buildings

MDPI

Article

Investigation of Real Estate Tax Leakage Loss Rates with ANNs

Mehmet Yılmaz [1] and Bülent Bostancı [2,*]

[1] Department of Architecture and Urban Planning, Tomarza Mustafa Akıncıoğlu Vocational School, Kayseri University, Tomarza 38900, Kayseri, Türkiye; myilmaz@kayseri.edu.tr
[2] Department of Geomatics, Faculty of Engineering, Erciyes University, Talas 38039, Kayseri, Türkiye
* Correspondence: bbostanci@erciyes.edu.tr

Abstract: In Türkiye, many changes have been made in the law within the past fifty years to determine the real estate tax value close to the real market value. However, the changes did not establish a fair valuation system for determining real estate tax. Despite the regulations and records of immovable properties with a geographic information system (GIS)-based inventory in recent years, the problem of leakage loss in real estate tax was still not resolved. Within the scope of this study, a mass appraisal model was created with a dataset of 499 independent sections including trading values from the last year in the district of Kayseri to determine the real estate tax leakage loss rates. Multiple regression analysis (MRA) and artificial neural network (ANN) methods, widely used in mass appraisal, were used in the analysis. Considering the analysis of the test data and the model performances, the ANN model was found to give better results than the MRA model. To conclude this study, the housing values obtained with the mass appraisal methods and the real estate tax values obtained with the existing system were compared, and a 3.7-fold difference was found between them.

Keywords: real estate tax; real estate tax value; mass appraisal; artificial neural networks; multiple regression analysis

check for updates

Citation: Yılmaz, M.; Bostancı, B. Investigation of Real Estate Tax Leakage Loss Rates with ANNs. *Buildings* **2023**, *13*, 2464. https://doi.org/10.3390/buildings13102464

Academic Editor: Pierfrancesco De Paola

Received: 5 June 2023
Revised: 14 July 2023
Accepted: 20 September 2023
Published: 28 September 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The collection of the real estate tax in Türkiye has been the responsibility of municipalities since 1986, and it is among the indispensable income sources of the municipalities. Today, municipalities face various problems in determining and collecting real estate tax. The most significant of these problems is the inability to correctly determine real estate tax values. While determining the tax value according to the laws, the square meter of the land and the square meter construction cost of buildings are considered. Land square meter unit values are determined every four years by the valuation commissions established by the municipality. Building construction costs are jointly announced annually by the Ministry of Environment, Urbanization, and Climate Change (MoEU) and the Ministry of Treasury and Finance (HMB).

The construction costs announced by the MoEU and HMB are used as a fixed value annually throughout Türkiye. Construction costs are not the same in every province. In line with the relevant laws, the tax value is calculated according to the cost method using the land and square meter values of the land and the building construction costs. Determining the market value according to the cost method is exceedingly difficult. In addition, the land and square meter values of the land are renewed annually according to the revaluation rate announced by HMB. Revaluation rates also do not reflect real market conditions.

Complaints about real estate tax mainly stem from inadequate valuation and unfair practices. In addition, disproportionate increases during each valuation period, such as 500% in some regions, damage the public trust of taxpayers. The Real Estate Tax Law General Communiqué serial no. 72, which went into force in 2017, limited this disproportionate increase. According to this communiqué, valuation commissions can increase the square land meter unit values by at most 50% in each valuation period compared with the

previous year. However, this practice will increase the difference between the market and real estate values in Türkiye, where inflation has rapidly increased.

The real estate tax value is determined according to the provisions of Tax Procedure Law No. 213 and Real Estate Tax Law No. 1319 (EVK). In addition, there is a Bylaw on the Appraisal of Tax Values to be Subject to Real Estate Tax (EVKT). The tax value calculation methods are far from scientific and do not include details about how to perform the valuation.

Due to these valuation problems, municipalities cannot collect much real estate tax. Additionally, in countries like Türkiye, where interest and inflation change frequently, it is impossible to find real estate's market value using only the cost method. Real estate valuation experts also do not prefer this method, except in exceptional cases. Mass appraisal methods should be used instead. For mass appraisal, the parameters that affect the value of real estate should first be determined. Information about real estate parameters should be recorded in an immovable database that minimizes human intervention. Real estate tax values can be calculated for each real estate using mass appraisal methods using the information in the database. Thus, personnel, appraisal costs, and time savings can be achieved. It should be remembered that no method will entirely give the market value of the real estate.

The purpose of this study is to determine the problems in appraisal in real estate tax in Türkiye; to determine real estate values with reliable methods; to ensure justice in tax, to achieve savings in terms of personnel, cost, and time in the assessment, accrual, and collection processes; and to increase the real estate tax revenues of the municipalities. For this purpose, this study includes a literature review; a description of the structure of mass appraisal methods, such as MRA and ANN; and an estimate of real estate tax values with mass appraisal applications followed by a discussion of the findings, conclusions, and recommendations.

2. Literature Review

In the literature, the following methods are used for mass appraisal models: multiple regression, hedonic regression, artificial neural networks, fuzzy logic (FL), geographic information system (GIS), analytical hierarchy process (AHP), random forest (RF), classification and regression trees (CART), machine learning, deep web, geographically weighted regression (GWR), ordinary least squares (OLS), and support vector machines (SVMs). Apart from these methods, advanced methods such as XGBoost [1], LightGBM [2,3], and deep learning [4,5] have emerged as a research topic for aggregate valuation in recent years.

The regression model is the most widely used method among practitioners and academics for modeling real estate prices [6,7]. Although it is a widely used method, it fails to effectively capture the non-linear relationship between real estate values and real estate characteristics [8–10]. To overcome the deficiencies of the regression approach, the ANN method, which gives more accurate and reliable estimations in real estate appraisal, has been used [11,12]. The method is highly accurate when there are sufficient data: it can effectively represent the non-linear relationship between real estate values and real estate characteristics [13], it can better predict outliers in the dataset [12], and it is impartial [14]. An ANN model is preferred to eliminate the deficiencies of the regression model; however, it has also been criticized in the literature for reasons such as lacking transparency [15–17] and requiring more extended training [18].

It should be noted that no appraisal model fully covers all real estate appraisal problems, as all appraisal models have pros and cons [19]. It can be used to estimate rough values in mass appraisal methods and to increase confidence in the valuation result in cooperation with property appraisers [20]. It is stated in the literature that the ANN method has immense potential to give accurate appraisal estimates [14]. In addition, some studies have reported that the ANN method is superior to the regression method [21–24]. Differences in the results of studies regarding ANN real estate appraisals may be due to the differences in data quality and model structures used in different real estate markets [25]. Data quality is

essential for developing accurate real estate appraisal models [26]. Data that may affect the value of real estate can be collected from different big data sources such as the Internet, remote sensing, and the Internet of Things (IoT) [27].

Mass appraisal primarily aims to create real estate tax, expropriation, and court appraisals. Methods that give results that cannot be explained or controlled by the courts are likely to be rejected by managers regardless of their statistical estimation ability. Values estimated using ANNs are not transparent enough to provide a clear appraisal model that is defensible against objections [16,28]. The practitioner cannot see the mathematical equation of the ANN model [12,25,29]. This is related to the non-linear nature of the ANN method [29]. However, an ANN is a valuable and powerful method under the right circumstances, especially in the context of mass appraisals for real estate tax purposes. The ANN method has been successfully used in various international real estate markets, giving faster and more accurate estimations [29,30]. An ANN can make house price appraisals after learning the fundamental relationships between input variables and the corresponding outputs. Borst first used the ANN method in real estate in 1991 [31]; the study examined the estimation accuracy of the ANN method in real estate appraisal and revealed that the method could give reliable and accurate valuation estimates [31]. After that study, the ANN method has been widely accepted in real estate appraisal.

ANNs have different parameters such as the optimum input variable, training, and test data rate, neural network model, number of hidden layers, number of neurons in the hidden layer, selection of activation and transfer function, selection of training algorithm, learning rate, and momentum term. For ANNs to run smoothly, these parameters must be determined correctly. Real estate valuation studies with ANNs were examined to determine the trend in the research area. In the literature, researchers have tried different approaches by changing ANN parameters according to their field of study. A summary of these studies, developed by Abidoye and Chan, is given in Table 1 [29].

It is known that many independent variables affect the values of real estate, and each of them has a different effect on the value [23,32]. In mass appraisal, 5 to 20 input variables are commonly used for the ANN architecture. The reviewed studies established ANN models with between 3 and 82 variables. The number of input variables must be sufficient for an ANN model to give an output with high accuracy. Too few input variables may not be sufficient for the ANN method. Too many input variables, on the other hand, may negatively affect model performance as they may be unnecessary. Because the most effective ANN topology is the one with the least number of neurons, the efficiency and accuracy of the model can be increased by optimizing the number of input variables [33,34].

In the literature, variable numbers are sorted using different methods, such as sensitivity and principal component analysis [33–37]. As a result, fewer variables are used than the number of variables determined at the beginning. Some studies state that decreasing the number of variables positively affects real estate valuation [33]. If the mass appraiser has a solid knowledge of the real estate market, they can intuitively identify the variables that affect the real estate value more [26]. The sample data used in the reviewed studies ranged from 88 to 65,302. The study conducted with 88 sample data showed that the ANN method does not require a lot of sample data to work correctly. In addition, a large-scale experimental study was conducted with the most extensive dataset in 2015. The study revealed that the ANN method can also be used with a large dataset.

Several training algorithms are used in ANNs, including derivative-based and meta-heuristic algorithms. The backpropagation (BP) algorithm is the most frequently used training algorithm in the mass appraisal of real estate. Most of the studies in the literature used the BP training algorithm [30,38,39]. BP uses the gradient descent search method to change the link weights to minimize errors between the actual and desired output. The algorithm is simple to implement and suitable for solving different problems. However, the BP algorithm has disadvantages, such as becoming stuck on local minima, more extended convergence, and user-dependent parameter settings. The literature states that the hybrid model, created by combining the convergence rates of derivative-based algorithms and the

ability of meta-heuristic algorithms to find the global optimum, can have advanced capabilities [17,34,37]. In addition, in models created with ANNs, the resilient backpropagation algorithm (RPROP) [13], cuckoo search algorithm (CS) [40], genetic algorithm (GA) [17], Powell–Beale conjugate gradient (PBCG) algorithm, and scaled conjugate gradient (SCG) algorithm [41] were also used.

It is ascertained that a single hidden layer is sufficient to achieve appropriate accuracy when using ANNs in any complex non-linear function [9]. The number of neurons in the hidden layers is at the user's discretion using trial and error. In a study conducted to find the optimal number of hidden neurons, it was recommended to start with a smaller number of neurons and increase the number until the desired result is achieved [42]. Studies in the literature show that the number of neurons in the hidden layer ranges from 3 to 20. The number of neurons was determined automatically by the software used in some studies [8,33,43]. Therefore, there is no consensus in the literature on the number of hidden neurons that should be included in an ANN model [13].

Each dataset used in modeling is divided into two parts for training and testing the developed models. In the literature, datasets are commonly split with a ratio of 80:20 to train and test models [44–46]. It is also recommended to use the cross-correction method, one of the data-splitting methods, to prevent over-learning while using ANNs [47].

To determine reliability, each model should be tested with IAAO ratio studies and other statistical accuracy tests. It is recommended to use software that offers robust statistical accuracy tests for these tests [17]. Software such as NeuroShell, SPSS, NeuralWare, Spreadsheet, NeuroSolutions, Alyuda, MATLAB, Trajan, DTREG, BKP—Neural Network Simulator, Encog3, SPSS, R programming, and WEKA have been used in the literature. The values estimated using an ANN may vary depending on the software used [48]. For this reason, MATLAB R2015a software was used in this study for the ANN model, which has been widely used in the literature, and SPSS 26 software was used for statistical tests. In the mass appraisal results obtained with the ANN, the mean absolute percentage error (MAPE) was found to range from 5 to 15 percent in general.

Mass appraisal is widely used in many countries to calculate real estate tax [26,49–51]. There are two critical obstacles to calculating real estate taxes using mass appraisal. First, there must be a sufficient number of real estates with known market value (dependent variable). On the other hand, the properties (independent variables) of the real estate for which the value will be estimated should be recorded [52,53]. Accurate and reliable information about dependent and independent variables is needed [52], so governments should invest in data management and analysis [54].

Table 1. Summary of Mass Appraisal Studies Conducted with ANN in the Last 10 Years.

