Sigma J Eng Nat Sci, Vol. 42, No. 3, pp. 667–678, June, 2024

Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2024.00062

Research Article

# An investigation on the estimation of the impact factors of pandemic deaths with artificial neural network and multiple regression algorithms: Covid-19 case

**İbrahim DEMİR[1], Murat SARI[2], Seda GÜLEN[3,*], Aniela BALACESCU[4]**

*[1]Turkish Statistical Institute, Cankaya, Ankara, 06100, Türkiye*
*[2]Department of Mathematical Engineering, Istanbul Technical University, Istanbul, 34467, Türkiye*
*[3]Department of Mathematics, Tekirdag Namik Kemal University, Tekirdag, 59030, Türkiye*
*[4]Constantin Brancusi University of Targu-Jiu,Targu Jiu, Gorj, 210135, Romania*

## ARTICLE INFO

*Article history*
Received: 20 July 2022
Revised: 14 November 2022
Accepted: 22 May 2023

**Keywords:**
Artificial Intelligence Modelling; COVID-19; Epidemic Disease; Forecasting Model; Regression Model

## ABSTRACT

This article aims to successfully estimate the number of deaths in a pandemic, with the appropriate implementation of two new modelling approaches, artificial neural network and multiple regression analysis. Then, these methods have been used comparatively to predict death cases for the future course of the COVID-19 outbreak. These approaches proposed for estimation appear to result in few errors and perform well in providing information on the course of deaths in the epidemic. The agreement between the predicted results by these methods, and the actual data proves the superiority of the proposed ones in forecasting accuracy in future cases. This is expected to provide significant benefits in increasing the effectiveness of health policies to be implemented within the scope of the measures to be taken for the future of this and similar epidemics. As this investigation reveals that the current modelling methods have undeniable advantages in predicting epidemic trends, using our models is believed to provide an accurate estimate of death rates and guide policymakers in formulating research, health, socio-economic and fiscal policies. All these findings can be widely regarded as significant milestones and essential guides for researchers examining potential future epidemic tendencies. In addition, although this epidemic is quite complex and varies from country-to-country and various factors, the proposed approaches offer a great opportunity to model the outbreak in other epidemics as well as in other countries.

**Cite this article as:** Demir I, Sarı M, Gülen S, Balacescu A. An investigation on the estimation of the impact factors of pandemic deaths with artificial neural network and multiple regression algorithms: Covid-19 case. Sigma J Eng Nat Sci 2024;42(3):667–678.

***Corresponding author.**
*E-mail address: sgulen@nku.edu.tr

*This paper was recommended for publication in revised form by Editor in Chief Ahmet Selim Dalkilic*

Published by Yıldız Technical University Press, İstanbul, Turkey

## INTRODUCTION

Human history has faced several disease outbreaks since its inception. More recently, at the end of 2019, the outbreak of a new coronavirus disease (COVID-19) which was caused by SARS-CoV-2 broke out in Wuhan, China, has spread rapidly to other countries over the world and was declared a global pandemic by the World Health Organization (WHO). This pandemic has affected millions of people around the world and was caused many deaths that exceeded million [1]. Besides, it has been seen that it has deeply affected the health, socio-economic and financial systems of the world that will be felt for years. Because the epidemiological characteristics of the disease have not fully been understood and no definitive treatment has yet been developed; governments, many researchers and clinicians are trying to unravel the complexity of the disease and explore various measures. In this context, researchers are focusing on the short and long-term predictions to assist decision makers in better prevention planning, healthcare delivery, and outbreak control.

Since there are no effective antiviral drugs for treating the COVID-19 disease, it caused a destructive effect in human civilization in terms of the health and safety of people with the rapid spreading. Therefore, in order to predict the course of the disease, take precautions such as isolation, quarantine and planning health services, many researchers in various fields of science increasingly have turned to computational methods. In recent studies, different approaches have been proposed to understand the dynamics of COVID-19 and manage it. While various studies [2-9] focused on compartmental models in epidemiology to understand the behavior of COVID-19 epidemic, some studies have used statistical and artifical intelligence-based methods [10-16] to forecast course of the disease.

Up till now, different statistical methods such as regression analysis models [17-20] and artificial intelligence based models [21-26] were used to forecast various epidemic diseases. With the current COVID-19 outbreak, regression algorithms, and artificial neural networks(ANNs) have been proposed to predict the COVID-19 disease **[27-32]**. Although many studies have been conducted on the COVID-19 outbreak in a short time since the beginning of the epidemic, and are being studied continuously, the multiple regression analysis and artificial neural networks applied to the COVID-19 pandemic in the literature are still very limited. This article aims to provide a perspective of COVID-19 outbreak and discover the spread and effects of the virus through artificial neural network modelling and regression analysis to present the results to policymakers in various fields of life.

As soon as the WHO declared the COVID-19 as a global epidemic, several countries including Turkey in the field of health immediately began to take preventive measures against the virus. As in the whole world, besides the implementation of several public and health policies, many academic researchers have focused on searching course of the disease in Turkey by using regression analysis and artificial neural network techniques. Of these, Nakip et al. [33] compared Linear regression(LR), Multi-Layer Perceptron(MLR) and Long-Short Term Memory(LSTM) techniques to determine the number of active cases in COVID-19 and concluded that the long term prediction of the number of the active cases in COVID-19 pandemic is not possible with high test accuracy for the considered models. They attributed this to the lack of sufficient number of samples. However, linear regression model has much better prediction performance with high generalization ability as compared to the complex models. Toga et al. [34] used Autoregressive Integrated Moving Average(ARIMA) and Artificial Neural Network (ANN) for predicting the numbers of infected cases, deaths and recovered cases. In that work, they observed that while ARIMA and ANN have almost same forecasting performance, ARIMA has high prediction accuracy. In addition to these results, they demonstrated the advantage of constructing models that can predict three variables simultaneously compared to the ARIMA model of ANN. Guleryuz [35] investigated the number of total case, the growth rate of total cases, the number of new cases, the number of total deaths, the increase rate of total deaths and the number of new deaths by using the Box-Jenkins Methods (ARIMA), Brown Exponential Smoothing model and RNN-LSTM. The results of that work concluded that, in the near future, pandemic will not show an increasing trend in the number of cases and ARIMA model can be used for forecasting a pandemic efficiently. Demir et al. [36] focused on tracking the spread behavior of the COVID-19 virus by using various models. Kuvvetli et al. [37] designed a predictive model based on an ANN model to forecast the future number of daily cases and deaths caused by COVID-19 in different countries' spreads including Turkey. Caglar and Ozen [38] investigated the number of cases and deaths in various countries including Turkey using the number of test, the number of seriously ill and recovered patients as parameters. In that study, various regression methods and an artificial neural network were used. Their results demonstrated that the optimal results for cases in Turkey are obtained with the corresponding methods. In addition, in this study, the similar features and differences of countries are mentioned. Kirbas et al. [39] modeled the data from cumulative confirmed COVID-19 cases in some European countries including Turkey using various techniques. In the corresponding work, it was determined that the LSTM approach has a much higher success than the other techniques.

