On the Prediction Performances of SVM and ANN Methods for Mass Appraisal Assessment: A Case Study from Ankara (Turkey)

Seckin YILMAZER and Sultan KOCAMAN, Turkey

Key words: Land Valuation, Mass appraisal, Machine learning, Support vector machine, Artificial Neural Networks

SUMMARY

Real estate is an important economic factor in many countries. The measurability of the size of the real estate sector plays an effective role in determining the gross national product of the countries and thus their economic scale. The real estate valuation activities are in general time-consuming and labor-intensive processes. In developed countries, mass and automatic valuation property approaches are often preferred to individual valuation based on conventional methods, especially when the outputs are utilized for the real estate tax calculations. The recent developments on data-driven machine learning methods facilitate the mass valuation studies even though the data quality issues introduce vulnerability to the processes. In this study, two commonly used machine learning methods, namely support vector machine (SVM) and artificial neural networks (ANN), were employed for the mass valuation of 2850 data samples composed of mostly residential properties and with few commercial ones. The study area is located in Mamak District of Ankara, Turkey, which is an urban expansion area. The property features and the price information were collected in a pilot project, which was previously carried out by the General Directorate of Land Registry and Cadastre, Turkey. According to the accuracy assessment results, the total variance explained from the ANN model was R^2 of 0.84 with an RMSE of 14.805 TRY. The SVM results yield to $R^2 = 0.76$ and RMSE = 16.397 TRY. Thus, ANN slightly outperforms SVM for the study dataset. The results show that the data-driven methods have the potential for modeling the real estate prices with a certain confidence level when reliable training data are available. The data, methodology and the results of the study are presented and discussed in this paper.

On the Prediction Performances of SVM and ANN Methods for Mass Appraisal Assessment: A Case Study from Ankara (Turkey)

Seckin YILMAZER and Sultan KOCAMAN, Turkey

1. INTRODUCTION

Despite the recent stagnation in the gross domestic product (GDP) values in European countries, the increase in real estate (RE) prices continued during the pandemic process (Deloitte, 2020). The share of housing investments within the overall investment scheme has also continued to increase in Turkey (Table A1 in Appendix, T.C.C. Strateji ve Bütçe Başkanlığı, 2021). Its size is even larger when associated sectors are taken into account. The rights and privileges granted to the RE owners have drawn attention of foreign capital to invest in this sector in Turkey, and facilitated the increase in the prices. Proper handling and registration of property prices have great importance for developing economies, as in Turkey, for various reasons such as the calculation of economic sizes, preparation of taxing schemes, transparency in the processes and financial transactions, etc.

The RE valuation is an important tool to provide a quantitative value of the properties for monitoring the sector and applying correct taxing schemes; and to measure the economic size of immovable assets. Occasionaly, the RE valuation service is required to validate mass properties quickly and reliably (Grover, 2016). Examples to such cases can be given for expropriation, enforcement bankruptcy, RE sales with auction, stock market instruments for RE shares, etc. In Turkey, a widely observed case is expropriation activities, for which rapid valuation is crucial to be able to start the planned project as soon as possible. The outcomes of the RE valuation are useful both for institutions and individuals, such as for the case of a simple house rental. Besides, the property taxes have an effect on almost all stakeholders.

While conventional single valuation methods are successfully used by experts for the value assessment of a single or a small number of properties; mass apprasial methods are required for the valuation of large numbers of REs. Although a number of mass appraisal approaches exists in the literature, such as Computer Assisted Mass Appraisal (CAMA) (McCluskey et al., 2013), data-driven machine learning (ML) based methods (e.g. Rafiei and Adeli, 2016; Yilmazer and Kocaman, 2020), and fuzzy logic (Bagnoli and Smith, 1998); no standardized procedures are currently available. The main requirement of the data-driven ML-based methods is the availability of reliable and accurate data, which directly affect the quality of the mass appraisal results (Crosby, 2000; Abidoye and Chan, 2018). Obtaining high quality data can be laborious, time-consuming and expensive. On the other hand, with the increase of such studies, the disadvantages can be largely overcome; and methodological aspects can also be clarified. The interpretability (Christensen and Sørensen, 2014; Liu et al., 2020), transparency (Eichholtz et al., 2011; Lind and Nordlund, 2014; Schulte et al., 2005), and the flexibility of the methods are also crucial for proper implementation and optimizing the prediction parameters which are effective and significant for the selected area.

In this study, two data-driven ML approaches were applied for the mass appraisal of a number of properties in a part of Mamak District of Ankara, Turkey. The area was selected for the variability in property object types and their prices. Mamak is an urban expansion area with diverse socio-economic conditions; and besides the old and newly constructed buildings, it has slum areas and urban transformation projects recently implemented in some parts. The artifical neural networks (ANN) and support vector machine (SVM) regression methods were implemented for the purposes of the study and applied to a reliable dataset with various property characteristics and prices. The total number of samples in the dataset were 2.850. Further details on the study area, the data description, the methodology and the results are provided in the following sections.