Study	Country	Sample/Number of Variables	Training: Test Ratio	Model Structure	Training Algorithm	Software Used
Lai (2011) [24]	Taiwan	2471/9	70:30	9-TE-1	BP	Alyuda
Hamzaoui and Perez (2011) [18]	Morocco	148/13	75:25	13-5-1	LM	MATLAB
Lin and Mohan (2011) [9]	USA	33,342/6	80:20	82-6-1	BP	-
Zurada et al. (2011) [7]	USA	16,366/18	-	-	-	-
Kontrimas and Verikas (2011) [49]	Lithuania	100/13	-	13-7-1	LM	MATLAB
Amri and Tularam (2012) [11]	Australia	7849/10	-	10-6-4-1	LM	-
McCluskey et al. (2012) [16]	Northern Ireland	2694/6	80:20	6-20-1	BP	-
Tabales et al. (2013) [44]	Spain	10,124/6	80:20	6-6-1	-	Trajan
McCluskey et al. (2013) [28]	Northern Ireland	2694/6	80:20	6/TE/1	BP	DTREG

Table 1. *Cont.*

Study	Country	Sample/Number of Variables	Training: Test Ratio	Model Structure	Training Algorithm	Software Used
Morano and Tajani (2013) [23]	Italy	85/6	80:20	6-13-1	-	BKP—Neural Network Simulator
Mimis et al. (2013) [55]	Greece	3150/9	-	9-5-1	-	-
Ahmed et al. (2014) [56]	Bangladesh	100/40	70:30	40-10-1	-	MATLAB
Vo (2014) [33]	Australia	7319/15	80:20	15-8-1	iRPROP +	Encog 3
Morano et al. (2015) [45]	Italy	90/7	80:20	7-13-1	-	BKP—Neural Network Simulator
Sampathkumar et al. (2015) [22]	India	204/13	80:20	13-3-1	LM	-
Feng and Jones (2015) [35]	England	65,302	-	-	-	SPSS 21
Güneş and Yıldız (2015) [50]	Türkiye	2447/10	80:20	10-10-1	-	-
Vo et al. (2015) [34]	Australia	-	-	14-7-1	iRPROP +	Encog 3
Yacim et al. (2016) [40]	South Africa	3494	-	-	BP, CSLM, CSBP	MATLAB
Abidoye and Chan (2017) [38]	Nigeria	321/11	80:20	11-5-1	BP	R programming software
Yalpır (2018) [32]	Türkiye	98/6	80/20	3-model	-	MATLAB
Morillo Balsera et al. (2018) [57]	Spain	9032/15	70:30	15-7-1	-	-
Yacim and Boshoff (2018a) [41]	South Africa	3242/18	70:30	18-20-1	PSOBP	WEKA
Yacim and Boshoff (2018b) [17]	South Africa	3232	70:30	-	LM, PBCG, SCG, BP	MATLAB and WEKA
Abidoye and Chan (2018) [8]	Hong Kong	321/11	80:20	11/5/1	BP	R programming software
Alexandridis et al. (2019) [58]	Greece	36,527/22	85:15	-	LM	-
Rahman et al. (2019) [39]	Malaysia	215/2	90:10	-	BP	-
Kang et al. (2020) [59]	South Korea	9435/33	70:30	-	GA	-
Yacim and Boshoff (2020) [60]	South Africa	3225/11	70:30	11/TE/1	-	-

Note: TE stands for trial and error; "-" indicates unavailable.

3. Methods

Many appraisals must be performed quickly to determine the real estate tax value. Attempts have been made to develop mass appraisal methods to make many appraisals based on mathematical models.

Mass appraisal is defined by the International Association of Assessing Officers (IAAO) in the USA as "the process of appraising a large number of real estates within a specified period using fully accurate up-to-date data, standardized methods, models, and statistical tests" [61]. According to this definition, a mass appraisal is the simultaneous appraisal of multiple properties using reliable data and methods.

To date, institutions and researchers related to mass appraisal have conducted many studies on the theoretical structure and standards, as summarized in Table 2. These studies present criteria and methods for mass appraisal, and their contents are explained in detail [62].

Table 2. Institutions that set and publish standards for mass appraisal [62].

Standard	Institution	Year (First Version)	Year (Latest Version)
SMARP	IAAO	1976	2017
RICS Red Book	RICS	1983	2017
IVS	IVSC	1990	2017
USPAP	AF	1987	2018

Mass appraisal involves estimating the relationship between market value and independent variables that are believed to affect market value. Model reliability depends on the quality and relevance of the data used. For this reason, it is necessary to check whether the ratio between the estimated and market values is within certain limits and whether the data used are accurate [26]. In other words, a mass appraisal can be performed when sufficient and reliable data about market values and properties of real estate are available. However, there are no reliable data sources in developing countries, and there are difficulties in obtaining accurate data. In cases where t sales prices do not reflect actual values, a different data source should be used or the obtained data should be verified with the help of other sources.

Mass appraisal practices in some countries have also been investigated. Lithuania has established an advanced central registration system for mass appraisal in this framework. The mass appraisal system determines the value of buildings since 2005 and lands since 2013. However, there are some problems with the adequacy of data for appraisal and the accuracy of data collected in the past. In Lithuania, a reappraisal is conducted every five years [63]. On the other hand, the Netherlands has been performing reappraisals annually using the mass appraisal system since 2007. Values produced annually using mass appraisal are also used for purposes other than taxation. The Netherlands has improved data quality by allowing taxpayers to check their real estate online [64]. A mass appraisal may not determine all real estate values in developing countries. For example, a mass appraisal can be performed in Moldova with only 12.5 percent of immovable properties [65].

Several studies were examined in detail during the literature review stage of this study, and it was observed that many methods were used for mass appraisal models. However, MRA and ANN models were prominent [62].

3.1. Multiple Regression Analysis (MRA)

Multiple regression analysis (MRA) is used when the value of an independent variable needs to be estimated based on two or more dependent variables. MRA is a statistical modeling method that considers all variables by performing many tests and analyses. Variables statistically significant for the model are considered, and other insignificant variables in explaining the dependent variable are ignored. MRA was first used by Rosen [66] to estimate the market value of real estate [33]. With this method, the real estate value can be estimated according to the variables that affect the real estate value [50,67]. The multiple linear regression model is generally defined as in Equation (1):

$$\begin{aligned} \textit{Sales Price} = \beta_0 \; & + \beta_1 \textit{Floor} + \beta_2 \textit{Total Number of Floors} + \beta_3 \textit{Front} + \beta_4 \textit{Number of Rooms} \\ & + \beta_5 \textit{Number of Bathrooms} + \beta_6 \textit{Number of Balconies} + \beta_7 \textit{Heating System} \\ & + \beta_8 \textit{Planned Areas} + \beta_9 \textit{Recreation Areas} + \beta_{10} \textit{Cellar} + \beta_{11} \textit{Security} \\ & + \beta_{12} \textit{Outdoor Parking} + \beta_{13} \textit{Indoor Parking} + \beta_{14} \textit{Indoor Swimming Pool} + \beta_{15}\text{Elevator} \\ & + \beta_{16}\text{Dressing Room} + \beta_{17}\text{Location} + \beta_{18}\text{Age of the Building} + \beta_{19}\text{Laundry Room} + \varepsilon \end{aligned} \quad (1)$$

Sale Price: Dependent variable.
β_i: Parameters to be estimated.
ε: Error term.

The estimated coefficients are determined using MRA. These coefficients indicate the effect of each independent variable on the real estate value. For example, five indepen-

dent variables that affect the real estate value the most can be explained using the MRA. Moreover, the statistical significance of each independent variable can be tested.

Although MRA is the most used method in mass appraisal practices, it has disadvantages. One of the disadvantages of MRA is that to obtain valid results, the dependent variable must be continuous; there must be more than two independent variables; the independence of the observations must be ensured; there must be a linear relationship between the dependent variable and each of the independent variables; the error variance must be constant; the data must not show multicollinearity; there must be no unusual points such as extremes or outliers; and the errors must be normally distributed [50,68]. However, it is well known that housing prices tend to be spatially dependent, as neighboring houses' physical and environmental characteristics are similar. Therefore, spatial autocorrelation should be considered to increase the accuracy of mass appraisal methods [69]. Some researchers include longitude and latitude values in the dataset or use location-sensitive (kriging, spatial econometric model (SEM), spatially varying coefficient (SVC), etc.) models to calculate spatial effects [70].

3.2. Artificial Neural Networks (ANNs)

ANNs are computer systems developed to automatically carry out various functions such as learning, memorizing, and relating information, which are the functions of the human brain. There are many types of ANNs. The backpropagation ANN is the most widely used type. This type gives excellent results in neural network prediction operations. ANNs are mathematical systems with good processors (neurons) connected by their weights. A processor is essentially an equation called a transfer function. The processor receives signals from neurons, combines and converts them, and returns digital outputs. Processors roughly correspond to neurons and are interconnected in a network. This structure constitutes the architecture of ANNs [71].

An ANN is an estimator that can manage non-linear relationships in real estate data. There is no generally accepted approach to designing ANN architectures to solve prediction problems. The essential elements of ANN are inputs, weights, addition (joining) function, activation function, and cell output. A simple ANN architecture is provided in Figure 1.

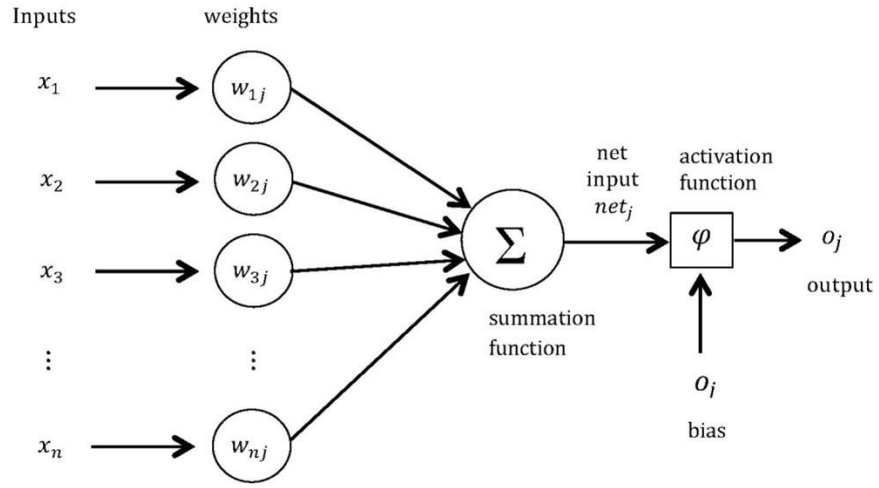


Figure 1. Simple neural network architecture [72].

An ANN consists of three layers: an input layer, an output layer, and at least one hidden layer for processing nonlinear elements. The values of the input variables representing the characteristics of the immovable are sent to the hidden layer via the input layer. The number of input data determines the network's input variables. The hidden layer is the layer between the input and output layers where the mathematical processing occurs. The hidden layer contains two operations: the addition (joining) and the activation function. Each input layer value is multiplied by a specified weight in this layer. These weight values

can be determined randomly or by the user initially. The sum of the products is calculated with a threshold value. More complex transfer functions can also be used instead of this simple addition. The activation function creates a result from the last determined value. The result of the transformation in the hidden layer is produced in the output layer, where the estimated real estate value is obtained. There are many activation/transfer functions used for transformation [17]. Kaastra and Boyd suggested the eight steps in Table 3 to develop an ANN model in general [38].

Table 3. Eight Steps in Designing a Neural Network Prediction Model.

Step	Title 2
Step 1:	Choosing the variables to be included in the model
Step 2:	Data collection from the study area
Step 3:	Analysis of the collected data
Step 4:	Splitting the dataset for training and testing the model
Step 5:	Determining the topological structure of the neural network
Step 6:	Model evaluation criteria and accuracy measurement
Step 7:	Training the ANN model
Step 8:	Application

It is challenging to create an ANN model for the problems encountered in daily life because the performance of an ANN model depends on many parameter settings. Also, there is no specific method for choosing the best settings. To develop an ANN model, it is necessary first to determine the number of input neurons, hidden layers (hidden neurons), and output neurons. The number of neurons in the input layer depends on the number of independent variables that will be used to develop the model. In the literature, input variables generally range from 5 to 20. In addition, Vo proposed an approach for removing unnecessary input variables and determining the optimum input variable. The number of output neurons is only one and expresses the value of the real estate [33].

Secondly, in determining ANN parameters, the network topology should be created by determining the number of hidden layers and neurons and, if necessary, a bias neuron in a hidden layer. Then, other decisions regarding the neural network design are made, including an activation function, training type, data normalization methods, and performance criteria for hidden neurons.