The research of the forecasting future cases of COVID-19 is very important for countries to take precautions and improving the healthcare in any respect such as mask, ventilation unit, drug supply etc. Since accurate forecasts provide information for taking preventive measures, this goal can be accomplished by selecting appropriate models and methods. Statistical models are important modelling techniques used in real-time epidemiological disease data analysis. In multiple regression, the regression model allows selection of the best coefficients for all characters. Besides statistical models, artificial intelligence-based models such as ANNs have been proposed as a powerful tool for processing huge datasets that can be analyzed for prediction and used to model different problems. Therefore, in this study, it is aimed to predict future death cases based on MLR and ANNs model, which are powerful tool for forecasting. Unlike many other studies aimed at forecasting the COVID-19 pandemic, this study has investigated the effects of the daily number of cases (DNC), active cases (AC), daytime testing count (DTC), daily number of deaths (DND), daily recovered cases (DCC), and active intubated number (AIN) in determining the number of deaths. In order to avoid the establishment of unnecessary and unsatisfactory prediction models, the prediction success of the approach using the MLR has been investigated by adding these variables one by one and the prediction success of the obtained model has been presented. Then, the ANN model has been established by using the variables that reached the maximum specificity coefficient in the MLR model. Thus, the models using the optimum input value have been obtained. The reason for choosing these two methods from the mentioned techniques is that they are easy to implement, they allow the different factors affecting the dependent variable to be controlled clearly at the same time, and they do not need any advanced software to apply them. Furthermore, ANN approaches are adaptable and suitable for modeling nonlinear data with a large number of inputs in regression situations. Moreover, unlike many other estimate techniques, ANNs do not place any constraints on the input variables. With the motivation offered by these advantages, the statistical and artificial intelligence-based methodologies have been proposed here to forecast the number of COVID-19 deaths in Turkey for the data set from March 11 to September 14. The accuracy of the results produced from both models has been measured comparatively with different criteria such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination ($R^2$). Thus, the number of deaths predicted by the proposed models has been seen to be in good agreement with the actual data. Although the regression models established make an accurate prediction for the number of deaths, it has been seen that artificial neural networks produce more successful results than the regression models. To best knowledge of the authors, multiple linear regression and neural networks emerge as the first research models used comparatively to predict future deaths caused by the COVID-19 in Turkey. Furthermore, it is strongly believed that this study will shed light on decision-makers developing control programs to reduce loss.

## MATERIALS AND METHODS

### Multiple Regression

One of the most used approaches in statistical modeling is regression analysis. Regression analysis, which forms the basis of the MLR model used in this study, determines the model of the relationship between variables and makes predictions for the future. The regression analysis using more than one independent variable is called multiple regression analysis. If a phenomenon is affected by more than one factor, investigation of cause and effect relationship can be investigated with a multiple regression analysis [40].

The MLR model can then be given as:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + \varepsilon_i \qquad (1)$$

where, $y_i$, $x_i$, $p$ and $\varepsilon_i$ are the dependent variable, the independent variables, the number of explanatory variable and model error, respectively. The parameter estimation of a regression model that is obtained by multiple regression analysis to be reliable is based on providing some assumptions about the model. In order to use the regression equation obtained by regression analysis for estimation purpose; error terms should show normal distribution depending on chance, mean of the expected value of errors should be equal to 0 and the variance should be homogeneous and equal to $\sigma^2$, the errors should be independent, no correlation between error terms and explanatory variables, and explonatory variables should be assumed to be independent of each other [41]. When these assumptions are not provided, it is proposed to change parameter estimation methods.

### Artificial Neural Network

The ANNs are computer systems by inspiring the properties of neural systems (derivation of knowledge, prediction etc.) [42]. The ANNs are generated by the aggregation of cells as in the biological systems and, generally, the architecture of ANNs consists of three-layers: input, hidden and output layers.

There can be more than one hidden layer in a network. It is not determined up to today how many hidden layers will be used in an ANN and how many cells are in each hidden layer. This situation, which depends on the problem, is solved by trial and error [43]. Since the network with several hidden neurons produce linear predictions, it cannot distinguish complex patterns. In addition, the high number of hidden neurons blocks the network to make generalization [43,44]. Since there are additional layer(s) between the input layer and the output layer in solving nonlinear problems, the network architecture is multilayered as shown in Figure 1. In the figure, DNC, AC, AIN, DTC and DND represent the daily number of cases, the active cases, the active intubated number, the daytime testing count and the daily number of death, respectively.