2. STUDY AREA AND DATA DESCRIPTION

Mamak District is among the rapidly growing districts of Ankara and located in 3.5 km away from Ankara City Center. The district has currently the fourth largest population size of Ankara. It has became a center of attraction with the urban transformations and large-scale restoration works carried out during the last 10 years. Mamak is surrounded by three major districts (Altindag in the North, Elmadag in the East, Cankaya and Elmadag in the South, Çankaya and Altindag in the West) as shown in Figure 1. It was observed that the proximity to the district boundaries also have an impact on the socio-economic structure of Mamak. While the housing prices are higher closer to the borders of Cankaya; the properties have lower economic values in the urban transformation areas near Kayas part of Mamak and Elmadag. According to the Land Registry and Cadastre Information System of Turkey (TAKBIS), Mamak District is ranked in the top five in Ankara for the land registry transactions in the period of Jan 2019-Jan 2020. The transaction types and their rates are shown in Figure 2. According to the land registry, there are a total of 326.000 independent sections subject to title deeds; which contains 291.000 residential and 22.000 commercial units. The remaining units can be classified as others.

In this study, the data were collected within the mass RE valuation pilot project carried out by the General Directorate of Land Registry and Cadastre (GDLRC) between 2010 and 2018 for various locations including Mamak District (GDLRC, 2014). Further verifications and processing were applied to the collected data within this study to increase the reliability and to ensure the temporal accuracy. For this purpose, court reports, data from valuation institutions and municipalities were obtained and analyzed. As a result, a dataset consisting of 40 variables (see Table A2 in Appendix) were obtained for a total of 3.400 RE properties. After the sectoral and regional analyses and careful investigations on the data samples and variables by the expert (first author); 20 of the variables were omitted from the model. The statistical R^2 and variable importance tests were also applied to the data; and 550 of the samples were determined as outliers and removed from the dataset. Finally, 2.850 samples were used in the study with 20 explanatory variables (shown in Table A3 in Appendix) for the modelling.

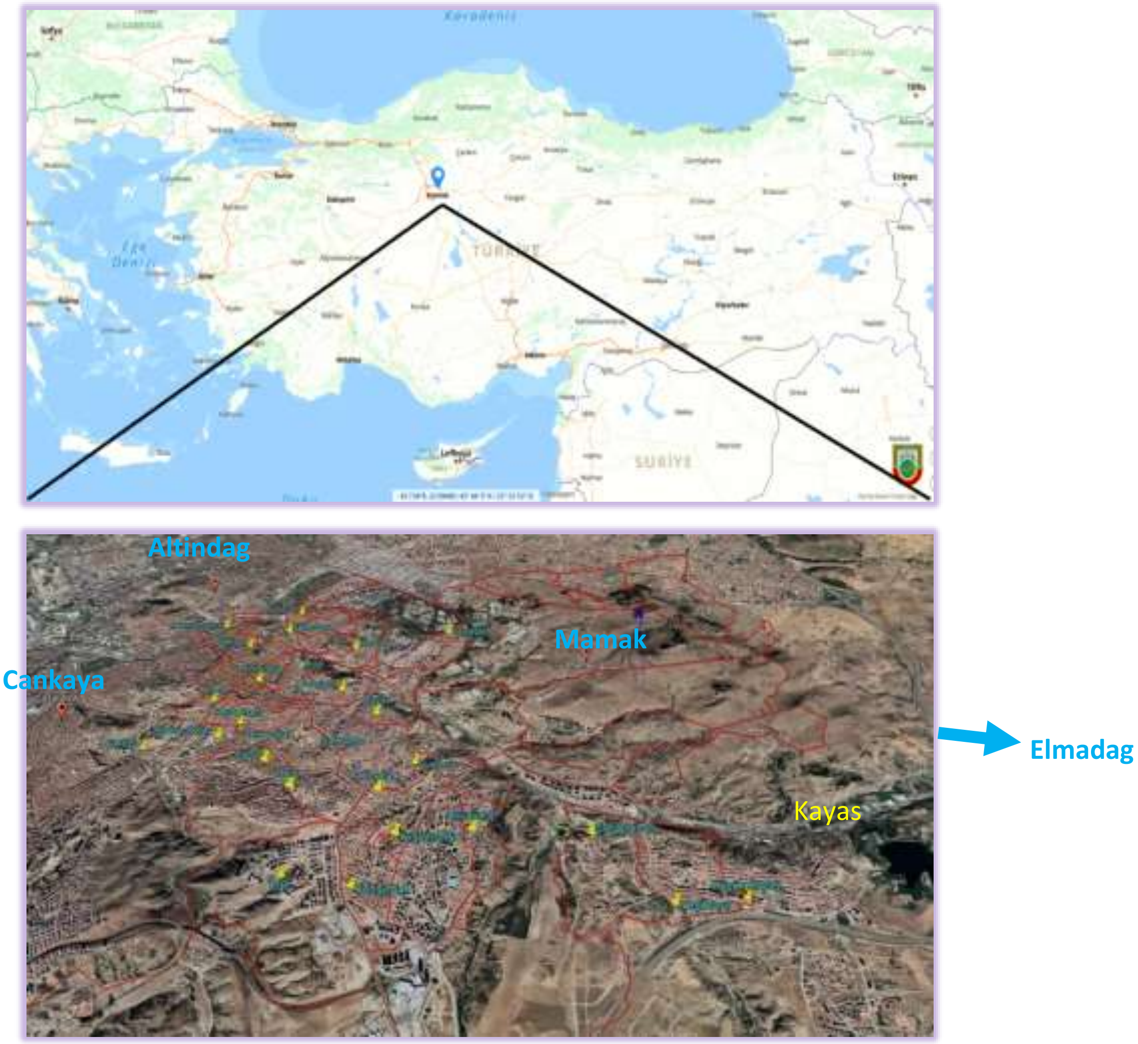

Figure 1. The location of the study area lat: 39.5509° lon: 32.5612° at the center (reference map: General Directorate of Mapping (GDM), Turkey; satellite image credit: Google Earth).