3.3. Study Area and Data Collection

The study area is the Melikgazi district in Kayseri Province, with the highest population in the province. According to 2020 TUIK data, the population of the district is 582,055. There are 283,945 independent sections in Melikgazi. Of these independent sections, 246,098 are residential. On average, there is one residence for 2.36 people in the district. Melikgazi was chosen over other districts because of its reliable, up-to-date, and numerical archive records.

The sample size should be sufficient so that the values estimated using the mass appraisal model accurately represent the study area. For this reason, neighborhoods with many independent sections were selected as the study area. The Köşk neighborhood, which was selected as a study area, is in the northeast of Kayseri Province. The neighborhood generally has a flat topography, consisting of 5–10 years old buildings with 8–15 floors, and is rapidly developing with new buildings being constructed. The social facilities in the region are sufficient. The Köşk neighborhood is a residential-dominated, planned, developing area that appeals to the middle- and high-income population with its medium and new buildings. It is rapidly developing and does not have infrastructure problems. It is about 2–4 km from the city center. The Alpaslan neighborhood is in the northeast of Kayseri Province. The neighborhood generally consists of 5–30-year-old buildings with 10–15 floors. Alpaslan is the most elite neighborhood in Kayseri Province, with large-area luxury residences and restaurants/cafes. Social facilities in the region are at an excellent

level. Alpaslan is a residential-dominated, planned area that appeals to the high-income population, has completed its development, and has no infrastructure problems. It is about 2–4 km from the city center. According to the 2020 data from the Turkish Statistical Institute, the population of Alpaslan is 24,220 people, and the population of Köşk District is 23,528. There are 9205 independent sections in the Alpaslan neighborhood and 8640 independent sections in the Köşk neighborhood. Figure 2 shows the map of the study area.

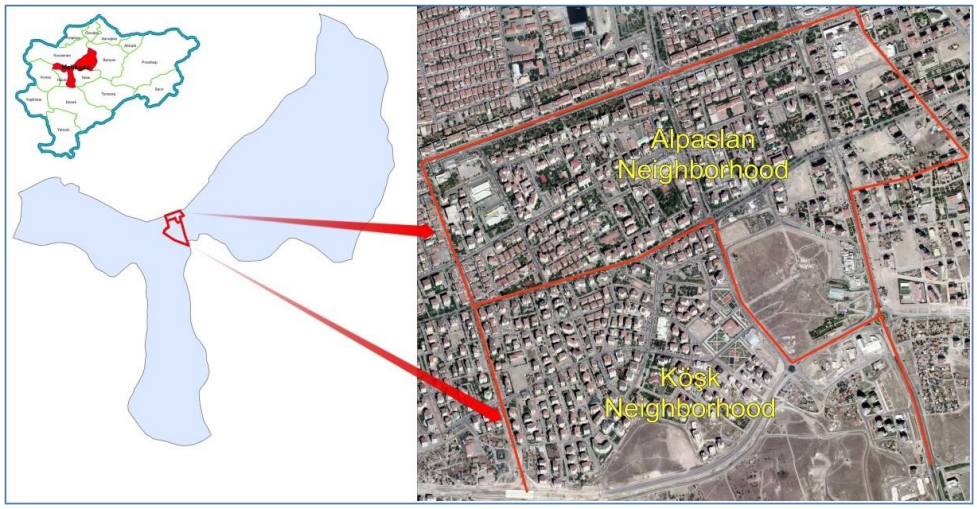


Figure 2. Map of the Study Area.

The dataset collected from the Alpaslan and Köşk neighborhoods for the mass appraisal model includes the housing sales values and properties from 1 January 2019 to 31 June 2021. A GIS-based database was created with the valuation data of 228 houses in the Alpaslan neighborhood and 271 houses in the Köşk neighborhood on the field research and from the reports of real estate appraisers. There are 422 parcels in the two neighborhoods, and data on at least one sale were obtained in 215. Using Kelley's program developed in the R language [73], a sufficient sample size was calculated as 217 with an effect size of 0.35, a statistical power of 0.80, a confidence interval of 0.05, and 18 independent variables. In addition to this sample size, several samples were needed to evaluate the models. Since a sufficient sample size could not be provided with one neighborhood, it was decided to combine the two neighborhoods. The data distribution by study area is presented in Figure 3.

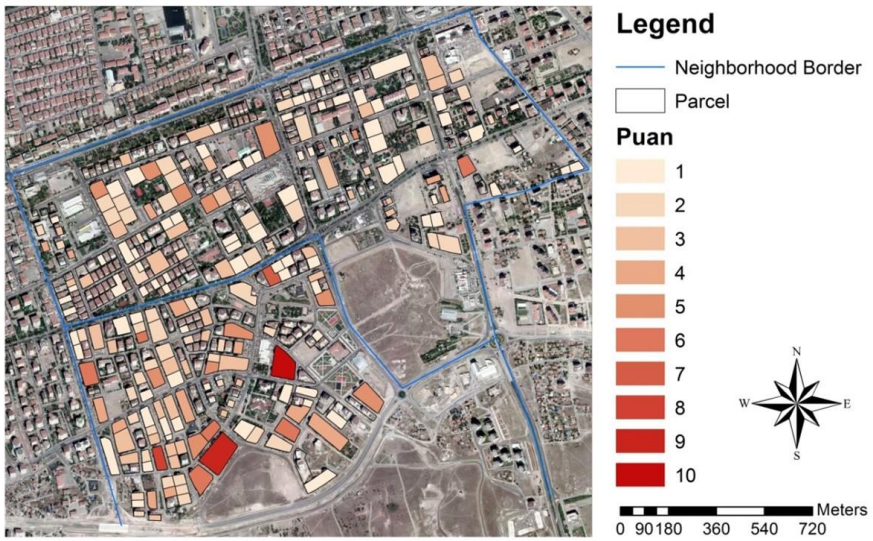


Figure 3. Data Distribution in the Study Area.

One of the significant issues in mass appraisal models is determining the variables that affect the value of real estate. The number and definitions of variables vary by country, city, and region. In the study, a literature review and opinions of experts who work in the field of real estate appraisal in Kayseri were used to determine the variables in the dataset. Initially, 20 variables were determined for each residence in the dataset. However, since there was no independent section with an outdoor swimming pool in the study area, this variable was excluded from the dataset. The names and descriptions of the variables are given in Table 4.

Table 4. Names and descriptions of the variables.

Order No	Variable	Description	Order No	Variable	Description
1	Floor	The floor where the independent section is located according to the architectural project.	11	Security	Whether there are security personnel in the parcel (yes or no).
2	Total number of floors	The total number of normal floors, excluding the duplex floor above the ground floor.	12	Outdoor parking	Whether there is an open car parking lot in the garden of the building (yes or no).
3	Front	The front of the independent section faces according to the architectural project.	13	Indoor parking	Whether there is a closed parking lot in the basement of the building (yes or no).
4	Number of rooms	The number of closed areas in the independent section surrounded by walls (except for the kitchen and en-trance halls).	14	Indoor swimming pool	Whether there is an indoor swimming pool in the parcel where the independent section is located.
5	Number of bathrooms	The total number of bathroom and shower areas in the independent section.	15	Elevator	The number of elevators that can be used to reach the independent section (not the total number of elevators in the building).
6	Number of balconies	The total number of balconies in the architectural project of the independent section (French balconies not included).	16	Dressing room	The number of elevators that can be used to reach the independent section (not the total number of elevators in the building).
7	Heating system	The type of heating used in the building.	17	Location	The total distance to the tram stop, hospital, city center, shopping mall, place of worship, main street, education area, and university area.
8	Planned areas	The area determined according to the gross area definition of the Zoning Regulation.	18	Age of the building	The number of years since the building was completed.
9	Recreation areas	The number of areas such as gyms, children's playgrounds, sauna, etc.	19	Laundry room	The number of laundry rooms in the architectural project of the independent section.
10	Cellar	The number of cellars in the architectural project of the independent section.			

There are verbal and numerical data in the dataset. Verbal data were converted into numerical data for the mass appraisal model. Eight types of verbal data affect the real estate value in the study area. These are the front, heating system, cellar, security, outdoor parking, indoor parking, outdoor swimming pool, and indoor swimming pool variables. Dummy variables in the form of 0–1 were used for the variables including cellar, security, outdoor parking, indoor parking, and indoor swimming pool. A 5-point Likert scale was used for the front and heating system variables.

3.4. Application

Normality, outlier, and multicollinearity analyses required for MRA were performed. In a regression analysis, outliers are usually defined as values of standardized error (i.e., outlier for y) greater than ±3. However, some researchers in the literature [74,75] suggest scanning for x and y outliers separately. Outliers in the x- and y-direction were examined using the statistical values listed in Table 5. Accordingly, a Mahalanobis distance greater than a 42.312 chi-square value and a larger leverage value greater than 0.08016 were accepted as outliers in the x-direction. In addition, studentized deleted residue values greater than the Bonferroni critical value were accepted as outliers in the y-direction. The critical value for Bonferroni's method was:

$$t\left(1-\frac{\alpha}{2n};n-p-1\right)=t(0.999799599;478)=3.56 \quad (2)$$

Table 5. Outlier Analysis Strategy [74,75].

Outliers	Statistical Value	Critical Value
x-outliers	Mahalanobis distance	The chi-square value at the α = 0.001 significance level and (k − 1) degrees of freedom
	Leverage value	$\frac{2p}{n}$
y-outliers	Studentized deleted residuals	α = 0.10 significance value and t-distribution value at n-k-2 degrees of freedom
	Cook's distance	The corresponding significance value in the F(p,n-p) distribution of Cook's distance > 0.50

k is the number of independent variables. p is the number of independent variables, including the regression constant coefficient (p = k + 1). n is the number of observations in the dataset.

The significance values of the F(20.479) distribution are minimal due to the sample size. Therefore, samples with three times the average Cook's distance value instead of significance values are considered outliers in the y-direction.

As shown in Figure 4, 39 data points, using the Mahalanobis distance statistical value, and 41 data points, using the leverage statistical value, were determined as outliers in the x-direction. In addition, 5 data points, according to studentized deleted residuals values, and 28 data points, according to Cook's distance values, were determined as outliers in the x-direction.

After determining the outliers in the x- and y-direction, the statistical values of DFFITS and DFBETAS were examined to determine whether these outliers were effective in the model [74]. DFFITS is the change in estimated value when a particular observation is removed from the model. DFBETAS, on the other hand, is the change in the regression coefficient resulting from removing a certain variable from the model. It should be noted that DFFITS and DFBETAS values with absolute values greater than one have an effect on the model. Outliers according to DFFITS and DFBETAS values are given in Table 6 and Figure 5.

Table 6. Effect of Outliers on the Model [74].

Dataset	DFFITS	DFBETAS
Small and medium dataset	>1	>1
Big dataset	$>2\sqrt{\frac{p}{n}}$	$>\frac{2}{\sqrt{n}}$

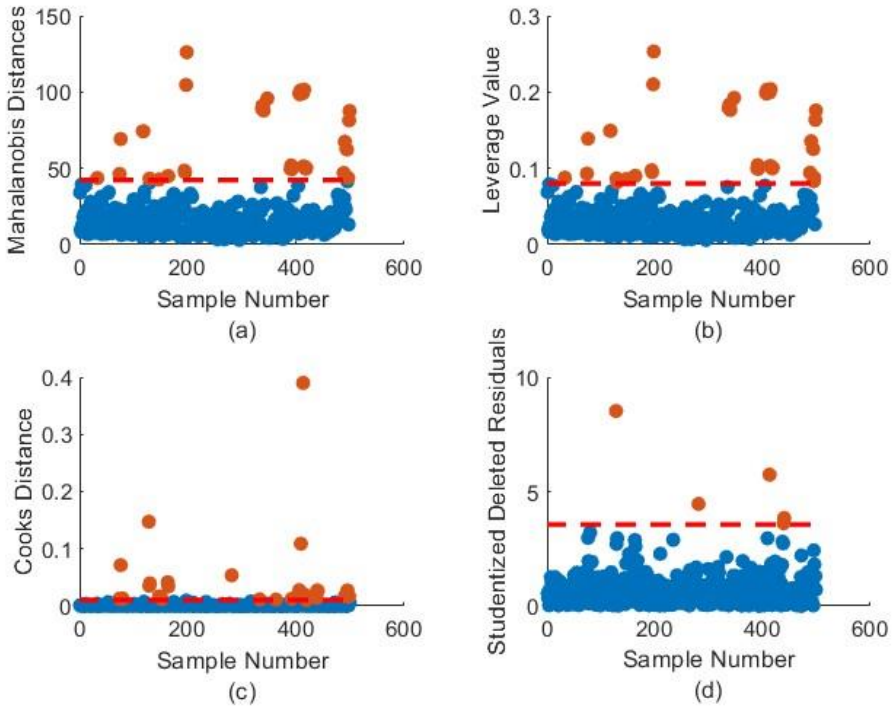


Figure 4. (**a**) Outliers in the x-direction according to Mahalanobis distances. (**b**) Outliers in the x-direction according to leverage values. (**c**) Outliers in the y-direction according to Cook's distances. (**d**) Outliers in the y-direction according to studentized deleted residual values. (Red dash line: critical value, red dots: outliers, blue dots: non outliers).