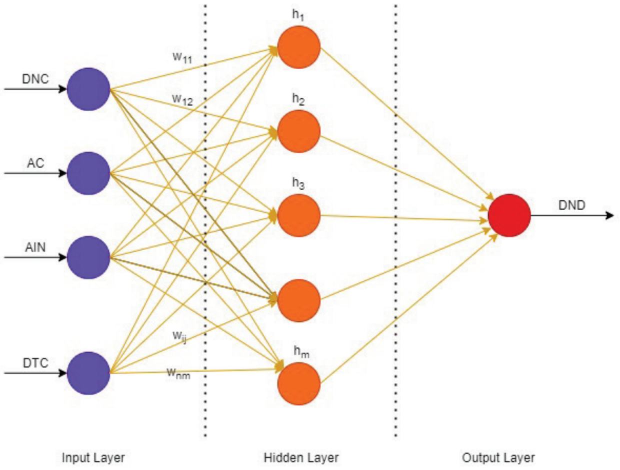

**Figure 1.** Feedforward network

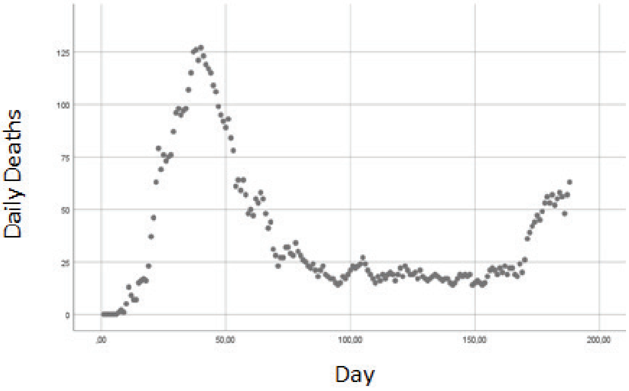

**Figure 2.** Variation in the number of deaths.

### Data Collection

The data are listed as new cases, active cases, daily deaths, daily tests, daily recovering patients, active intubated numbers, and active intensive care patients from March 11, when the first case was identified, to September 14. Data were taken daily from PRT-DTO [45] and RTM-COVID-19 Information Page [46].

### Data Analysis and Application

In this study, 188 COVID-19 data reported for Turkey between 11 March and 14 September have been used. Two forecasting algorithms, the MLR and ANNs, have been applied to predict the number of future death by using the mentioned data. When the data reported by the Ministry of Health has been examined, it is seen that the data involved daily number of cases (DNC), active cases (AC), daytime testing count (DTC), daily number of death (DND), daily recovered cases (DCC) and active intubated number (AIN). However, sharing of some of the data was stopped some time later, for example, active intensive care patients. The prediction model for the DND has been established by the DNC, AC, DTC, AIN data that are considered to affect the DND. Since the number of intensive care patients was not shared sometime later and there were total of 140 pieces of data, this variable has not been incorporated into the model. Since the number of daily recovery does not affect the number of daily death, it has not been incorporated into the model. The course of the DND can be shown as in Figure 2 from the decleration of the first case by the Ministry of Health. The number of deaths, which had increased rapidly at the beginning of the epidemic, started to decrease with the measures taken. While the number of deaths remained stable during the summer days, it seems that it started to increase after September.

As noted earlier, daily new cases (DNC), active cases (AC), active intubated cases (AIN) affect the number of death cases (DND) directly and indirectly. The basic statistics related to these data has been presented in Table 1. In the first few days, since some data cannot be observed/obtained, this data was taken as zero to avoid data loss. When the statistics are analyzed, it is seen that the data is non-normal regarding both values of skewness and kurtosis. The values obtained by dividing kurtosis and skewness values by their standard deviations must be in the interval [-1.96,1.96]. Here, the variable AIN provides the normality by fulfilling this condition. Since these values of the other variables are all positive and greater than 2, it is understood that they have a right skewed and pointed structure.

In the stage of determination of the input of the established model, whether the relationship between the number

**Table 1.** Descriptive statistics of the variables

| | | | | | | Skewness | | Kurtosis | |
|---|---|---|---|---|---|---|---|---|---|
| | N | Minimum | Maximum | Mean | Std. Deviation | Statistics | Std. Error | Statistics | Error |
| DND | 188 | 1 | 127 | 39.12 | 31.244 | 1.323 | 0.180 | 0.713 | 0.358 |
| DNC | 188 | 0 | 5138 | 1557.66 | 1032.972 | 1.611 | 0.177 | 2.545 | 0.353 |
| AC | 188 | 1 | 80808 | 25893.63 | 19683.057 | 1.239 | 0.177 | 0.971 | 0.353 |
| DTC | 188 | 1738 | 117113 | 47429.25 | 27140.104 | 0.923 | 0.180 | 0.446 | 0.358 |
| AIN | 188 | 0 | 1351 | 575.06 | 333.098 | 0.291 | 0.177 | -0.691 | 0.353 |

**Table 2.** The relationship between the variables

| | DND | DNC | AC | DTC | AIN |
|---|---|---|---|---|---|
| DND | 1 | 0.900** | 0.865** | -0.047 | 0.730** |
| DNC | 0.900** | 1 | 0.714** | -0.068 | 0.688** |
| AC | 0.865** | 0.714** | 1 | -0.139 | 0.498** |
| DTC | -0.047 | -0.068 | -0.139 | 1 | 0.570** |
| AIN | 0.730** | 0.688** | 0.498** | 0.570** | 1 |

**Correlation is significant at the 0.01 level (2-tailed)

of daily deaths and other variables is linear, has been analyzed by the scatter plot (Figure 3), the power of the relationship between variables and their directions have been analyzed by the Pearson correlation. Correlations of the variables are presented in Table 2. According to the table, except for the variable of the number of deaths and the variable DTC, there is a high correlation between other variables.

The relationship between the variable DND and the variables DNC, AC, AIN with high correlation is statistically significant ($p < 0.05$). These relationships are positive and strong. The prediction can be obtain by using these variables and DND. Here, there is a negative and very weak relationship between the DNC and DTC, and this relationship is not statistically significant and is at the %5 significance level. At first glance, since there is no relationship between this variable and the DND, it may seem impossible to build the model. Due to the large number of data, this variable can also be included in the time series and multiple regression model. However, due to the large amount of data available, this variable can be included in the MLR model. By adding the variable to the model, it will be analyzed by the adjusted R-square whether it contributes significantly. Fifteen various models can be established by one dependent variable and four independent variables. Some of them can be statistically insignificant and even though some of them are statistically significant, the model explanatory power may be low. Therefore, in establishing the regression model, predictive power has been investigated with the variables added to the model one by one using the stepwise method. In this study, the model using optimum number of data has also been tried to be discovered. Because, if the number of deaths can be estimated using a small enough number of data, the model that contains a large number of inputs is unnecessary.

How independent variables affect the dependent variable DND, will first be examined with multiple time series. This investigation has been carried out separately for each of the 1,3,5,7,10,12 and 15 lags. In establishing ANN models, the methods that are based on the rules of %80, %10, %10 or %70, %15, %15 intended for training, validation, and test set in literature are proposed to test the generalization ability of the model [47]. %70, %15, and %15 of the 188-piece dataset have been used in training, validation, and testing, respectively.

