maine. Edu/gisweb/spatdb/urisa/ur9072. html. Accessed, 3(10), 03.

Luo, Jun, and Yehua Dennis Wei. "A geostatistical modeling of urban land values in Milwaukee, Wisconsin." Geographic Information Sciences 10.1 (2004): 49-57.

Löchl, M. (2006). Real estate and land price models for UrbanSim's Greater Zurich application. Arbeitsberichte Polyprojekt Zukunft urbane Kulturlandschaften, 6.

Lee, S. H., Kim, J. H., & Huh, J. H. (2021). Land Price Forecasting Research by Macro and Micro Factors and Real Estate Market Utilization Plan Research by Landscape Factors: Big Data Analysis Approach. Symmetry, 13(4), 616.

Choi, J., & Kim, Y. (2021). Prediction of the land price based on deep learning and residual kriging. The Korean Data & Information Science Society, 32(3), 475-485.

Lee, C. (2021). Predicting land prices and measuring uncertainty by combining supervised and unsupervised learning. International Journal of Strategic Property Management, 25(2), 169-178.

Tsutsumi, M., Shimada, A., & Murakami, D. (2011). Land price maps of Tokyo metropolitan area. Procedia-Social and Behavioral Sciences, 21, 193-202.

Al-Duhaidahawi, Z. S., Almuhanna, R. R., Abdabas, A. Y., & Al-Jameel, H. A. (2020). Traffic Assignment of Al-Kufa City Using Trans CAD. In IOP Conference Series: Materials Science and Engineering (Vol. 978, No. 1, p. 012016). IOP Publishing.

Curriero, F. C. (2006). On the use of non-Euclidean distance measures in geostatistics. Mathematical Geology, 38(8), 907-926.

Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2007). Regression. Springer-Verlag Berlin Heidelberg.

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30(1), 79-82.

Azeez, O. S., Pradhan, B., Shafri, H. Z., Shukla, N., Lee, C. W., & Rizeei, H. M. (2019). Modeling of CO emissions from traffic vehicles using artificial neural networks. Applied Sciences, 9(2), 313.

Azeez, O. S., Pradhan, B., & Shafri, H. Z. (2018). Vehicular CO emission prediction using support vector regression model and GIS. Sustainability, 10(10), 3434.

Wang, D., Li, V. J., & Yu, H. (2020). Mass appraisal modeling of real estate in urban centers by geographically and temporally weighted regression: a case study of Beijing's core area. Land, 9(5), 143.

Wang, C. H., & Chen, N. (2017). A geographically weighted regression approach to investigating the spatially varied built-environment effects on community opportunity. Journal of transport geography, 62, 136-147.

Cellmer, R., Kobylińska, K., & Bełej, M. (2019). Application of hierarchical spatial autoregressive models to develop land value maps in urbanized areas. ISPRS International Journal of Geo-Information, 8(4), 195.

Reda Yaagoubi, Kamil Faisal, Yehia Miky (2021). Land value assessment based on an integrated Voronoi-geographically weighted regression approach in Makkah, Saudi Arabia, Journal of King Saud University - Science, Volume 33, Issue 7.