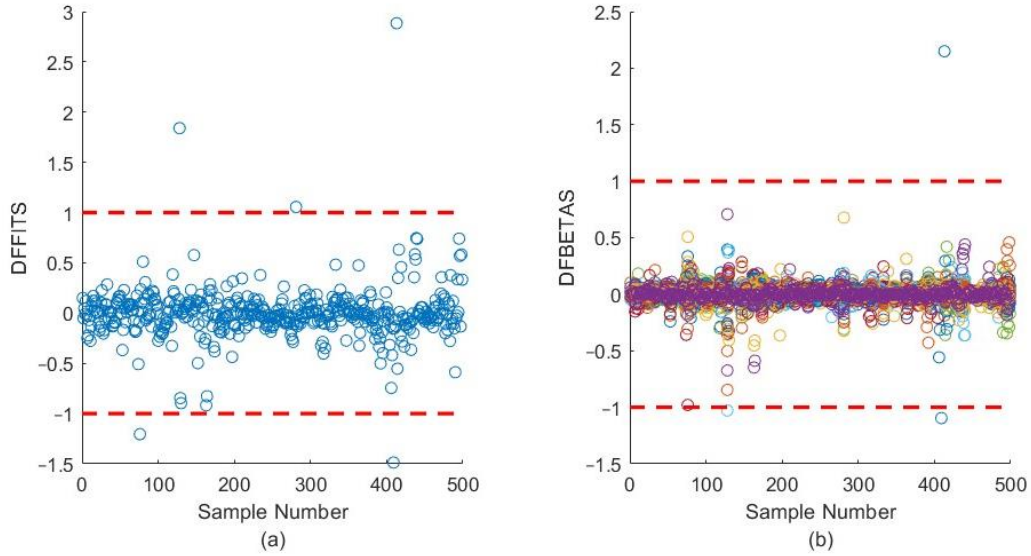


Figure 5. (**a**) Outliers with an effect on the model according to DFFITS values. (**b**) Outliers with an effect on the model according to DFBETAS values.

DFFITS is the change in the predicted value when a particular observation is removed from the model. DFBETAS, on the other hand, is the change in the regression coefficient due to removing a certain variable from the model. DFFITS and DFBETAS values with absolute values greater than one were determined as outliers having an effect on the model.

As a result of the outlier analysis, 16 samples were excluded from the dataset because they negatively affected the model results.

Normality analysis uses Shapiro–Wilk tests, Kolmogorov–Smirnov tests, skewness, and kurtosis values. According to Shapiro–Wilk and Kolmogorov–Smirnov tests, it was

determined that all variables were not normally distributed because p-significance values were less than 0.05. Kim [76] assumes that independent variables with a skewness value of less than ±7 or a kurtosis value of less than ±2 are normally distributed. According to the values listed in Table 7, the floor, front, number of bathrooms, age of the building, outdoor parking, and location variables were normally distributed. In addition, since the number of balconies and cellar variables are at limit values, it is assumed that these variables are normally distributed.

Table 7. Kurtosis and Skewness Values of the Independent Variables.

Independent Variable	Skewness	Kurtosis
Floor	0.227	1.142
Total number of floors	2.118	4.662
Front	0.555	0.289
Number of bathrooms	0.342	0.337
Number of balconies	0.563	2.053
Heating system	2.34	3.490
Elevator	1.74	2.928
Age of the building	0.599	0.208
Laundry room	8.897	77.480
Cellar	2.048	2.914
Security	5.466	27.995
Outdoor parking	1.955	1.830
Indoor parking	2.370	3.632
Indoor swimming pool	10.966	118.729
Recreation areas	4.204	17.627
Dressing room	2.975	7.796
Location	0.184	0.681
Planned areas–number of rooms	1.384	5.197

Data conversion operations can be performed on variables that do not show normal distribution. However, data conversion processes are not appropriate since these variables are dummy variables in the form of 0–1. Therefore, these variables were accepted without data transformation.

For multicollinearity analysis, the relationships between dependent and independent variables and between independent variables and other independent variables were examined. Pearson's correlation coefficient and Belsley's collinearity diagnostics table were used. Multicollinearity was assumed if the variance ratio is determined as 0.50 or more for two or more variables in rows with a condition index greater than 30 [77]. In Figure 6, in the row with a condition index of 52.7, the number of rooms and planned areas independent variables with variance ratios greater than 0.5 are observed.

In addition, when the Pearson correlation coefficients in Figure 7 were examined, it was determined that there was a strong positive correlation of 0.83 between the number of rooms and planned areas variables. These variables were combined into a single variable using principal component analysis to remove the relationship between the planned areas and the number of rooms variables.

Then, the dataset was randomly divided into 80% for training and 20% for testing. Removing the nonsignificant variables in the regression model and determining the variables that best reflect the relationship between the independent and dependent variables was necessary. Stepwise feature selection methods were used to determine the variables to include in the model. According to the data in Tables 8 and 9, these variables were determined using the stepwise method on the training dataset.

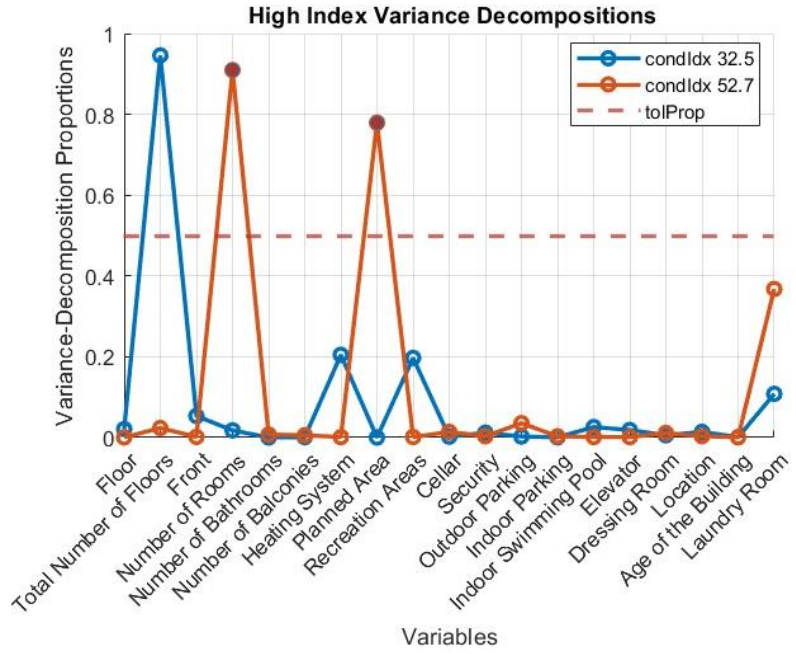


Figure 6. Belsley collinearity diagnostics.

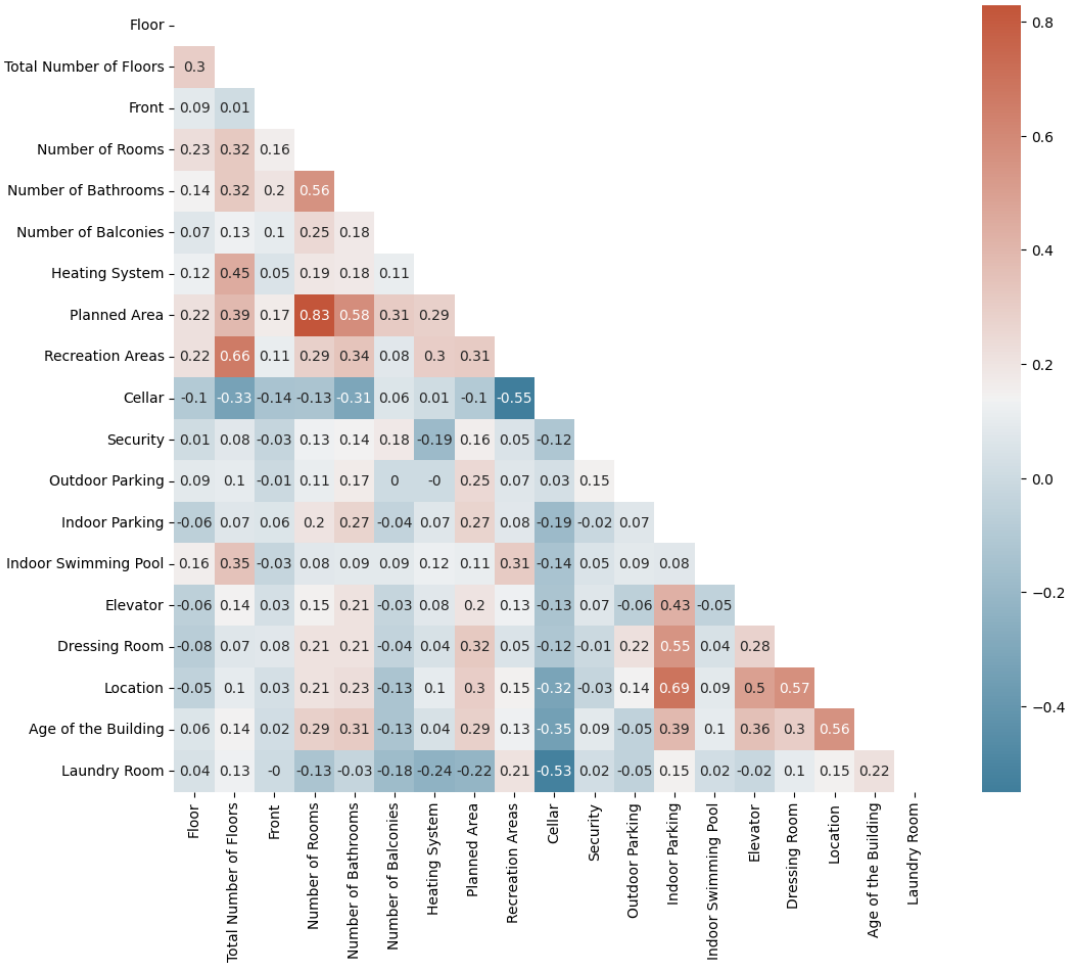


Figure 7. Correlation Coefficients between Independent Variables.

Table 8. Variable Selection with the stepwise feature selection Method.

Model	Adj. R^2	Std. Error of the Estimate	Durbin–Watson
1	0.575	137,069.701	
2	0.688	117,488.522	
3	0.713	112,766.073	
4	0.721	111,013.172	
5	0.729	109,421.613	
6	0.739	107,381.913	
7	0.749	105,463.346	
8	0.752	104,735.857	
9	0.754 *	104,219.486	1.430

* Predictors in the model: (constant), planned areas–number of rooms, age of the building, dressing room, heating system, front, indoor swimming pool, security, floor, indoor parking.

Table 9. Multiple Regression Analysis Training Dataset Model Summary.

Model	Adj. R^2	Std. Error of the Estimate	Change Statistics					Durbin–Watson
			R^2 Change	F Change	df1	df2	Sig. F Change	
1	0.737	107,941.915	0.741	159.201	7	389	0.000	1.850

Dependent variable: market value.

Although the indoor swimming pool variable does not have a negative relationship with the dependent variable, the regression coefficient is negative. The dataset has an indoor swimming pool in only four samples. Therefore, the model could not accurately predict the coefficient of this variable. When the indoor swimming pool variable was removed from the model, the safety variable ($p < 0.05$) was not significant. Therefore, these variables were excluded from the model. As a result, a seven-variable model was created, and the model results are given in Tables 9 and 10.

Table 10. Multiple Regression Analysis Training Dataset Model Coefficients.