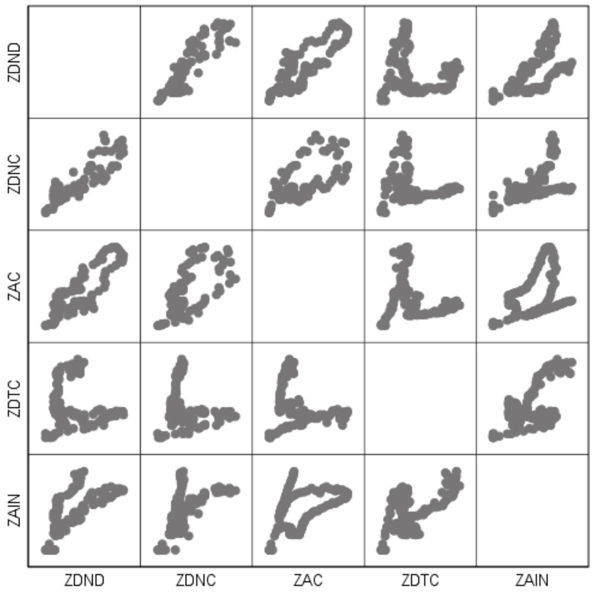

**Figure 3.** The skewness plot of the relationship between the variables.

The reason for the separation of the data set as training, test, and validation is for searching whether the model memorizes the data. If the model memorizes, the model cannot make an accurate prediction for the data that is not in the dataset. For this reason, the data is investigated by partitioning three parts. At this stage, Levenberg-Marquardt artificial neural network algorithm and sigmoid activation function programmed in MATLAB. The Levenberg-Marquardt algorithm has been preferred since it does not contain multiple parameters that can be determined by a trial-error method such as learning rate and momentum ratio as other feed-forward backward propagation networks. Furthermore, it converges faster compared to other algorithms and gives more reliable results.

The data is presented to the network by scaling between 0 and 1 with the relation $\frac{x-x_{max}}{x_{max}-x_{min}}$ due to the properties of the sigmoid function. Hence, by eliminating the distances between the data, it is provided that the network convergences faster in the training stage.

In the training of the network, the number of cells in the hidden layer, the initial Marquardt ($\mu_o$)parameter, and the iteration number of the network have been determined by trial and error so that the total squared error is minimum. In addition, in this study, an optimum number of iterations has been used in the training phase in order not to lose the generalization ability of the models, that is, to prevent the test period performance from deteriorating due to overfitting. The ANN model here has four inputs, fifteen cell hidden layers, and an output layer.

To compare the MLR and ANN models, in establishing the multiple regression model, all variables have been included in the model by the stepwise method and the ANN model has been established by using the variables, which reaches maximum specificity coefficient.

After the models are established, their performances are compared by different statistical criteria. Based on these criteria, the best model is determined. In the literature, the most used comparison criterion in the comparison of the regression and ANN models is the values that give the relationship between the estimated values and the actual data. These are the values of mean absolute error (MAE), mean squared error (MSE), root-mean-squared error (RMSE) and the coefficient of determination ($R^2$). It is preferred that the MAE, MSE and RMSE values are close to zero and $R^2$ is close to one. As these values are close to the desired values, the prediction converges to the real value accurately. The mentioned values are calculated by the following formulae [12,14]:

$$MSE = \frac{1}{N}\sum\left(Y - \hat{Y}\right)^2, \tag{2}$$

$$R\,MSE = \sqrt{\frac{1}{N}\sum\left(Y - \hat{Y}\right)^2}, \tag{3}$$

$$R^2 = \frac{\sum\left(\hat{Y} - \bar{Y}\right)^2}{\sum(Y - \bar{Y})^2}, \tag{4}$$

$$MAE = \frac{1}{N}\sum\left|Y - \hat{Y}\right|^2, \tag{5}$$

where, $Y$, $\hat{Y}$ and $N$ represent the observed value, the prediction value and the number of data, respectively.

## RESULTS AND DISCUSSION

The analysis was first started by the regression. Here, the number of deaths and other variables have been considered as dependent and independent variables, respectively, and then models have been run within this framework. In this study, as indicated before, the models have been established by the stepwise method. The models have been established with the variables obtained by the min-max method in order to get rid of the bias of the scales. The $R^2$ MSE, RMSE and MAE values of the models established with these variables are given in Table 3. Since one variable is incorporated into the model at each stage, four models have been established by the regression analysis. When the stepwise method is not preferred, fifteen various models need to be established. Hence, by means of this method, both the models are seen to be significant and the variables enter the model one by one. Besides, the models with the highest explanatoriness have been determined. The summary of the model results is presented in Table 3. Notice

**Table 3.** Model summary

| Model | Variables | R-square | MSE | RMSE | MAE |
|---|---|---|---|---|---|
| 1 | DNC | 0.81 | 0.0096515 | 0.098242 | 0.06211 |
| 2 | DNC, AC | 0.91 | 0.0039074 | 0.062509 | 0.048978 |
| 3 | DNC, AC, AIN | 0.94 | 0.0034675 | 0.058886 | 0.046416 |
| 4 | DNC, AC, AIN, DTC | 0.96 | 0.002642 | 0.0514 | 0.038823 |
| 5 ANN | DNC, AC, DTC, AIN | 0.99 | 0.0010478 | 0.032374 | 0.022440 |

**Table 4.** Model summary with raw data

| Model | Variables | R-square | MSE | RMSE | MAE |
|---|---|---|---|---|---|
| 1 | DNC | 0.806 | 187.2682 | 13.6846 | 10.4043 |
| 2 | DNC, AC | 0.911 | 87.9789 | 9.37971 | 7.4599 |
| 3 | DNC, AC, AIN | 0.936 | 60.4447 | 7.77479 | 5.9308 |
| 4 | DNC, AC, AIN, DTC | 0.958 | 41.7284 | 6.45975 | 5.0467 |
| 5 ANN | DNC, AC, DTC, AIN | 0.999 | 13.98 | 3.73 | 2.6907 |

that the results of the ANN models have been added to the last row of the table.