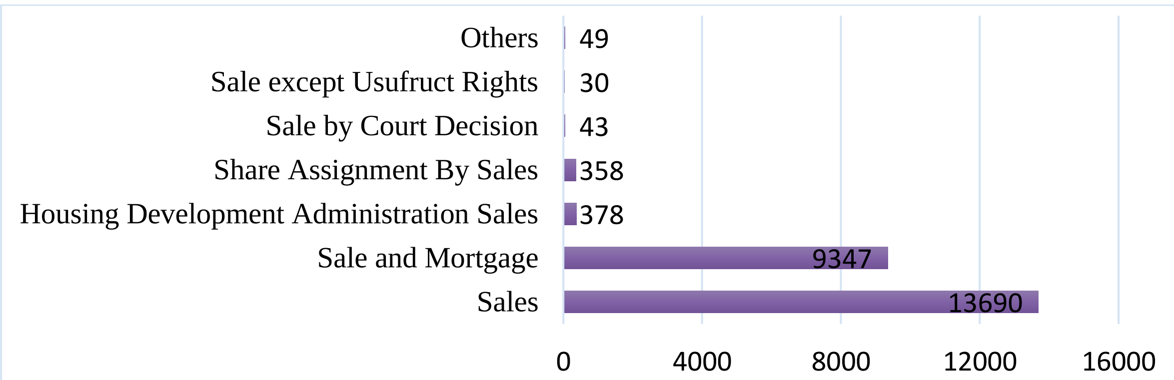

Figure 2. The rates of Land Registry Transaction Types in Mamak District, Ankara, between Jan 2019-Jan 2020.

3. METHODOLOGY

In this study, the mass appraisal for the Mamak District was carried out using a total of 2.850 samples and the ANN and the SVM regression methods. The methodological details and the number of samples used for the model training, validation and testing are provided in the following sub-sections.

3.1 Artificial Neural Networks (ANN)

The ANN is an information and decision support system developed to imitate the characteristics of human brain to storage and process information, and use it in decision-making processes (Rosenblatt, 1958). ANNs are inspired by the data processing tecnique of the human brain (Worzala et al., 1995). The algorithm have been used by many disciplines (e.g. Dayhoff and DeLeo, 2001; Olden and Jackson, 2002), and preferred in the mass appraisal studies as well (e.g. Abidoye and Chan, 2017; Hamzaoui and Perez, 2011) due to its abilities to produce accurate results by dealing with complex data structures, to contribute to the solution using hidden layers, to successfully use back propagation and error minimization techniques, and to catch nonlinear interactions between variables (Goh, 1995). Besides the positive aspects of the ANN method, there are also some shortcomings. Since it is a black-box method (Dayhoff and DeLeo, 2001; Olden and Jackson, 2002; McCluskey et al., 2013) and produces results with hidden layers; both the model and the obtained results cannot be explained transparently. Moreover, the training time can be lengthy and the model can easily encounter overfitting or can generate incorrect local solutions.

The target is the sale price of the REs, while the independent variables used here can be categorized as residential unit features, building features, and zoning area features (Table A3). The study dataset was split into three subsets randomly for utilizing in the ANN method. As shown in Table 1, 70%, 15% and 15% of the data samples were employed for training, validation and test, respectively. As a part of the ANN implementation in the Matlab software (Mathworks, 2021), within the back propagation process, actual data and estimated data were compared, and the initial parameters were updated iteratively. In back propagation based methods, such as feed forward back propagation, the success of the improvement process varies according to the reliability and accuracy of the dependent and independent variables. The basic schematic diagram of the feed forward backpropagation in the ANN structure implemented in this study is depicted in Figure 3. The parameters used here are also shown in the Table 1.

Table 1. Data split and hyperparameterization in the applied ANN process.

Training algorithm	Feed Forward Back Propagation
Number of hidden layers	10
Number of training samples	2010 (70% of all samples)
Number of validation samples	420 (15% of all samples)
Number of test samples	420(15% of all samples)
Number of iterations	15

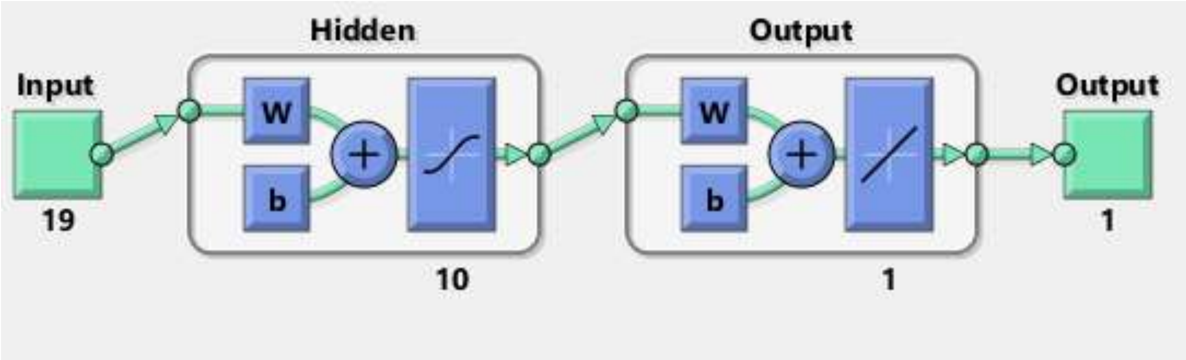

Figure 3. Basic schematic diagram of feed forward backpropagation in the ANN implementation used here.