Model	Unstandardized Coeff.		Std. Coeff. Beta	t	Sig.	Collinearity Statistics	
	B	Std. Error				Tolerance	VIF
(Constant)	613,108.887	36,005.383		17.028	0.000		
Floor	3880.125	1362.778	0.077	2.847	0.005	0.918	1.090
Front	6499.500	1849.291	0.094	3.515	0.000	0.936	1.068
Heating system	28,405.289	8519.026	0.090	3.334	0.001	0.918	1.089
Age of the building	−8170.494	840.435	−0.271	−9.722	0.000	0.858	1.166
Indoor parking	43,908.426	17,577.915	0.069	2.498	0.013	0.863	1.159
Dressing room	116,780.533	21,303.917	0.164	5.482	0.000	0.739	1.353
Planned areas–number of rooms	120,577.164	5905.766	0.597	20.417	0.000	0.779	1.284

Predictors: (constant), planned areas–number of rooms, age of the building, indoor parking, front, heating system, floor, dressing room. Dependent variable: market value.

The Durbin–Watson (DW) test was used to determine autocorrelation. When the sample order in the dataset is changed, the DW statistical value and model coefficients change. Based on the DW critical value table, the lower (dL) and upper (dU) limits of the critical value are 1.728 and 1.801, respectively, according to the number of 397 observations in the training dataset, seven independent variables, and the 0.01 significance level. There is no autocorrelation when the DW value is between dU (1801) and 4 dU (2199). As shown in Figure 8, the DW value was calculated for 100 different dataset combinations. Most of these dataset combinations provide the desired DW value.

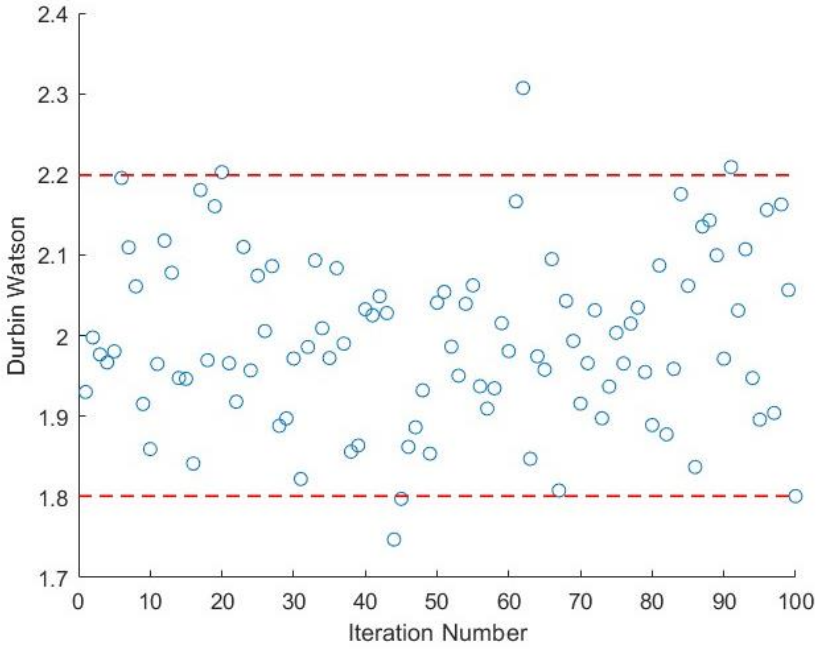


Figure 8. DW values for different dataset combinations.

A regression model was established with all independent variables using the training dataset. A model with seven independent variables and an adjusted R^2 coefficient of 0.737 was established. The equation of the model is given below:

$$\text{Sale Price} = 613{,}108.88 + \text{Planned Areas} - \text{Number of Rooms} \times 120{,}577.16 - \text{Building Age} \times 8170.49 + \text{Dressing Room} \times 116{,}780.53 + \text{Heating System} \times 28{,}405.28 + \text{Front} \times 6499.50 + \text{Floor} \times 3880.12 + \text{Indoor Parking} \times 43{,}908.42$$

Performance criteria and ratio operations were performed by comparing these values with the sales prices. The results of these studies are explained in the findings section. A multiple regression analysis was conducted to adjust the dataset to the mathematical model and control the ANN analysis.

An ANN consists of three layers: an input layer, an output layer, and at least one hidden layer for processing non-linear elements. The number of these layers should be determined at the beginning of the ANN architecture. For the input layer, seven different models were created using all the variables in the dataset (18 independent variables), the variables that were identified as being significant using the multiple regression data analysis (7 independent variables), and the variables in which the planned areas and number of rooms variables were combined with the principal component analysis. Different models were created by changing only the number of independent variables without changing the ANN parameters. While creating these models, the coefficients of the variables in the multiple regression analysis were considered. Variables with low regression coefficients were removed from the model, and the ANN results were examined. In addition, the order of importance of the independent variables was determined using SPSS sensitivity analysis. With the sensitivity analysis, while $n - 1$ independent variables remained constant, changes in the estimated real estate value were calculated when an independent variable changed. Table 11 lists the order of importance of the independent variables from the highest to the lowest.

This table shows that the study area's average independent section size (planned area) significantly impacts house prices, followed by building age and number of rooms. In addition, factors with the most negligible effect on housing prices seem to be outdoor parking, number of bathrooms, and dressing room.

Table 11. Order of Importance of the Independent Variables.

Independent Variable	Importance	Independent Variable	Importance
Planned areas	0.294	Floor	0.032
Age of the building	0.170	Indoor swimming pool	0.029
Number of rooms	0.054	Cellar	0.026
Recreation areas	0.054	Indoor parking	0.025
Elevator	0.048	Location	0.023
Security	0.039	Heating system	0.021
Front	0.038	Dressing room	0.019
Total number of floors	0.038	Number of bathrooms	0.013
Number of balconies	0.034	Outdoor parking	0.009
Laundry room	0.033		

In the models created with different independent variables, the dataset was divided into two groups, i.e., the training and test datasets, using 5-fold cross-validation. The models were trained using the Levenberg–Marquardt training algorithm, where epochs = 50, goal = 1×10^4, and min_grad = 1×10^4 were determined as stopping criteria at the end of the trials. On the other hand, the tansig function, which is widely used in the literature, was used for the activation function. The model results according to these parameters are given in Table 12.

Table 12. Model Results According to Different Independent Variables.

Model	Dataset	MAPE	RMSE	R^2	Adj. R^2
1	Training	0.0786	6.55×10^4	0.8954	0.8901
	Test	0.1419	1.25×10^5	0.6099	0.5148
2	Training	0.0942	7.84×10^4	0.847	0.8396
	Test	0.1372	1.25×10^5	0.5947	0.5023
...	Training	0.1027	8.56×10^4	0.8151	0.8103
	Test	0.131	1.27×10^5	0.606	0.5607
6 [a]	Training	0.1031	8.59×10^4	0.8196	0.8158
	Test	0.1315	1.11×10^5	0.6886	0.6606
7	Training	0.6499	8.37×10^4	0.8287	0.8256
	Test	0.1331	1.13×10^5	0.6752	0.6499

[a] Independent variables: floor, front, heating system, age of the building, indoor parking, dressing room, planned areas–number of rooms.

The best results were obtained with the independent variables in Model 6. The trial-and-error method found the number of hidden neurons for this model. Values for the number of hidden neurons were tested, starting from one to the number of input layers. The performance criteria values according to the hidden neuron numbers are given in Table 13.

Table 13. Determining the Number of Hidden Neurons for the Best Model.

Number of Hidden Neurons	Dataset	MAPE	RMSE	R^2	Adj. R^2
1	Training	0.102	85,777	0.834	0.830
	Test	0.112	90,836	0.707	0.681
2	Training	0.095	78,290	0.848	0.845
	Test	0.119	125,496	0.638	0.605
3	Training	0.103	84,426	0.832	0.828
	Test	0.124	100,444	0.717	0.692
4	Training	0.100	84,673	0.824	0.821
	Test	0.127	107,401	0.724	0.699
5	Training	0.091	75,019	0.849	0.846
	Test	0.144	158,837	0.540	0.498
...	Training	0.191	146,569	0.395	0.383
	Test	0.210	164,698	0.293	0.230
10	Training	0.146	115,142	0.606	0.598
	Test	0.170	140,252	0.478	0.431

For Model 6, the best number of hidden neurons was four. This model attempted to address the over-learning problem using various activation functions and stopping criteria between the training and test datasets. The results given in Table 14 were obtained with the Levenberg–Marquardt training algorithm.

Table 14. Levenberg–Marquardt Training Algorithm Results.

Model	Dataset	MAPE	RMSE	R^2	Adj. R^2
LM Algorithm	Training	0.1110	9.1229 × 104	0.7131	0.7686
	Test	0.1253	1.0765 × 105	0.7961	0.6873

The Levenberg–Marquardt training algorithm could not solve the over-learning problem, and the model's success could not be increased. The parameters used with the ready dataset (house dataset) in the MATLAB software were tested. The ready dataset results are given in Table 15.

Table 15. MATLAB Ready Data Results.

Model	Dataset	MAPE	RMSE	R^2	Adj. R^2
house_dataset	Training	0.1114	3.0663 × 103	0.8879	0.8842
	Test	0.1297	3.8047 × 103	0.8263	0.8006

The ready dataset with the parameters used for the model's training was successful. Accordingly, it was determined that the reason for the low performance values was the data. Instead of using the Levenberg–Marquardt training algorithm, other derivative-based training algorithms were used to increase the model's success. In addition, the sigmoid, hyperbolic tangent, and ReLU functions were used as activation functions. All possibilities for these parameters were tested using many experiments. Ultimately, the conjugate gradient backpropagation with Powell–Beale restarts (CGB) and conjugate gradient backpropagation with Polak–Ribiére updates (CGP) training algorithms gave the best results. The results of CGB and CGP training algorithms are given in Tables 16 and 17.

Table 16. CGB and CGP Training Algorithm Results.

Model	Dataset	MAPE	RMSE	R^2	Adj. R^2
CGB	Training	0.1156	9.6755 × 104	0.7717	0.7669
	Test	0.1211	1.0623 × 105	0.7228	0.6979
CGP	Training	0.1156	9.6760 × 104	0.7711	0.7664
	Test	0.1261	1.0597 × 105	0.7192	0.6940

Table 17. CGB Cross-Validation Results.

Dataset	Adj. R^2				
	1	2	3	4	5
Training	0.7683	0.7878	0.7516	0.7684	0.7585
Test	0.7048	0.6928	0.7155	0.7025	0.6737

As a result, the ANN model was established with a CGB training algorithm, four hidden neurons, 5-fold cross-validation, 100 epochs, a tansig function between the input layer and hidden layer, and purely activation function parameters between the hidden layer and the output layer.

In the MRA model, it is possible to evaluate whether the independent variables have a negative or positive relationship with the dependent variable, the size of the regression coefficients, and their contribution to the value. However, in the ANN model, the direction

of the variables and their effects on the value cannot be explained by the neuron weights produced in the model results. For this reason, the ANN model was tested with the assumptions listed below. The test results are given in Table 18.

Table 18. ANN Model Test Results.

Test	Floor	Front	Heating System	Age of the Building	Indoor Parking	Indoor Swimming Pool	Dressing Room	Planned Areas	Number of Rooms	Result
1 [a]	5	SW	3	10	0	0	0	150	4	625,649.58
	5	SW	3	10	0	0	0	200	5	885,663.84
	5	SW	3	10	0	0	0	250	6	1,021,507.29
2 [b]	0	SW	3	10	0	0	0	200	5	830,375.19
	1	SW	3	10	0	0	0	200	5	843,953.53
	7	SW	3	10	0	0	0	200	5	895,575.45
3 [c]	5	SW	3	5	0	0	0	200	5	940,851.35
	5	SW	3	10	0	0	0	200	5	885,663.84
	5	SW	3	15	0	0	0	200	5	839,748.75
4 [d]	5	SW	3	10	0	0	0	150	4	546,102.45
	5	SW	3	10	0	0	0	150	4	605,484.34
	5	SW	3	10	0	0	0	150	4	625,649.58
5 [e]	8	SW	3	10	0	0	0	200	5	897,196.63
	8	SW	3	10	1	0	0	200	5	877,519.88
6 [f]	10	SW	3	10	0	0	0	150	4	662,453.93
	10	SW	3	10	0	0	1	150	4	657,428.65
7 [g]	10	SW	3	10	0	0	0	150	4	662,453.93
	10	SW	3	10	0	1	0	150	4	675,935.28

[a] The apartment with larger planned areas should have more value. [b] There should be a 10–15% difference between the ground and upper floors. [c] The older apartment should have less value. [d] The price difference between apartments on different fronts should not exceed 5%. [e] The value of the apartment with indoor parking should be higher. [f] The value of the apartment with a dressing room should be higher. [g] The value of the apartment with an indoor swimming pool should be higher.