It is seen from the table that the ANN model is the best model according to the comparison values of the models. Because, the $R^2$ is close to one and the other values are close to zero. While the model is established, it is not exactly understood how small the error since the values that are normalized to the interval [0,1] has been used. When the models are established with raw data, it is better seen how small the error values of the ANN model are compared to the other models (Table 4).The mean squared error (MSE) of the first model is approximately 13 times greater than that of the ANN model, the MSE of the second model is approximately 6 times greater than that of the ANN model, the MSE of the third model is approximately 4 times greater than that of the ANN model, and finally the MSE of the fourth model is approximately 3 times greater than that of the ANN model.This means that the DND predicted by the ANN model produces far less error. The superiority of the ANN model over the competitors is also clearly seen from the RMSE and MAE values.

It is seen that all of the models with $R^2$, MSE, RMSE and MAE values above are statistically significant. The regression coefficients created by the raw data, the normalized coefficients, *t* values and the significance values of these models are given in Table 5.

Considering the last model in Table 5, when the value of each of the DNC, AC and AIN variables is increased by 10000, it is seen that 83 people, 7 people and 480 people died, respectively. As the number of DTC increases, the number of deaths decreases. This situation can be explained as follows: As the number of tests increases, the number of patients is determined accurately and since these people are taken under control by the contact tracing team, infecting others is prevented. According to the standardized beta coefficient, while the variable AIN is the most effective variable, the AC follows it.

As will be noted, the relationship between the dependent variable and the independent variables has been found to be high except for the DTC variable, as seen in Table 5. Since there is no linear relationship between variables, the relationship between the dependent variable and the DTC is weak. At the same time, considering the regression models, this variable is seen to have a negative effect on the dependent variable as noticed from the correlation. The correlation table and the graph show that there is another nonlinear relationship between the two variables. Since the other variables have high correlations and a linear relationship with the dependent variable, as can be seen in the figure, the MLR can be established by these variables and the dependent variable. Besides, as seen in the figure and the correlation table, there is a high relation between the variables, except for the DTC. For these and similar reasons, even if the regression results have a high prediction capacity, they may be expected to cause some difficulties in making accurate estimations. Higher or lower than expected erroneous estimations can be made especially in the extreme values, that are not present in the data, of the variables. Hence, these situations cause the assumptions of the regression model to be broken and the prediction of the regression analysis to be approached with suspicion. In

**Table 5.** Coefficients

| | **Unstandardized Coefficients** | | **Standardized Coefficient** | | |
|---|---|---|---|---|---|
| **Model** | **B** | **Error** | **Beta** | **t** | **Sig.** |
| 1 (Constant) | -5.4351 | 1.949 | | -2.788 | 0.006 |
| DNC | 0.0277 | 0.001 | 0.895 | 26.966 | 0.000 |
| 2 (Constant) | -9.2961 | 1.351 | | -6.882 | 0.000 |
| DNC | 0.0179 | 0.001 | 0.580 | 18.498 | 0.000 |
| AC | 0.0007 | 0.000 | 0.455 | 14.525 | 0.000 |
| 3 (Constant) | -14.7156 | 1.301 | | -11.312 | 0.000 |
| DNC | 0.0135 | 0.001 | 0.436 | 13 .973 | 0.000 |
| AC | 0.0007 | 0.000 | 0.455 | 17.222 | 0.000 |
| AIN | 0.0212 | 0.002 | 0.218 | 8.606 | 0.000 |
| 4 (Constant) | -6.5004 | 1.423 | | -4.568 | 0.000 |
| DNC | 0.0083 | 0.001 | 0.268 | 8.362 | 0.000 |
| AC | 0.0007 | 0.000 | 0.409 | 18.102 | 0.000 |
| AIN | 0.0480 | 0.004 | 0.494 | 13.175 | 0.000 |
| DTC | -0.0003 | 0.000 | -0.254 | 8.911 | 0.000 |

a. Dependent variable: DND

**Table 6.** Model summary with raw data

| | | DNC | DNC, AC | DNC, AC, DTC | DNC, AC, DTC, AIN |
|---|---|---|---|---|---|
| Original | RMSE | 13.6846 | 9.3797 | 9.3096 | 6.5395 |
| | MAE | 10.4043 | 7.4599 | 7.5152 | 5.2128 |
| | R | 0.9002 | 0.9544 | 0.9551 | 0.9781 |
| Lag 1 | RMSE | 12.9589 | 9.8579 | 9.7777 | 6.9051 |
| | MAE | 9.9803 | 7.5689 | 7.5983 | 5.4225 |
| | R | 0.9091 | 0.9485 | 0.9493 | 0.9751 |
| Lag 3 | RMSE | 11.8834 | 10.9075 | 10.9070 | 8.1014 |
| | MAE | 9.5281 | 8.5696 | 8.5715 | 6.3477 |
| | R | 0.9238 | 0.9362 | 0.9362 | 0.9653 |
| Lag 5 | RMSE | 11.4706 | 11.4478 | 11.3936 | 9.2975 |
| | MAE | 9.5712 | 9.4621 | 9.4198 | 7.3293 |
| | R | 0.9292 | 0.9294 | 0.9301 | 0.9541 |
| Lag 7 | RMSE | 13.4385 | 13.1321 | 12.8899 | 11.2015 |
| | MAE | 10.4757 | 10.6916 | 10.5691 | 8.5087 |
| | R | 0.9012 | 0.9059 | 0.9095 | 0.9325 |
| Lag 10 | RMSE | 16.6834 | 16.6331 | 15.8598 | 13.9642 |
| | MAE | 12.9933 | 13.0220 | 12.8759 | 10.6869 |
| | R | 0.8110 | 0.8445 | 0.8598 | 0.8932 |
| Lag 12 | RMSE | 21.2968 | 18.9044 | 17.5715 | 15.5298 |
| | MAE | 16.1722 | 15.0353 | 14.8178 | 12.3495 |
| | R | 0.7293 | 0.7944 | 0.8254 | 0.8666 |
| Lag 15 | RMSE | 25.2913 | 21.9237 | 19.6170 | 17.5939 |
| | MAE | 19.2869 | 17.1932 | 16.6672 | 14.3336 |
| | R | 0.5916 | 0.7152 | 0.7803 | 0.82789 |

such cases, using ANNs that are independent of assumptions can give more accurate results.

The multiple time series analysis results are shown in Table 6. While performing this analysis, the models have been created by preserving the regression structure. The model established as lag free, the results produced by this model for all variables are for $R = 0.97$. The rest of the models have had lower results.