3.2 Support Vector Machine (SVM) Regression

The SVM Regression (SVM-R) is a ML algorithm used for maximizing mathematical function according to the given data (Noble, 2006). Although it is generally used in classification studies, it can also be used for the regression problems as well (Lam et al., 2009). The SVM-R is a comparative method, which seeks the local minima (Shin et al., 2005; Lam et al., 2009). The method was used in a number of RE appraisal studies in the literature (e.g. Lin and Chen, 2011; Yacim and Boshoff, 2020).

The SVM-R can be used both for simple linear regression problems and for explaining non-linear problems with a kernel function. The linear SVM-R can solve the problems successfully if the independent variables are in linear interaction with the target; but may fail in complex and multidimensional input space such as RE data. The kernel function converts low-dimensional data to a higher dimensional space. According to Noble (2006), the best approach for the selection of the kernel function in most cases is to use trial and error. Within the scope of this study, after using trial and error of kernel functions, the Gaussian Radial Basis Function (RBF) Kernel Based SVM-R was found optimal and applied to the data. The method provides non-linear mapping, finds the infinite size support vectors, and calculates the similarity of each sample to a certain point with the normal distribution. The data split ratio and the hyperparameterization used for SVM-R here is provided in Table 2. In SVM-R, the cross validation ensures adequate number of samples via best fit margin between the target and the independent variables. Therefore 2.430 training samples were used in 5-fold cross validation and 420 of samples were used for test (Table 2). The training and test samples were selected randomly. During the model building phase, the difference between each prediction and actual value was tried to be minimized by adjusting C and γ coefficients. Finally, the iteration stops when the adjusted convergence criterion rate or value is achieved according to the specified algorithm. The box contraints were selected automatically and the number of iterations was limited to 15 for this study; since the use of RBF increased the training time, which may also lead to over-training.

Table 2. Data split and hyperparameterization in the applied SVM Regression method.

Training algorithm	K=5 Cross Validation
Kernel	Gaussian – Radial Basis Function
Kernel Scale	4.4
Epsilon	Automatic
Box Constraints- C	Automatic
Number of training samples	2430 (85% of all samples)
Number of test samples	420 (15% of all samples)
Number of iterations	15

3.3 Performance Validation

The results of the ANN and SVM-R methods were compared here using R^2 and root mean square error (RMSE) as performance metrics. R^2 is the total variance explained from the model and measures the correlation between the outputs and the targets. The RMSE is a measure of deviation of the estimated values from the observed values. The formulas are provided in Equations 1 and 2.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\,Actual\ Price_i - \ Estimated\ Price_i)^2}{N}} \qquad \textit{Equation 1}$$

$$R^2 = 1 - \left(\frac{MSE}{\sigma y^2}\right) \qquad \textit{Equation 2}$$

where;

- N is the number of cases,
- MSE is the mean square error (square of RMSE)
- σ_y is the standard deviation obtained from the sample set y

4. RESULTS AND DISCUSSION

In the present study, a mass appraisal was conducted with two ML methods, i.e. ANN and SVM-R methods, on a dataset consisting of 20 independent variables and 2.850 data samples in the Mamak District of Ankara, Turkey. In ANN, the best validation performance was obtained at the 9th epoch with an $MSE = 2.85x10^8$ as shown in Figure 4. The best fitting graphs for training, validation, and test data and also for overall are given in Figure 5. The SVM-R results are presented in Figure 6.

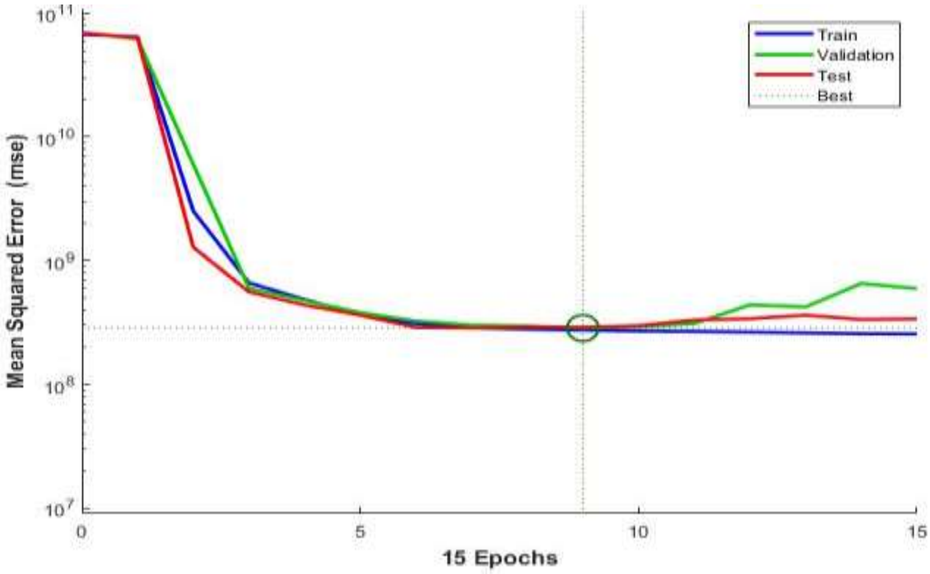

Figure 4. ANN validation performances at different epochs for the training, test and validation datasets. The best validation performance was achived at the 9th epoch (MSE =$2.85x10^8$).