The ANN model was tested with the seven items above. The test results are similar to the results obtained with the MRA model. Although the indoor parking independent variable did not have a negative relationship with the dependent variable, the price of the apartment with indoor parking was lower. Since the same results were obtained for the parking garage variable in both methods, it was considered that the dataset caused this situation.

3.5. Findings

In the application stage of this study, MRA and ANN were used as the mass appraisal method. In these methods, a dataset consisting of 499 samples was created. As a result of outlier analysis on the dataset, it was determined that nine samples significantly affected the results of the model. These nine samples were excluded from the model for this reason. The remaining 490 samples were divided randomly into two groups: 80% as the training set and 20% the a test set.

The dataset was created with 19 independent variables that were considered to affect the real estate value. However, in the multicollinearity analysis, it was determined that there was a strong positive relationship with a value of 0.83 between the planned areas and the number of room variables. Hence, these two variables were turned into a single variable using factor analysis, and the model was rearranged as 18 independent variables.

Moreover, although some independent variables did not have a negative relationship with the dependent variable, the sign of the regression coefficients was negative, or the significance values were above 0.05. For this reason, the variable elimination and addition (stepwise), variable elimination (backward elimination), and variable addition methods in the SPSS software were used to remove the nonsignificant variables from the regression model and to determine the variables that best reflect the relationship between the independent variables and the dependent variable. According to the variable selection methods, the effects of the independent variables including planned areas–number of rooms, age of the building, dressing room, heating system, indoor swimming pool, security, front, floor,

and indoor parking were found to have a significant effect on the real estate value. On the other hand, other variables were not significant since they did not contribute significantly to the value of the real estate. Models were created using these variables in MRA and ANN methods. The best results were obtained with the same independent variables in both mass appraisal methods. The performance criteria and model results commonly used in mass appraisal studies are given in Table 19.

Table 19. Model Results According to Performance Criteria.

Model	Dataset	MAPE	RMSE	R^2	Adj. R^2
MRA	Training	0.1271	106,848.80	0.7412	0.7365
	Test	0.1270	111,148.74	0.6037	0.5711
ANN	Training	0.1156	96,755.49	0.7717	0.7669
	Test	0.1211	106,233.32	0.7228	0.6979

The ANN model gives better results compared with the MRA model. The multiple regression model gives the worst out-of-sample estimates (test set) results.

In addition to the performance criteria, it is recommended to determine the percentage of immovables that fall within the internationally acceptable margin of error (±0–10%) and outside this margin of error. However, no reliable data sources exist in developing countries such as Türkiye. Therefore, obtaining mass appraisal results with an error margin of ±10 may not be possible. A 15% margin of error is also acceptable for developing countries. The model results according to these error margins are given in Table 20.

Table 20. Model Results According to Different Margins of Error.

Model	Dataset	10% Margin of Error	15% Margin of Error	20% Margin of Error
MRA	Training	49.12%	68.51%	78.34%
	Test	46.24%	69.89%	78.49%
ANN	Training	51.53%	71.17%	83.16%
	Test	44.90%	72.45%	80.61%

According to the data in Table 20, approximately 71% of the real estate can be determined using the mass appraisal method within a 15% margin of error. It is possible to increase these values by increasing the data quality.

The International Association of Assessing Officers (IAAO) recommends performing a ratio analysis to test and analyze mass appraisal models. Whichever method is used in the mass appraisal, the results should be subjected to ratio analysis. In ratio analysis, the different ratios between the sales value of the real estate and the values estimated using the models are examined. With these ratios, the appraisal value of each variable, uniformity (horizontal and vertical equality), and reliability measurements are calculated. The ratio analysis results of the mass appraisal methods used in this study are given in Table 21.

Table 21. Ratio Analysis Results of Mass Appraisal Methods.

Model	Dataset	Median	COD	COV	PRD	PRB
MRA	Training	1.0058	12.636	15.8295	1.0243	−0.0941
	Test	1.0171	12.257	15.2882	1.0223	−0.1636
ANN	Training	1.0028	11.486	14.2636	1.0197	−0.0848
	Test	0.9875	12.256	16.1683	1.0324	−0.1503

According to the Standard for Proportion Studies, the median ratio must be between 0.90 and 1.10; the coefficient of dispersion (COD) must be between 5 and 10 for new or similar residences and between 5 and 15 for older or more heterogeneous areas; PRD must

be between 0.98 and 1.03; and PRB must be between −0.05 and +0.05. According to the values in Table 21, only the PRB value does not meet the requirement.

The PRB is a vertical equity indicator that measures the relationship between valuation/market value ratios and value in percent. It determines whether the ratios behave similarly for low-value, medium-value, and high-value real estate. The dataset has immovables with different values between TRY 200,000 and TRY 1,600,000. Furthermore, samples between TRY 1,600,000 and TRY 1,950,000 were determined as outliers in the outlier analysis. In the model results, real estate values were produced between TRY 180,000 and TRY 1,100,000. Since there are not enough high-value samples in the study area, the error rate is higher in high-value immovables. Therefore, the desired PRB ratio could not be achieved.

It has been understood that the ANN model in this study generally meets the mass appraisal standards, and the results are dependable. Within the study area, there are a total of 17,845 independent residential sections, of which 9205 are in the Alpaslan neighborhood and 8640 in the Köşk neighborhood. There are 7470 independent sections in the buildings whose sales value is obtained. It was found that 310 of these independent sections are exempt from real estate tax. For this reason, these independent sections were not considered in determining the real estate tax leakage loss rates. The obtained ANN model also determined the value of the other 7160 (40.12%) independent sections. The real estate tax values (base) of 7160 independent sections in 2020 were obtained from the Real Estate and Expropriation Directorate of Melikgazi Municipality. The values produced by the mass appraisal models, the real estate tax values, and the differences between these values are given in Table 22 based on five different independent sections.

Table 22. Examples of Value Generated Using the Mass Appraisal Models and Differences.

Sample No	Plot/Parcel No	Independent Section No	Planned Areas	ANN (TL)	MRA (TL)	Difference (TL) ANN-MRA	EVD	Difference (TL) ANN-EVD
1	854/1205	8	106	236.508	212.112	24.397	89.835	146.674
2	854/1205	9	106	226.080	251.109	−25.028	89.835	136.246
3	12193/1	46	175	907.798	1002.372	−94.574	715.561	192.237
4	12193/1	39	165	907.031	1001.982	−94.951	518.020	389.011
5	12193/1	52	176	924.257	1016.016	−94.205	586.289	321.121
...								

In provinces with metropolitan municipalities, the real estate tax is charged at the rate of two per thousand over the real estate tax value of residential immovables. The total real estate tax base value, total real estate tax amount, and collection values in Melikgazi Municipality and those obtained using the mass appraisal model are presented in Table 23.

Table 23. Base, Accrual, and Collection Values of Immovables.

Transaction	Melikgazi Municipality (TL)	Mass Appraisal Model (TL)	Difference (%)
Base	1,626,651,033	4,704,029,611	289.18
Accrual	3,253,302	9,408,059	289.18
Collection (78%)	2,537,575	9,408,059	370.75

According to the data on 7160 independent sections in the table, there is a difference of approximately 2.89 times between the Melikgazi Municipality values and those estimated using the MRA + ANN method. In addition, the real estate tax collection rate in Melikgazi Municipality in 2020 was 78%. Accordingly, the difference between the model's municipal tax revenue and tax revenue is 3.7 times.

Melikgazi Municipality received TRY 6,127,017 real estate tax income in the working area in 2020, TRY 3,487,396 from the Alpaslan neighborhood, and TRY 2,639,621 from the Köşk neighborhood. There are a total of 17,845 independent sections in the Alpaslan

and Köşk neighborhoods. A total of 7160 (40.12%) of these independent sections were determined with the mass appraisal models. According to the mass appraisal model, a real estate tax income of TRY 9,408,059 should be received from the independent section of the Alpaslan and Köşk neighborhoods, corresponding to 40.12%. When this value is proportioned to the entire study area, it is predicted that a total of TRY 23,447,879 real estate tax income can be collected from the Alpaslan and Köşk neighborhoods.

In 2020, TRY 72,215,401 tax (real estate, advertisement, advertisement, environment, and cleaning) was collected in Melikgazi Municipality. Real estate tax revenue in the Alpaslan and Köşk neighborhoods constituted 8.48% of the municipality's total tax revenue. Accordingly, because of the real estate tax revenue of TRY 23,504,284, which is expected to be collected from the Alpaslan and Köşk neighborhoods for the entire municipality, the real estate tax income of Melikgazi Municipality can be TRY 276,365,810.

The real estate value maps of the 7160 independent sections were created with the values obtained from the ANN, MRA, and real estate tax values. The geostatistical analysis/field interpolation tool was used in ArcGIS 10.7 software to create the real estate value maps. Independent sections have different square meters and values on the same parcel. For this reason, square meter unit values were calculated instead of the values of independent sections. The square meter unit values were divided into 13 classes at 500 TL/m^2 intervals. Real estate maps produced according to these classes are presented in Figure 9.

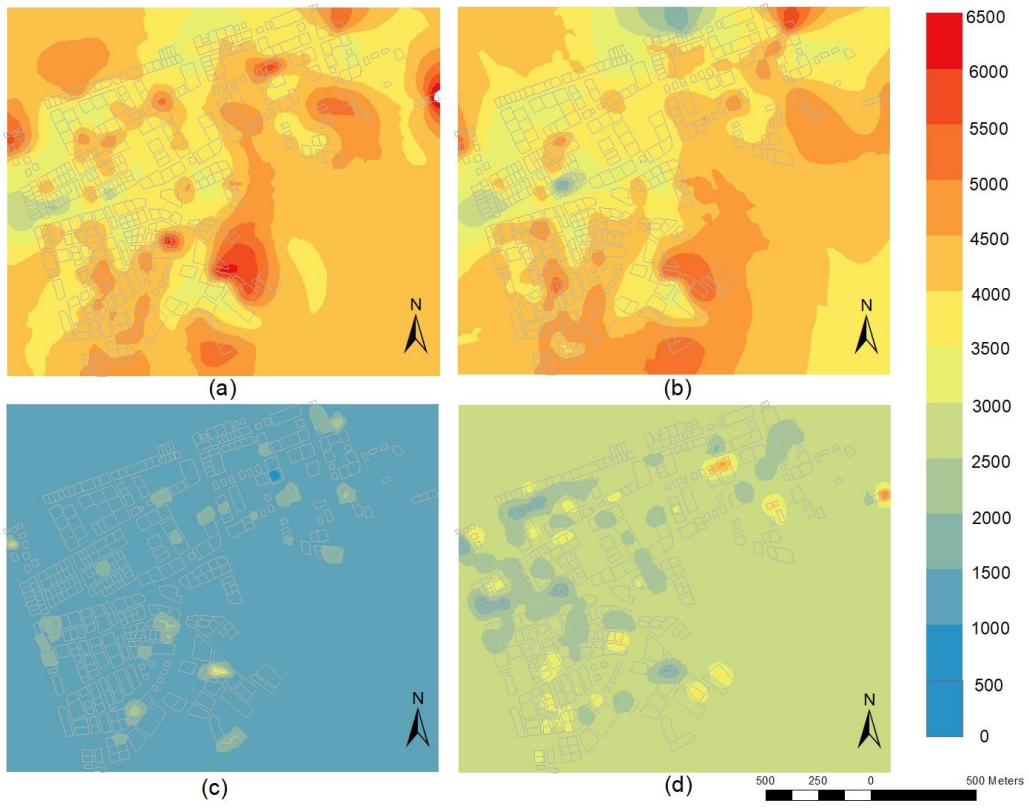


Figure 9. (**a**) Real estate value map with ANN values. (**b**) Real estate value map with MRA values. (**c**) Real estate value map with property tax values. (**d**) Real estate value map with the difference between ANN and property tax values.