When the ANN results are analyzed, it is seen that $R^2$ values are 0.99 for the train, test, and validation results. Hence, it is understood that the ANN model does not memorize the data, on the contrary, it trains the data. The values of MSE, RMSE, MAE of the ANN model are presented in Tables 3 and 4. Since the obtained $R^2$ values in the multiple regresssion models are higher, the ANN models are the best representation of the data and these models make more accurate predictions. The stages of the ANN analysis results are presented in Figures 4 and 5.

When the results are examined, it is remarked that the ANN model results are more successful than that of the MLR models. In this work, the ANN model has been established with the variables in the fourth model that has the most explanatory regression models. It is used a single hidden layer and between 10 and 20 cells. Since there is not significant variation in the result of analysis, in this paper, the model results with one hidden layer and fifteen cells has preferred.

In the application of artificial neural network, the Levenberg-Marquardt algorithm that is one of the feed-forward back-propagation network algorithm has been used. This algorithm is preferred since it is not included more than one parameter that is determined trial and error method, it converges faster than other algorithms and gives reliable results. The correlation and the scatter plots between the variables have been utilized to decide the inputs of the MLR and ANN. The variables that are entered to the model have been determined by the stepwise method and the variables have been entered into the model one by one. The reason for this is to find the variable and the variables that explain the model in the best way and to see how to change the explanatoriness of the model by adding a variable to it. The ANN analysis has been performed with the variables in the most explanatory model obtained from the regression. While the fourth model is the best model among the regression models, it has been concluded that the ANN model is the best among all models as it has no statistical

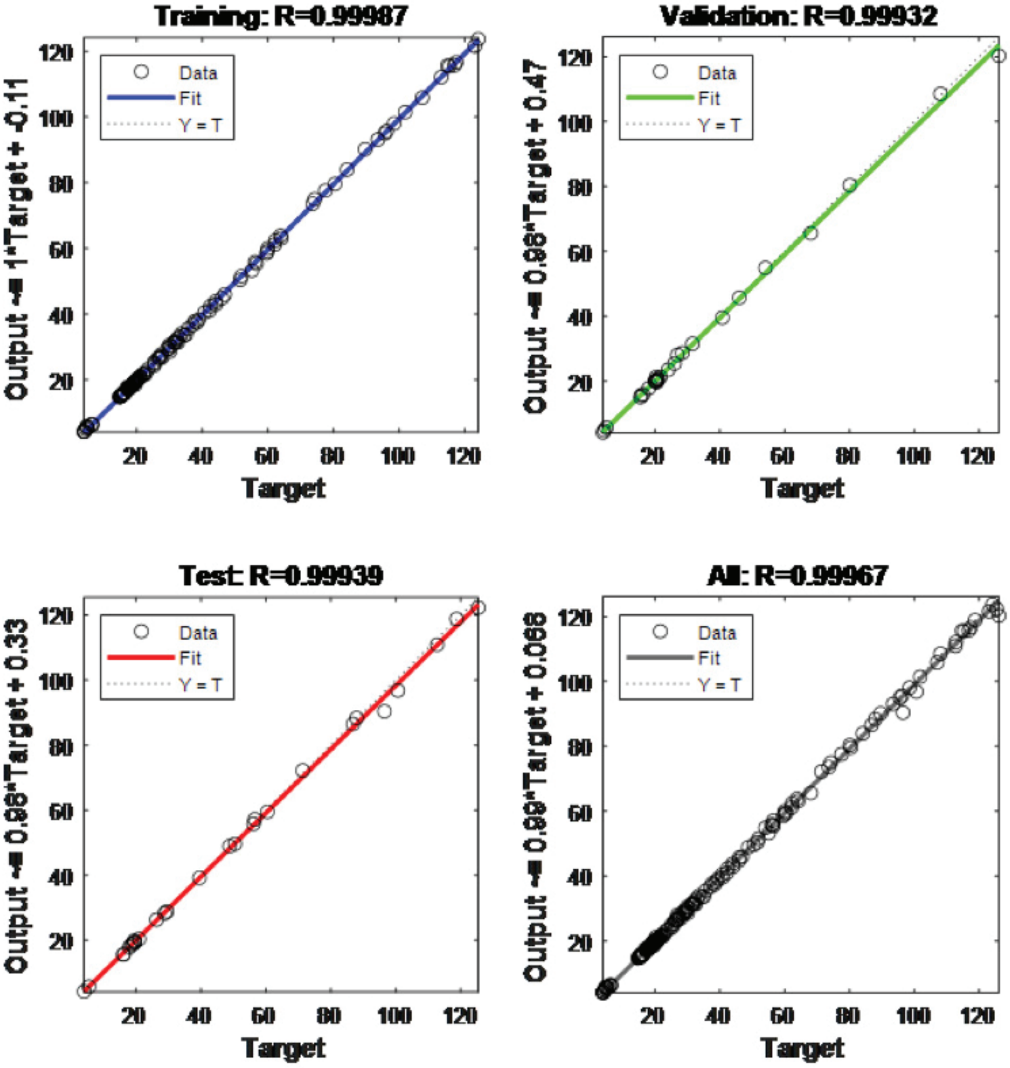

**Figure 4.** The ANN analysis results

assumptions and has an explanatory value of 0.99\%. It has been observed that the ANN which is easily used in the linear and nonlinear models is superior to other models since it does not have a statistical assumption and has %99.9 explanatoriness. In addition, even if the number of deaths is not declared, the number of deaths can be estimated with an accuracy of %99.9 with the ANN model. Consequently, the ANN model has been seen to be more successful to predict the daily number of deaths according to the reported COVID-19 data. The ANN model presents a very good result for establishing a prediction model in cases where the independent variables have multiple connections and there is no linear relationship between the independent variable and the dependent variable in multiple regression analysis. When the results are compared, it has been observed that the ANN model produces more accurate solutions than of the regression models.