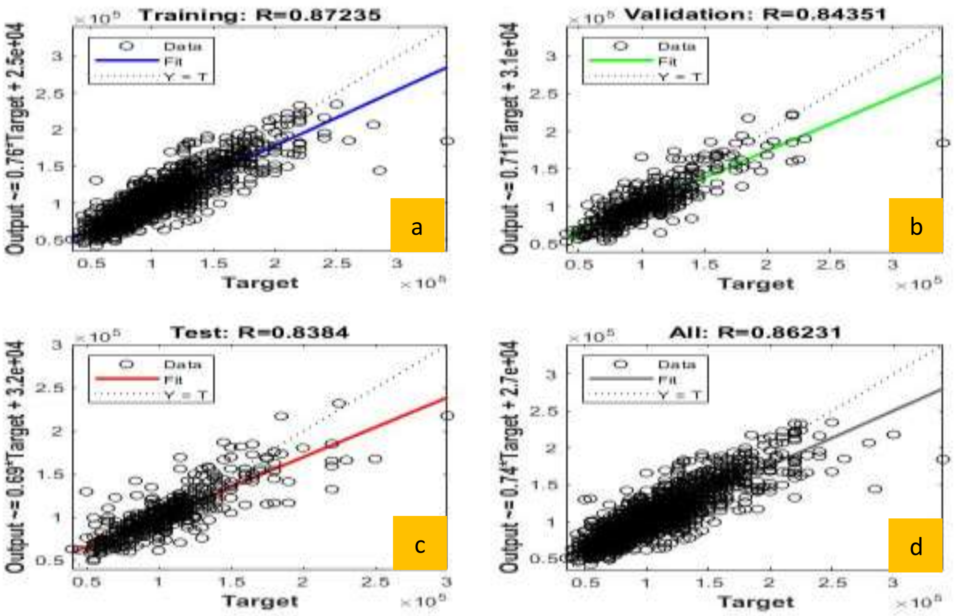

Figure 5. The fitting graphs for (a) training, (b) validation, (c) test, and (d) overall data obtained from the ANN.

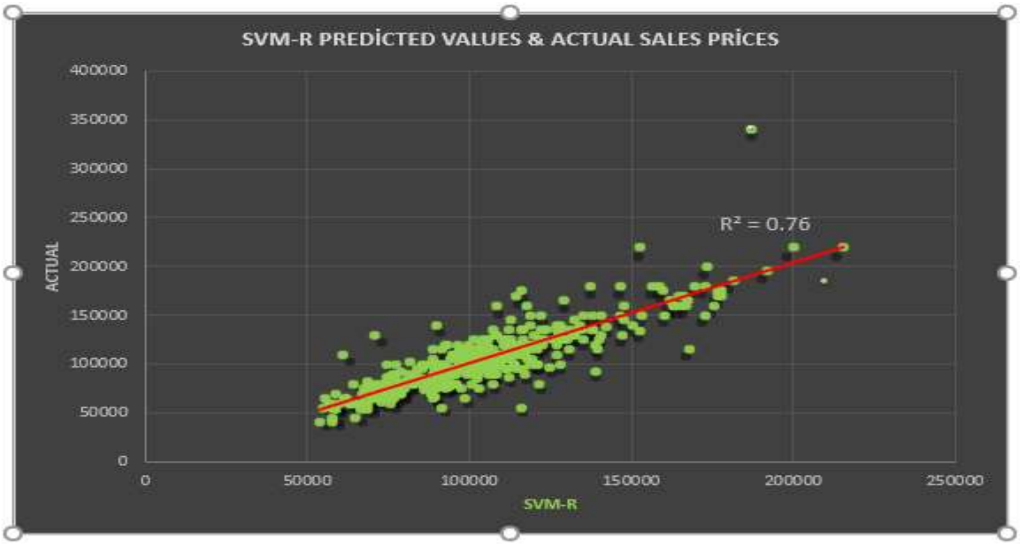

Figure 6. The predicted and actual prices distrubution graph obtained from the SVM-R method.

When the results obtained from the ANN and the SVM-R are compared, the total variance explained from the ANN model was found better (i.e. $R^2 = 0.84$) with an RMSE of 14.805 TRY. The SVM-R results yield to $R^2 = 0.76$ and RMSE = 16.397 TRY. Thus, the ANN slightly outperforms SVM-R for the study dataset as shown in Figure 7. On the other hand, acceptable and comparable outputs were obtained from both methods. However, besides the positive aspects of both methods, some concerns can be raised regarding the model building phase. In particular, explaining the model was difficult and complicated. In addition, since 10 hidden layers were used in the training phase of ANN, the model could not be explained efficiently and considered as black-box.