Figure 9a shows that the square meter unit values of the independent sections in the study area are generally between 3000 and 5000 TRY/m^2 according to the ANN model. While the square meter unit values approach 3000 TRY/m^2 in old buildings, it approaches 5000 TRY/m^2 in new buildings. In addition, the square meter unit value of a small number of buildings is approximately 6000 TRY/m^2. As shown in Figure 9b, there are smooth

transitions between the values of the MRA model and the square meter unit values on the map. It was determined that the MRA could not catch up with the rapid price changes in the study area. In Figure 9c, the square meter unit values of real estate tax values are generally between 1000 and 2000 TRY/m^2. However, 3000 and 5000 TRY/m^2 values are seen in new luxury buildings. In Figure 9d, the real estate value map was created with the differences between the values obtained using the ANN model and the real estate tax values. Generally, the differences are between 2000 and 3000 TRY/m^2. However, it was determined that exceptionally low real estate tax was collected from the independent sections indicated with yellow and red parcels. In addition, it was observed that the real estate tax is closer to the market value in the newly built houses, which are indicated with dark blue.

4. Conclusions and Suggestions

The results of this study are related to the Melikgazi Municipality and its neighborhoods. However, similar mass appraisal models can easily be created based on any municipality and neighborhood. Data must be recorded for the models created for mass appraisal to be applied to the entire real estate. Determining the characteristics of all immovable properties and recording them in a database is a time-consuming and costly process. For example, in the National Address Database, there are 51,265 residential buildings in Melikgazi. For developing countries, obtaining data from different institutions and combining them in a standardized database takes a long time. To solve this problem, a location-based national building inventory information system should be established.

The revenue budget of Melikgazi Municipality in 2020 was TRY 41,486,242,382. The real estate tax income determined within the scope of this study was TRY 276,365,810 based on market values. This value corresponds to approximately 66% of the municipality's 2020 income. In 2020, TRY 72,215.401 in tax collection was made in Melikgazi Municipality. According to this, it was determined that the Melikgazi Municipality had a real estate tax leakage loss of approximately TRY 200,000,000 in 2020. According to these data, tax losses of municipalities can be prevented using mass appraisal methods. In this way, local governments' revenues can be increased, and their dependence on the central government can be reduced.

Real estate tax is also collected from land, commercial, and residential properties. Since the values of land and commercial immovables are higher than those of residences, the tax leakage loss rates of these immovables are higher. In addition, losses occur due to reasons such as a collection of land tax from independent sections with condominium servitude and invalid exemption transactions until the residence is purchased. These losses are only real estate tax losses. Furthermore, losses occur in taxes and fees based on the real estate tax value. When these fees and taxes are included, the actual losses are well above the value determined in this study.

The loss in real estate tax revenues of local governments was determined with a reliability of 71% using the model created for the two neighborhoods. The results of this study also revealed that there are significant problems in tax collection. The fact that this loss in real estate tax is due to the high rate of the real estate tax should not be ignored. The central government should lower real estate tax rates, and taxes should be levied on market values.

The success of ANN models also depends on data quality. No matter which method is applied as a mass appraisal method, creating a successful mass appraisal model is impossible using poor-quality data. The sales prices used in this study cover the years 2020–2021. The low housing loan interest rates applied in July and August 2020, the inflationary effect on housing prices, and the impact of the COVID-19 pandemic caused sudden increases of 50–60%. Even during the pandemic, when immovable values increased rapidly, the value of approximately 71% of immovables can be determined with a 15% margin of error using mass appraisal methods.

Real estate appraisals in Türkiye are performed by different institutions such as real estate appraisers, public institution appraisal commissions, and court experts. The data, appraisal methods, and definitions these institutions use differ significantly. Therefore, there are problems with a lack of reliable data and data standards in Türkiye. Legal regulations and a standard valuation database should be established to provide a transparent and traceable real estate market and create a mass appraisal model for Türkiye.

Determining the real estate tax value using mass appraisal cannot be performed by provincial and district municipalities lacking a standard real estate database infrastructure. The central government should perform a mass appraisal to determine real estate taxes accurately. The legal basis for the responsibilities of central and local governments for mass appraisal should be established. Institutions such as land registry offices, tax offices, municipalities, and banks should come together to determine the real estate tax values. In addition, for the mass appraisal model to be successfully implemented in Türkiye, it is necessary to develop a procedure for determining the variables that affect the value of the real estate and establishing the standards for these variables.

Another issue in ANNs is obtaining the market values of the independent sections for which the dataset is created. Accurate, dependable, and sufficient samples should be collected for each neighborhood. Reliable data sources are limited in developing countries such as Türkiye. The "Digital Service Tax Law and the Proposal of Amendment in Some Laws and the Decree-Law No. 375", submitted to Parliament in September 2019, attempted to legalize the real estate appraisal report requirement. However, during the meetings held at the General Assembly of the Grand National Assembly of Türkiye, Articles 30 and 31 regarding the appraisal report obligation were removed from the bill. An appraisal report should be required in trading transactions to provide a sufficient sample of mass appraisal models. However, the spread of appraisal reports can be encouraged because the amount paid for the appraisal report is covered by the state instead of the people who make the transaction. In addition, for all immovables sold in Türkiye (residence, land, land, etc.), the title deed fee is 4% of the sales value reported in the title deed transactions (2% for the buyer and 2% for the seller, separately). Values are declared lower than the market value of the immovables since fees are charged in the purchase and sale transactions. Citizens can be encouraged to make correct statements by determining lower tax rates in title deed transactions related to purchase and sale.

In the current institutional system, the only valuations in Türkiye based on international standards are those made by real estate appraisers affiliated with the CMB of Türkiye. The market values appraised in these valuation reports can also be used for mass appraisal models. However, very few valuation reports are issued in some provinces or districts of Türkiye. For this reason, sufficient sample size may not be provided in the models created using only the values appreciated in the valuation reports in the mass appraisal model to apply all around Türkiye.

Another issue in mass appraisal is the selection of appraisal methods. ANNs have been successfully applied in various fields, such as the real estate sector. However, they have some shortcomings when used in real estate valuation. These are:

Data Availability and Quality: ANNs need high-quality data to be trained effectively. Inaccurate or incomplete data can negatively affect performance. The accuracy of the neural network data can be limited to improve the quality of real estate data, especially for certain geographical areas or property types.

Lack of Interpretability: ANNs are often considered black box models because they provide limited interpretability. While they can predict real estate values based on input characteristics, they cannot determine the variables contributing to the valuation.

Sensitivity to Input Variables: Selecting relevant variables that accurately capture the characteristics that influence real estate value is crucial. However, as real estate valuation involves multiple factors such as location, size, economic conditions, and market trends, it can be difficult to identify the most appropriate dataset.

Overfitting: ANNs can be prone to overfitting, especially when the training dataset is relatively small or not representative of the entire real estate market.

ANN-based models can analyze exceptionally substantial amounts of data. In this way, complex relationships and factors affecting property tax leakage loss rates that may not be visible to valuation experts can be identified. Appraisers can integrate the findings from ANN analysis into their valuation models. By including tax leakage loss rates, they can better estimate the impact of tax-related variables on property values.

ANN-based analysis can reveal differences in tax leakage loss rates in different geographical regions of countries. Experts can adjust a region's valuation data based on property tax leakage loss rates. Appraisers can contribute to solving the problem by highlighting areas with high tax leakage loss rates and communicating their findings to valuation stakeholders, local administrators, and policymakers. This can lead to more fair and efficient tax policies that benefit the real estate sector and the economy.

In summary, investigating property tax leakage loss rates with ANN can help valuation experts in areas such as developing models that improve the accuracy of the valuation model, making regional comparisons, performing sector-specific risk assessment, and contributing to policy development. For appraisal practices to be conducted objectively and scientifically, the structure of land and land commissions should be removed from the hands of local and central administrations, and the participation of experienced real estate appraisers in the region should be ensured. In addition, the appraisal for taxation purposes should be made on a parcel basis not on a street-by-street basis. When it is desired to determine the market value based on parcels, the appraisal cost increases, but more accurate results can be obtained in a short time.

As a result, with the MRA + ANN models, a reliable and accurate database system, a more equitable taxation system, and a real estate tax system that increases the revenues of local and central governments can be established.

Author Contributions: Conceptualization, M.Y. and B.B.; methodology, M.Y. and B.B.; software, M.Y.; validation, B.B. and M.Y.; formal analysis, M.Y.; investigation, M.Y.; resources, M.Y.; data curation, M.Y. and B.B.; writing—original draft preparation, M.Y.; writing—review and editing, B.B.; visualization, M.Y. and B.B.; supervision, B.B.; project administration, B.B.; funding acquisition, M.Y. and B.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from Melikgazi Municipality and appraisers. The data presented in this study can be obtained from the authors upon request, with the permission of the relevant institutions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Guliker, E.; Folmer, E.; van Sinderen, M. Spatial Determinants of Real Estate Appraisals in The Netherlands: A Machine Learning Approach. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 125. [CrossRef]
2. Hong, J.E.; Kim, W.S. Combination Of Machine Learning-Based Automatic Valuation Models For Residential Properties In South Korea. *Int. J. Strateg. Prop. Manag.* **2022**, *26*, 362–384. [CrossRef]
3. Iban, M.C. An explainable model for the mass appraisal of residences: The application of tree-based Machine Learning algorithms and interpretation of value determinants. *Habitat Int.* **2022**, *128*, 11. [CrossRef]
4. Lee, C. Enhancing the performance of a neural network with entity embeddings: An application to real estate valuation. *J. Hous. Built Environ.* **2022**, *37*, 1057–1072. [CrossRef]
5. Zhou, X.; Tong, W. Learning with self-attention for rental market spatial dynamics in the Atlanta metropolitan area. *Earth Sci. Inform.* **2021**, *14*, 837–845. [CrossRef]
6. Bostancı, B. Taşınmaz Geliştirmede Değer Kestirim Analizleri ve Istanbul Konut Alani Örneğinde bir Uygulama. Ph.D. Thesis, Yıldız Teknik Üniversitesi, Esenler, İstanbul, 2008.
7. Zurada, J.; Levitan, A.; Guan, J. A comparison of regression and artificial intelligence methods in a mass appraisal context. *J. Real Estate Res.* **2011**, *33*, 349–388. [CrossRef]
8. Abidoye Rotimi, B.; Chan Albert, P.C. Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network. *Pac. Rim Prop. Res. J.* **2018**, *24*, 71–83. [CrossRef]