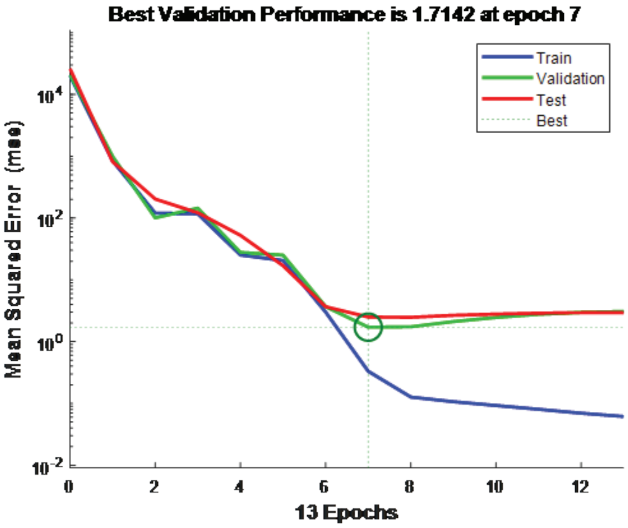

**Figure 5.** The ANN analysis results.

In addition, SIR and SEIR models could also have been used to construct a model among the data of interest. However, the high explanatory power of the currently obtained model does not necessitate the use of those approaches.

The ANN that is an alternative to the MLR analysis has been applied in various fields and gives better results than of the MLR analysis [48-51]. Studies in the literature have shown that ANN overcomes the constraints of the MLR models [49,52-54]. One of the most important assumptions of the MLR models is that the relationship between the dependent variable and the independent variables is linear. On the other hand, in many studies, this assumption can not be provided. However, the ANN does not make any assumptions about the relationship between the variables, so both linear and nonlinear relationships can be analyzed. Furthermore, this method does not affect from multicollinearity. Besides, the ANN provides consistent results even if it works with missing data, does not make assumptions or works with data that does not conform to normal distribution [51].

## CONCLUSION

- In the present study, the relationship between the daily number of case (DNC), active case (AC), daytime testing count (DTC), daily number of deaths (DND) and active incubated patient number (AIN) and the effect of other variables on the DND have been modelled by using the MLR and ANN approaches.
- Effective estimation of COVID-19 deaths based on various parameters has been made comparatively for the first time in this study using ANN and multiple regression algorithms.
- The predicted death numbers have been found to be in good agreement with the actual data for both models proposed here.
- Even though the regression algorithms seem to have very explanatory variables to predict the number of deaths in the model, it has been concluded that ANN produces a model with the highest explanatoriness.

Hence, it is concluded that the ANN model has been seen to be more successful to predict the daily number of deaths according to the reported COVID-19 data.

This study can be expanded to estimate COVID-19 in a bit more realistic manner by incorporating additional characteristics such as the number of cases vaccinated. Therefore, models containing vaccination data may be one step closer to real-world relevance in forecasting COVID-19 spread behaviour.

## AUTHORSHIP CONTRIBUTIONS

The authors equally contributed to this work.

## CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## ETHICS

There are no ethical issues with the publication of this manuscript.

## REFERENCES

[1] World Health Organization. Coronavirus disease (COVID-2019) situation reports. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed on May 7, 2024.

[2] Wu JT, Leung K, Leung GM. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. Lancet 2020;395:689−697. [CrossRef]

[3] Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, et al. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med 2020;26(6):855−860. [CrossRef]

[4] Maier BF, Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science 2020;368:742−746. [CrossRef]

[5] Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 2020;15:e0230405. [CrossRef]

[6] Turkyilmazoglu M. Explicit formulae for the peak time of an epidemic from the SIR model. Physica D 2021;422:132902. [CrossRef]

[7] Turkyilmazoglu M. An extended epidemic model with vaccination: Weak-immune SIRVI. Physica A 2022;598:127429. [CrossRef]

[8] Turkyilmazoglu M. A restricted epidemic SIR model with elementary solutions. Phsysica A 2022;600:127570. [CrossRef]

[9] Tunc H, Sari M, Kotil SE. Effect of sojourn time distributions on the early dynamics of COVID-19 outbreak. Nonlinear Dyn 2023;111:11685−11702. [CrossRef]

[10] Saba AI, Elsheikh AH. Forecasting the prevalence of COVID-19 outbreak in Egypt using nonlinear autoregressive artificial neural networks. Process Saf Environ Prot 2020;141:1−8. [CrossRef]

[11] Ardabili SF, Mosavi A, Ghamisi P, Ferdinand F, Varkonyi-Koczy AR, Reuter U, et al. COVID-19 outbreak prediction with machine learning. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3580188. Accessed on May 7, 2024.

[12] Benvenuto D, Giovanetti M, Vassallo L, Angeletti S, Ciccozzi M. Application of the ARIMA model on the COVID-2019 epidemic dataset. Data Brief 2020;29:105340. [CrossRef]

[13] Al-Qaness MAA, Ewees AA, Fan H, Abd El Aziz M. Optimization method for forecasting confirmed cases of COVID-19 in China. J Clin Med 2020;9:674. [CrossRef]

[14] Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792. [CrossRef]

[15] Ceylan Z. Estimation of COVID-19 prevalence in Italy, Spain, and France. Sci Total Environ 2020;729:138817. [CrossRef]

[16] Bapir SY, Kareem SM. COVID-19 and functionality: By providing social distancing of indoor common spaces in residental building. J Stud Sci Eng 2021;1:36–45. [CrossRef]

[17] Xue H, Bai Y, Hu H, Liang H. Influenza activity surveillance based on multiple regression model and artificial neural network. IEEE Access 2017;6:563–575. [CrossRef]

[18] Liu X, Jiang B, Gu W, Liu Q. Temporal trend and climate factors of hemorrhagic fever with renal syndrome epidemic in Shenyang City, China. BMC Infect Dis 2011;11:331. [CrossRef]

[19] Balkhy HH, Abolfotouh MA, Al-Hathlool RH, Al-Jumah MA. Awareness, attitudes, and practices related to the swine influenza pandemic among the Saudi public. BMC Infect Dis 2010;10:42. [CrossRef]

[20] Ahmad S, Mehfuz S, Mebarek-Oudina F, Beg J. RSM analysis based cloud access security broker: A systematic literature review. Cluster Comput 2022;25:3733–3763. [CrossRef]

[21] Guan P, Huang DS, Zhou BS. Forecasting model for the incidence of hepatitis A based on artificial neural network. World J Gastroenterol 2004;10:3579–3582. [CrossRef]

[22] Faisal T, Taib MN, Ibrahim F. Neural network diagnostic system for dengue patients risk classification. J Med Syst 2012;36:661-676. [CrossRef]