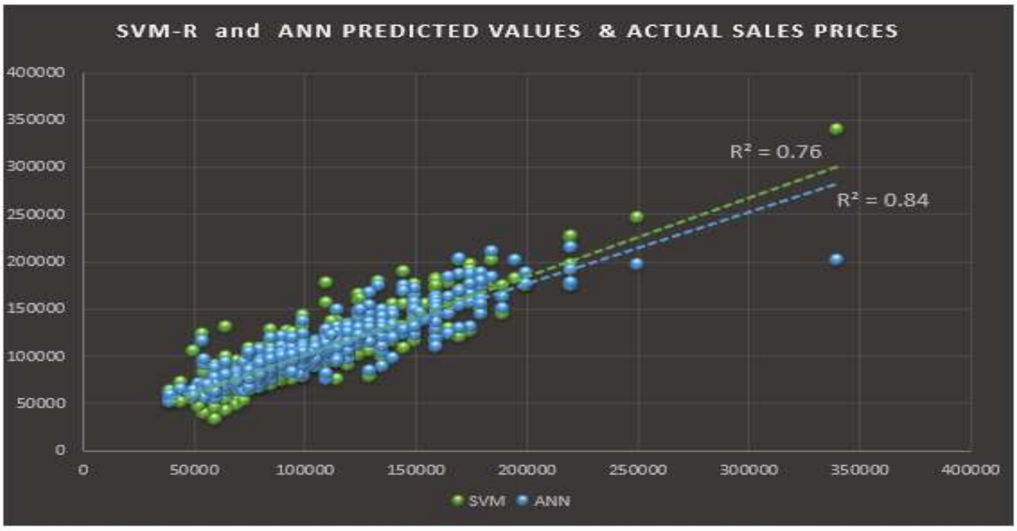

Figure 7. Distrubution graph for SVM-R and ANN predicted values & actual sales prices

5. CONCLUSIONS

In the present study, a set of residential and commercial properties were appraised using two ML methods, ANN and SVM Regression. The study aimed at investigating the mass appraisal potential of ML algorithms in complex and high dimensional data structures. The physical and legislative attributes of the properties together with their prices were collected in a pilot project by GDLRC, Turkey; and verified within the current study by further analysis and testing. The study dataset includes 2.850 data samples with 20 independent variables (Table A3), which were reduced from the initial list with 40 attributes (Table A2). According to the results, both the ANN and the SVM Regression methods provided acceptable and comparable prediction performances. The better R^2 value obtained from the ANN and the SVM regression methods were 0.84 and 0.76, respectively. The main issues regarding the applied methods are the transparency of the model parameters and the complexity in training. In addition, the applicability of the methods in other areas with different geographical, urban and socio-economic characteristics is subject to investigation as future work.

ACKNOWLEDGEMENTS

The authors sincerely thank to Pilot Appraisal Project Team of GDLRC for collecting and providing the data.

REFERENCES

Abidoye, R.B., Chan, A.P., 2017. Artificial neural network in property valuation: application framework and research trend. *Property Management,* 35(5), 554-571. DOI: 10.1108/PM-06-2016-0027

Abidoye, R.B., Chan, A.P., 2018. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pacific Rim Property Research Journal* 24, 71-83. DOI: 10.1080/14445921.2018.1436306

Bagnoli, C., Smith, H., 1998. The theory of fuzz logic and its application to real estate valuation. *Journal of Real Estate Research* 16, 169-200. DOI: 10.1080/10835547.1998.12090941

Christensen, B., Sørensen, T.V., 2014. Interpretability versus out-of-sample prediction performance in spatial hedonic models. European Real Estate Society (ERES). Available at: https://eres.architexturez.net/system/files/pdf/eres2014_96.content.pdf (accessed on 28 Feb 2021)

Crosby, N., 2000. Valuation accuracy, variation and bias in the context of standards and expectations. *Journal of Property Investment & Finance,* 18(2), 130-161. DOI: 10.1108/14635780010324240

Dayhoff, J.E., DeLeo, J.M., 2001. Artificial neural networks: opening the black box. *Cancer: Interdisciplinary International Journal of the American Cancer Society* 91, 1615-1635. DOI: 10.1002/1097-0142(20010415)91:8+<1615::AID-CNCR1175>3.0.CO;2-L

Deloitte, 2020. Property Index: Overview of European Residential Markets, 9th Edition, July 2020. Available at: https://www2.deloitte.com/content/dam/Deloitte/at/Documents/presse/deloitte-property-index-2020.pdf (accessed on 28 Feb 2021).

Eichholtz, P.M., Gugler, N., Kok, N., 2011. Transparency, integration, and the cost of international real estate investments. *The Journal of Real Estate Finance and Economics* 43, 152-173.

GDLRC, 2014. Pilot Mass Appraisal Report of Mamak. Technical Report. General Directorate of Land Registry and Cadastre, Ankara.

Goh, A.T., 1995. Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering* 9, 143-151. DOI: 10.1016/0954-1810(94)00011-S

Grover, R., 2016. Mass valuations. *Journal of Property Investment & Finance.* Vol. 34 No. 2, pp. 191-204. https://doi.org/10.1108/JPIF-01-2016-0001

Hamzaoui, Y.E., Perez, J.A.H., 2011. Application of artificial neural networks to predict the selling price in the real estate valuation process. Proceedings of the 2011 10th Mexican International Conference on Artificial Intelligence, November 2011, pp. 175-181. DOI: 10.1109/MICAI.2011.14

Lam, K.C., Yu, C., Lam, C.K., 2009. Support vector machine and entropy based decision support system for property valuation. *Journal of Property Research* 26, 213-233. DOI: 10.1080/09599911003669674

Lin, H., Chen, K., 2011. Predicting price of Taiwan real estates by neural networks and support vector regression, Proc. of the 15th WSEAS Int. Conf. on Syst*, pp. 220-225.*