9. Chun Lin, C.C.; Mohan Satish, B. Effectiveness comparison of the residential property mass appraisal methodologies in the USA. *Int. J. Hous. Mark. Anal.* **2011**, *4*, 224–243. [CrossRef]
10. Limsombunchai, V.; Gan, C.; Lee, M. House price prediction: Hedonic price model vs artificial neural network. *Am. J. Appl. Sci.* **2004**, *1*, 193–201. [CrossRef]
11. Amri, S.; Tularam, A. Performance of Multiple Linear Regression and Nonlinear Neural Networks and Fuzzy Logic Techniques in Modelling House Prices. *J. Math. Stat.* **2012**, *8*, 419–434. [CrossRef]
12. Mora-Esperanza, J.G. Artificial intelligence applied to real estate valuation: An example for the appraisal of Madrid. *Catastro* **2004**, *1*, 255–265.
13. Cechin, A.; Souto, A.; Gonzalez, M.A. Real estate value at Porto Alegre city using artificial neural networks. In Proceedings of the Proceedings. Vol. 1. Sixth Brazilian Symposium on Neural Networks, Rio de Janeiro, Brazil, 25 November 2000; pp. 237–242.
14. Tay Danny, P.H.; Ho David, K.H. Artificial Intelligence and the Mass Appraisal of Residential Apartments. *J. Prop. Valuat. Invest.* **1992**, *10*, 525–540. [CrossRef]
15. Dimopoulos, T.; Bakas, N. Sensitivity analysis of machine learning models for the mass appraisal of real estate. Case study of residential units in Nicosia, Cyprus. *Remote Sens.* **2019**, *11*, 3047. [CrossRef]
16. McCluskey, W.; Davis, P.; Haran, M.; McCord, M.; McIlhatton, D. The potential of artificial neural networks in mass appraisal: The case revisited. *J. Financ. Manag. Prop. Constr.* **2012**, *17*, 274–292. [CrossRef]
17. Yacim, J.A.; Boshoff, D.G.B. Impact of artificial neural networks training algorithms on accurate prediction of property values. *J. Real Estate Res.* **2018**, *40*, 375–418. [CrossRef]
18. Hamzaoui, Y.E.; Perez, J.A.H. Application of Artificial Neural Networks to Predict the Selling Price in the Real Estate Valuation Process. In Proceedings of the 2011 10th Mexican International Conference on Artificial Intelligence, Puebla, Mexico, 26 November–4 December 2011; pp. 175–181.
19. Pagourtzi, E.; Metaxiotis, K.; Nikolopoulos, K.; Giannelos, K.; Assimakopoulos, V. Real estate valuation with artificial intelligence approaches. *Int. J. Intell. Syst. Technol. Appl.* **2007**, *2*, 50–57. [CrossRef]
20. Renigier-Biłozor, M.; Źróbek, S.; Walacik, M. Modern Technologies in the Real Estate Market—Opponents vs. Proponents of Their Use: Does New Category of Value Solve the Problem? *Sustainability* **2022**, *14*, 13403. [CrossRef]
21. Özkan, G.; Yalpir, Ş.; Uygunol, O. An investigation on the price estimation of residable real-estates by using ANN and regression methods. In Proceedings of the 12th Applied Stochastic Models and Data Analysis International Conference (ASMDA), Chania, Greece, 29 May–1 June 2007; pp. 1–8.
22. Sampathkumar, V.; Santhi, M.H.; Vanjinathan, J. Evaluation of the trend of land price using regression and neural network models. *Asian J. Sci. Res.* **2015**, *8*, 182–194. [CrossRef]
23. Morano, P.; Tajani, F. Bare ownership evaluation: Hedonic price model vs artificial neural network. *Int. J. Bus. Intell. Data Min.* **2013**, *8*, 340–362. [CrossRef]
24. Lai, P.Y. Analysis of the mass appraisal model by using artificial neural network in Kaohsiung city. *J. Mod. Account. Audit.* **2011**, *7*, 1080–1089.
25. Lenk, M.M.; Worzala, E.M.; Silva, A. High-tech valuation: Should artificial neural networks bypass the human valuer? *J. Prop. Valuat. Invest.* **1997**, *15*, 8–26. [CrossRef]
26. Grover, R. Mass valuations. *J. Prop. Invest. Financ.* **2016**, *34*, 191–204. [CrossRef]
27. Wei, C.K.; Fu, M.C.; Wang, L.; Yang, H.B.; Tang, F.; Xiong, Y.Q. The Research Development of Hedonic Price Model-Based Real Estate Appraisal in the Era of Big Data. *Land* **2022**, *11*, 334. [CrossRef]
28. McCluskey, W.J.; McCord, M.; Davis, P.T.; Haran, M.; McIlhatton, D. Prediction accuracy in mass appraisal: A comparison of modern approaches. *J. Prop. Res.* **2013**, *30*, 239–265. [CrossRef]
29. Abidoye Rotimi, B.; Chan Albert, P.C. Artificial neural network in property valuation: Application framework and research trend. *Prop. Manag.* **2017**, *35*, 554–571. [CrossRef]
30. Ge, J.X. Housing Price Models for Hong Kong. Ph.D. Thesis, University of Newcastle, Callaghan, Australia, 2004.
31. Borst, R.A. Artificial neural networks: The next modelling/calibration technology for the assessment community. *Prop. Tax J. (Int. Assoc. Assess. Off.)* **1991**, *10*, 69–94.
32. Yalpir, Ş. Enhancement of parcel valuation with adaptive artificial neural network modeling. *Artif. Intell. Rev.* **2018**, *49*, 393–405. [CrossRef]
33. Nguyen, V.T. A New Conceptual Automated Property Valuation Model for Residential Housing Market. Ph.D. Thesis, Victoria University, Melbourne, Australia, 2014.
34. Vo, N.; Shi, H.; Szajman, J. Sensitivity analysis and optimisation to input variables using winGamma and ANN: A case study in automated residential property valuation. *Int. J. Adv. Appl. Sci.* **2015**, *2*, 19–24.
35. Feng, Y.; Jones, K. Comparing multilevel modelling and artificial neural networks in house price prediction. In Proceedings of the 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM), Fuzhou, China, 8–10 July 2015; pp. 108–114.
36. Ünel, F.B.; Yalpir, Ş. Reduction of mass appraisal criteria with PCA and integration to GIS. *Int. J. Eng. Geosci.* **2019**, *4*, 94–105. [CrossRef]
37. Vo, N.Y.; Shi, H.; Szajman, J. Optimisation to ANN inputs in automated property valuation model with Encog 3 and winGamma. *Appl. Mech. Mater.* **2014**, *462–463*, 1081–1086.

38. Abidoye Rotimi, B.; Chan Albert, P.C. Modelling property values in Nigeria using artificial neural network. *J. Prop. Res.* **2017**, *34*, 36–53. [CrossRef]
39. Rahman, S.N.A.; Maimun, N.H.A.; Razali, M.N.; Ismail, S. The artificial neural network model (ANN) for Malaysian housing market analysis. *Plan. Malays.* **2019**, *17*, 1–9.
40. Yacim, J.A.; Boshoff, D.G.B.; Khan, A. Hybridizing cuckoo search with levenberg-marquardt algorithms in optimization and training of anns for mass appraisal of properties. *J. Real Estate Lit.* **2016**, *24*, 473–492. [CrossRef]
41. Yacim, J.A.; Boshoff, D.G.B. Combining BP with PSO algorithms in weights optimisation and ANNs training for mass appraisal of properties. *Int. J. Hous. Mark. Anal.* **2018**, *11*, 290–314. [CrossRef]
42. Kwok, T.Y.; Yeung, D.Y. Constructive algorithms for structure learning in feedforward neural networks for regression problems. *IEEE Trans. Neural Netw.* **1997**, *8*, 630–645. [CrossRef]
43. Abidoye Rotimi, B.; Chan Albert, P.C. Achieving property valuation accuracy in developing countries: The implication of data source. *Int. J. Hous. Mark. Anal.* **2018**, *11*, 573–585. [CrossRef]
44. Tabales, J.N.M.; Ocerin, C.J.M.; Carmona, F.J.R. Artificial neural networks for predicting real estate prices. *Rev. Metodos Cuantitativos Para Econ. Empresa* **2013**, *15*, 29–44.
45. Morano, P.; Tajani, F.; Torre, C.M. Artificial intelligence in property valuations: An application of artificial neural networks to housing appraisal. In Proceedings of the 11th International Conference on Energy, Environment, Ecosystems and Sustainable Development (EEESD '15), Canary Islands, Spain, 10–12 January 2015; pp. 23–29.
46. Lam, K.C.; Yu, C.Y.; Lam, K.Y. An Artificial Neural Network and Entropy Model for Residential Property Price Forecasting in Hong Kong. *J. Prop. Res.* **2008**, *25*, 321–342. [CrossRef]
47. Gloudemans, R.; Sanderson, P. The potential of artificial intelligence in property assessment. *J. Prop. Tax Assess. Adm.* **2021**, *18*, 9.
48. Worzala, E.; Lenk, M.; Silva, A. An exploration of neural networks and its application to real estate valuation. *J. Real Estate Res.* **1995**, *10*, 185–201. [CrossRef]
49. Kontrimas, V.; Verikas, A. The mass appraisal of the real estate by computational intelligence. *Appl. Soft Comput.* **2011**, *11*, 443–448. [CrossRef]
50. Güneş, T.; Yıldız, Ü. Mass valuation techniques used in land registry and cadastre modernization project of Republic of Türkiye. In Proceedings of the FIG Working Week: From the Wisdom of the Ages to the Challenges of the Modern World, Sofia, Bulgaria, 17–21 May 2015.
51. Hong, J.; Choi, H.; Kim, W.-s. A house price valuation based on the random forest approach: The mass appraisal of residential property in South Korea. *Int. J. Strateg. Prop. Manag.* **2020**, *24*, 140–152. [CrossRef]
52. Grover, R.; Walacik, M. Property valuation and taxation for fiscal sustainability—Lessons for Poland. *Real Estate Manag. Valuat.* **2019**, *27*, 35–48. [CrossRef]
53. Yılmaz, M. Emlak Vergisi Kayıp Kaçak Oranlarının Toplu Değerleme ile Araştırılması: Kayseri Örneği. Master's Thesis, Erciyes Üniversitesi, Kayseri, Turkey, 2021.
54. Grover, R.; Walacik, M.; Buzu, O.; Güneş, T.; Raskovic, M.; Yıldız, Ü. Barriers to the use of property taxation in municipal finance. *J. Financ. Manag. Prop. Constr.* **2019**, *24*, 166–183. [CrossRef]
55. Mimis, A.; Rovolis, A.; Stamou, M. Property valuation with artificial neural network: The case of Athens. *J. Prop. Res.* **2013**, *30*, 128–143. [CrossRef]
56. Ahmed, S.; Rahman, M.M.; Islam, S. House Rent Estimation in Dhaka City by Multi Layer Perceptions Neural Network. *Int. J. u-e-Serv. Sci. Technol.* **2014**, *7*, 287–300. [CrossRef]
57. Morillo Balsera, M.C.; Martínez-Cuevas, S.; Molina Sánchez, I.; García-Aranda, C.; Martinez Izquierdo, M.E. Artificial neural networks and geostatistical models for housing valuations in urban residential areas. *Geogr. Tidsskr.-Dan. J. Geogr.* **2018**, *118*, 184–193. [CrossRef]
58. Alexandridis, A.K.; Karlis, D.; Papastamos, D.; Andritsos, D. Real Estate valuation and forecasting in non-homogeneous markets: A case study in Greece during the financial crisis. *J. Oper. Res. Soc.* **2019**, *70*, 1769–1783. [CrossRef]
59. Kang, J.; Lee, H.J.; Jeong, S.H.; Lee, H.S.; Oh, K.J. Developing a forecasting model for real estate auction prices using artificial intelligence. *Sustainability (Switzerland)* **2020**, *12*, 2899. [CrossRef]
60. Yacim, J.A.; Boshoff, D.G.B. Neural networks support vector machine for mass appraisal of properties. *Prop. Manag.* **2020**, *38*, 241–272. [CrossRef]
61. IAAO. *Standard on Mass Appraisal of Real Property*; International Association of Assessing Officers: Kansas, MO, USA, 2017.
62. Wang, D.K.; Li, V.J. Mass appraisal models of real estate in the 21st century: A systematic literature review. *Sustainability* **2019**, *11*, 7006. [CrossRef]
63. Almy, R. Property taxation and valuation in Lithuania. *Land Tenure J.* **2016**, *15*, 29–44.
64. Kuijper, M.; Kathmann, R. Property valuation and taxation in the Netherlands. *Land Tenure J.* **2016**, *15*, 47–61.
65. Buzu, O. Property assessment and taxation in the Republic of Moldova. *Land Tenure J.* **2016**, *15*, 63–81.
66. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]
67. Yıldız, Ü. Gayrimenkul Bilimlerinde Kitlesel Değerleme Uygulamaları ve Türkiye için Model Önerisi. Master's Thesis, Ankara Üniversitesi, Ankara, Turkey, 2014.

68. Hoffmann, J.P.; Shafer, K. *Linear Regression Analysis: Assumptions and Applications*, 2nd ed.; NASW Press: Washington, DC, USA, 2010.
69. McCord, M.; Lo, D.; Davis, P.; McCord, J.; Hermans, L.; Bidanset, P. Applying the Geostatistical Eigenvector Spatial Filter Approach into Regularized Regression for Improving Prediction Accuracy for Mass Appraisal. *Appl. Sci.* **2022**, *12*, 10660. [CrossRef]
70. Geerts, M.; vanden Broucke, S.; De Weerdt, J. A Survey of Methods and Input Data Types for House Price Prediction. *ISPRS Int. J. Geo-Inf.* **2023**, *12*, 200. [CrossRef]
71. Elmas, Ç. *Yapay Zeka Uygulamaları*; Seçkin Yayıncılık: Ankara, Turkey, 2018.
72. Atalay, M.; Çelik, E. Büyük veri analizinde yapay zeka ve makine öğrenmesi uygulamaları. *J. Mehmet Akif Ersoy Univ. Soc. Sci. Inst.* **2017**, *9*, 155–172. [CrossRef]
73. Kelley, K. Methods for the Behavioral, Educational, and Social Sciences: An R package. *Behav. Res. Methods* **2007**, *39*, 979–984. [CrossRef]
74. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. *Applied Linear Statistical Models*, 5th ed.; McGraw-Hill: New York, NY, USA, 2005.
75. Sullivan, J.H.; Warkentin, M.; Wallace, L. So many ways for assessing outliers: What really works and does it matter? *J. Bus. Res.* **2021**, *132*, 530–543. [CrossRef]
76. Kim, H.-Y. Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restor. Dent. Endod.* **2013**, *38*, 52–54. [CrossRef] [PubMed]
77. Belsley, D.A. A Guide to Using the Collinearity Diagnostics. *Comput. Sci. Econ. Manag.* **1991**, *4*, 33–50. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.