[23] Elveren E, Yumuşak N. Tuberculosis disease diagnosis using artificial neural network trained with genetic algorithm. J Med Syst 2011;35:329–332. [CrossRef]

[24] Baridam B, Irozuru C. The prediction of prevalence and spread of HIV/AIDS using artificial neural network-the case of rivers State in the Niger Delta, Nigeria. Int J Comput Appl 2012;44:42-45. [CrossRef]

[25] Aburas HM, Cetiner BG, Sari M. Dengue confirmed-cases prediction: A neural network model. Expert Syst Appl 2010;37:4256–4260. [CrossRef]

[26] Koliopoulos TK, Papakonstantinou D, Ciarkowska K, Antonkiewicz J, Gambus F, Mebarek-Oudina F, et al. Green Designs in Hydraulics - Construction Infrastructures for Safe Agricultural Tourism and Sustainable Sports Tourism Facilities Mitigating Risks of Tourism in Crisis at Post COVID-19 Era. In: de Carvalho JV, Liberato P, Peña A, editors. Advances in Tourism, Technology and Systems. Smart Innovation, Systems and Technologies. 2nd ed. New York: Springer; 2022. p. 37–47. [CrossRef]

[27] Chaurasia V, Pal S. COVID-19 pandemic: ARIMA and regression model-based worldwide death cases predictions. SN Comput Sci 2020;1:288. [CrossRef]

[28] Amar LA, Taha AA, Mohamed MY. Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt. Infect Dis Model 2020;5:622–634. [CrossRef]

[29] Niazkar HR, Niazkar M. Application of artificial neural networks to predict the COVID-19 outbreak. Glob Health Res Policy 2020;5:50. [CrossRef]

[30] Melin P, Monica JC, Sanchez D, Castillo O. Multiple ensemble neural network models with fuzzy response aggregation for predicting COVID-19 time series: The case of mexico. Healthcare (Basel) 2020;8:181. [CrossRef]

[31] Tamang SK, Singh PD, Datta B. Forecasting of Covid-19 cases based on prediction using artificial neural network curve fitting technique. Glob J Environ Sci Manag 2020;6:53–64.

[32] Moftakhar L, Seif M, Safe MS. Exponentially increasing trend of infected patients with COVID-19 in Iran: A comparison of neural network and arima forecasting models. Iran J Public Health 2020;49:92–100. [CrossRef]

[33] Nakip M, Copur O, Guzelis C. Comparative study of forecasting models for COVID-19 outbreak in Turkey. Available at: https://www.iitis.pl/sites/default/files/pubs/Covid_19_Forecasting%20%281%29.pdf. Accessed on May 7, 2024.

[34] Toğa G, Atalay B, Toksari MD. COVID-19 prevalence forecasting using Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Networks (ANN): Case of Turkey. J Infect Public Health 2021;14:811–816. [CrossRef]

[35] Guleryuz D. Forecasting outbreak of COVID-19 in Turkey; Comparison of Box-Jenkins, Brown's exponential smoothing and long short-term memory models. Process Saf Environ Prot 2021;149:927–935. [CrossRef]

[36] Demir E, Canitez MN, Elazab M, Hameed AA, Jamil A, Al-Dulaimi AA. Assessing the spreading behavior of the Covid-19 epidemic: A case study of Turkey. Available at: https://acikerisim.istinye.edu.tr/xmlui/handle/20.500.12713/3242. Accessed on May 7, 2024.

[37] Kuvvetli Y, Deveci M, Paksoy, T, Garg H. A predictive analytics model for COVID-19 pandemic using artificial neural networks. Decis Anal J 2021;1:100007. [CrossRef]

[38] Caglar O, Ozen F. A comparison of Covid-19 cases and deaths in Turkey and in other countries. Netw Model Anal Health Inform Bioinform 2022;11:45. [CrossRef]

[39] Kırbaş İ, Sözen A, Tuncer AD, Kazancıoğlu FŞ. Comparative analysis and forecasting of COVID-19 cases in various European countries with ARIMA, NARNN and LSTM approaches. Chaos Solitons Fractals 2020;138:110015. [CrossRef]

[40] Weisberg S. Applied Linear Regression. 3th ed. New Jersey: John Wiley & Sons, Inc; 2005. [CrossRef]

[41] Fox J. Applied Regression Analysis: Linear Models and Related Methods. California: SagePublication; 1997.

[42] Skapura DM. Building Neural Networks. 1st ed. Boston: Addison-Wesley; 1995.

[43] Haykin S. Neural Networks: A Comprehensive Foundation. New York: Macmillian College Publishing Company Inc; 1994.

[44] Chaudhuri BB, Bhattacharya U. Efficient training and improved performance of multilayer perceptron in pattern classification. Neurocomput 2000;34:11–27. [CrossRef]

[45] PRT-DTO, 2020. Presidency of The Republic of Turkey Digital Transformation Office. Available at: https://cbddo.gov.tr/ Accessed on May 7, 2024.

[46] T.C Sağlık Bakanlığı - Covid-19 Bilgilendirme Platformu. Günlük Covid-19 aşı tablosu. Available at: https://covid19.saglik.gov.tr/. Accessed on May 7, 2024.

[47] Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: The state of the art. Int J Forecast 1998;14:35–62. [CrossRef]

[48] Spangler WE, May JH, Vargas LG. Choosing datamining for multiple classification: Representational and performance measurement implications for decision support. J Manag Inf Syst 1999;16:37–62. [CrossRef]

[49] Uysal M, Roubi SE. Artificial neural networks versus multiple regression in tourism demand analysis. J Travel Res 1999;38:111–118. [CrossRef]

[50] Fadlalla A, Lin CH. An analysis of the applications of neural networks in finance. Interfaces 2001;31:112–122. [CrossRef]

[51] Nguyen N, Cripps A. Predicting housing value: a comparison of multiple regression analysis and artificial neural networks. J Real Estate Res 2001;22:313–336. [CrossRef]

[52] Gorr WL. Research prospective on neural network forecasting. Int J Forecast 1994;10:1–4. [CrossRef]

[53] Hill T, Remus W. Neural network approach for intelligent support of managerial decision making. Decis Support Syst 1994;11:449–459. [CrossRef]

[54] Wray B, Palmer A, Bejou D. Using neural network analysis to evaluate buyer-seller relationship. Eur J Market 1994;28:32–48. [CrossRef]