Lind, H., Nordlund, B., 2014. A Transparent Two-Step Categorization of Valuation Methods. *Appraisal Journal 82.*

Liu, Z., Cao, J., Xie, R., Yang, J., Wang, Q., 2020. Modeling Submarket Effect for Real Estate Hedonic Valuation: A Probabilistic Approach. *IEEE Transactions on Knowledge and Data Engineering.* DOI: 10.1109/TKDE.2020.3010548

Mathworks, 2021. Matlab. https://www.mathworks.com/products/matlab.html (accessed on 28 Feb 2021)

McCluskey, W.J., McCord, M., Davis, P., Haran, M., McIlhatton, D., 2013. Prediction accuracy in mass appraisal: a comparison of modern approaches. *Journal of Property Research 30, 239-265.*

Noble, W.S., 2006. What is a support vector machine? *Nature biotechnology* 24, 1565-1567. DOI: 10.1038/nbt1206-1565

Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: a randomization approach for understanding variable contributions in artificial neural networks. *Ecological modelling* 154, 135-150. DOI: 10.1016/S0304-3800(02)00064-9

Rafiei, M.H., Adeli, H., 2016. A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management 142, 04015066.* DOI: 10.1061/(ASCE)CO.1943-7862.0001047

Rosenblatt,F.,1958.The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain. *Journal of Psychological Review*, 65(6), 386-408. DOI: 10.1037/h0042519

Schulte, K.W., Rottke, N., Pitschke, C., 2005. Transparency in the German real estate market. *Journal of Property Investment & Finance* 23(1), 90-108. DOI: 10.1108/14635780510575111

Shin, K. S., Lee, T. S., & Kim, H. J., 2005. An application of support vector machines in bankruptcy prediction model. *Expert systems with applications* 28(1), 127-135. DOI: 10.1016/j.eswa.2004.08.009

T.C.C. Strateji ve Bütçe Başkanlığı, 2021. Report on Gross Fixed Investments by Sectors. *Technical Report.* Available at: https://www.sbb.gov.tr/temel-ekonomik-gostergeler/#1594716589132-d3a64e97-2238 (accessed on 28 Feb 2021).

Worzala, E., Lenk, M., Silva, A., 1995. An exploration of neural networks and its application to real estate valuation. Journal of Real Estate Research, 10(2), 185-201. DOI: 10.1080/10835547.1995.12090782

Yacim, J.A., Boshoff, D.G.B., 2020. Neural networks support vector machine for mass appraisal of properties. *Property Management* 38(2), 241-272. DOI: 10.1108/PM-09-2019-0053

Yilmazer, S., Kocaman, S., 2020. A mass appraisal assessment study using machine learning based on multiple regression and random forest. *Land Use Policy* 99, 104889. DOI: 10.1016/j.landusepol.2020.104889

APPENDIX

Table A1. Gross Fixed Investments By Sectors (T.C.C. Strateji ve Bütçe Başkanlığı, 2021)

Share in Total (1)		Percentage Change (2)		
2019 (3)	2020 (4)	2019 (3)	2020 (4)	
				Public Sector
6.0	5.1	-50.0	-27.3	Agriculture
4.2	2.7	12.7	-44.0	Mining
0.6	0.8	13.9	11.2	Manufacturing
6.7	10.6	31.6	36.9	Energy
40.1	37.1	-28.8	-20.5	Transport. & Commun.
0.3	0.3	-67.5	-5.1	Tourism
1.1	1.4	-28.4	13.2	Housing
11.1	10.3	-19.4	-20.3	Education
5.6	5.1	-11.4	-22.3	Health
24.4	26.8	-39.1	-5.6	Other Services
11.1	11.2	-30.6	-13.7	Economic
13.3	15.6	-44.8	1.1	Social
100.0	100.0	-28.7	-14.0	**Total**
141,307	132,837	**49325.2806**	**42419.0749**	**Total - Million TRY**
				Private Sector
1.2	1.5	16.0	36.0	Agriculture
2.1	2.0	9.0	12.4	Mining
23.8	23.8	-2.8	13.0	Manufacturing
1.2	1.2	-16.7	12.0	Energy
29.9	29.7	-9.5	13.6	Transport. & Commun.
3.2	3.9	32.0	40.0	Tourism
31.1	30.6	-10.1	7.5	Housing
2.1	2.1	-9.4	9.1	Education
2.8	2.6	-6.7	5.3	Health
2.6	2.6	-9.9	12.0	Other Services
100.0	100.0	-7.0	12.1	**Total**
1,029,706	1,263,030	**401963.209**	**450681.15**	**450131.8529**

Source: Presidency of the Rebuplic of Turkey Presidency of Strategy and Budget

(1) At Current Prices
(2) 2009=100 Chained Volume
(3) Realization Estimate
(4) Programme

Table A2. The initial list of variables prepared by the expert.

Nmb	*ID*	*Variable Descriptive*	*Variable Type - In/Out*
1	IU-ID	'IndividualUnit_TitleID'	Descriptive - Out
2	BALC	'Balcony'	Independent - In
3	BL	'BuildingLicence'	Independent - In
4	D-SLN	'DependentSaloonOrNot'	Independent - In
5	D-STTS	'DevelopingStatus'	Independent - In
6	DIS-BAZ	'DistanceToBazaar'	Independent - In
7	DIS-BUS	'DistanceToBusStop'	Independent - In
8	DIS-CEN	'DistanceToCenter'	Independent - In
9	DIS-COL	'DistanceToCollege'	Independent - In
10	DIS-CUL	'DistanceToCulturalArea'	Independent - In
11	DIS-GAR	'DistanceToGarbageCollection'	Independent - In
12	DIS-HOS	'DistanceToHospital'	Independent - In
13	DIS-ROD	'DistanceToMainRoad'	Independent - In
14	DIS-MAL	'DistanceToMall'	Independent - In
15	DIS-MRK	'DistanceToMarket'	Independent - In
16	DIS-MOS	'DistanceToMosque'	Independent - In
17	DIS-SCHL	'DistanceToPimarySchool'	Independent - In
18	DIS-REC	'DistanceToRecreationalArea'	Independent - In
19	DIS-UNI	'DistanceToUniversity'	Independent - In
20	ELVTOR	'Elevator'	Independent - In
21	FR-GARD	'FrontGarden'	Independent - In
22	ISCOPLX	'IsIntheComplexStyleorNot'	Independent - In
23	ROADFR	'MainRoadFrontageorNot'	Independent - In
24	H-MAX	'MaxConstHeightOnaPlot'	Independent - In
25	NM-BATH	'NumberOfBaths'	Independent - In
26	NM-FACA	'NumberOfFacade'	Independent - In
27	NM-FLOR	'NumberOfFloors'	Independent - In
28	NM-PARK	'NumberOfParkingArea'	Independent - In
29	NM-ROOM	'NumberOfRooms'	Independent - In
30	OCC-PER	'OccupancyPermit'	Independent - In
31	ONWH-FL	'OnWhichFloor'	Independent - In
32	CL-PARK	'ParkingOrNot'	Independent - In
33	NC	‘Numberofcompartment’	Independent - In
34	R-SQM	'ResidenceGrossArea'	Independent - In
35	R-OPSQM	'ResidenceGrossOpenSpace'	Independent - In
36	SGARDN	'SideGardenWidth'	Independent - In

37	STWIDTH	'StreetWidth'	Independent - In
38	PLOC	‘ParcelLocation’	Independent - In
39	CAREA	‘Constructionright’	Independent - In
40	PRICE	‘Price’	Dependent – In

Table A3. Variables used in the modeling phase of the study.

Number	*ID*	*Variable Descriptive*	*Variable Type - In/Out*
1	R-SQM	'ResidenceGrossArea'	Independent - In
2	R-OPSQM	'ResidenceGrossOpenSpace'	Independent - In
3	NM-BATH	'NumberOfBaths'	Independent - In
4	NM-ROOM	'NumberOfRooms'	Independent - In
5	ONWH-FL	'OnWhichFloor'	Independent - In
6	ELVTOR	'Elevator'	Independent - In
7	ISCOPLX	'IsIntheComplexStyleorNot'	Independent - In
8	NC	‘Numberofcompartment’	Independent - In
9	NM-FLOR	'NumberOfFloors'	Independent - In
10	CAREA	‘Constructionright’	Independent - In
11	DIS-CUL	'DistanceToCulturalArea'	Independent - In
12	DIS-HOS	'DistanceToHospital'	Independent - In
13	DIS-ROD	'DistanceToMetro'	Independent - In
14	DIS-MAL	'DistanceToMall'	Independent - In
15	DIS-MRK	'DistanceToMarket'	Independent - In
16	DIS-SCHL	'DistanceToTrain'	Independent - In
17	DIS-REC	'DistanceToRecreationalArea'	Independent - In
18	DIS-UNI	'DistanceToUniversity'	Independent - In
19	PLOC	‘ParcelLocation’	Independent - In
20	PRICE	‘Price’	Dependent – In

Abbreviations : *N = Number, IU =Individual Unit; Dis=Distance*

The first 5 variables are the variable of the residency, the second 5 variables are the properties of the building, the variables between 11 and 19 are spatial variables, and all variables are listed hierarchically

BIOGRAPHICAL NOTES

Seckin Yilmazer is a geomatics engineer with a M.Sc. degree from Ankara University Department of Real Estate Appraisal and Development. He is currently a Ph.D. candidate at

Hacettepe University, Department of Geomatics Engineering and working as Internal Auditor at the General Directorate of Land Registry and Cadastre, Turkey.
Sultan Kocaman, Dr. Sc. ETH Zurich, holds a Ph.D. degree on photogrammetry and remote sensing, and has practical experience in the fields of cadastral surveying, GIS and spatial DBMS. She is currently an Assoc. Prof. at Hacettepe University, Department of Geomatics Engineering, Ankara, Turkey. Her research interests include the application of geospatial technologies and machine learning algorithms to various problem domains related to Earth Observation, natural hazards and land information management.

CONTACTS

Seckin YILMAZER
General Directorate of Land Registry and Cadastre 06100, Cankaya
Ankara
TURKEY
Tel. +90 (312) 413 60 00
seckinyilmazer@hacettepe.edu.tr
https://www.tkgm.gov.tr/icdenetim-bb/ic-denetci-2

Dr. Sultan KOCAMAN
Hacettepe University, Department of Geomatics Engineering, 06800, Beytepe
Ankara
TURKEY
Tel. +90 (312) 297 69 90
sultankocaman@hacettepe.edu.tr
https://avesis.hacettepe.edu.tr/sultankocaman/