Prediction of Citrus Fruits Production using Artificial Neural Networks and Linear Regression Analysis

Elifcan GÖÇMEN[1*] , Yusuf KUVVETLİ[2]

[1] Department of Industrial Engineering, Faculty of Engineering, Munzur University, Tunceli/TURKEY

[2] Department of Industrial Engineering, Faculty of Engineering, Çukurova University, Adana/TURKEY

Abstract

Accurate and timely prediction of fruits production plays a significant role in the agriculture industry. Therefore, it is very important to predict of citrus fruits production. In this study, prediction the production amount of different citrus fruits for a city of Turkey (Adana) is aimed. Orange, mandarin and bitter orange are included as citrus products and the production amounts of ten years are used as dataset. Artificial neural network (ANN) and linear regression analysis are performed for predicting the production amounts. A feed forward neural network is proposed with regarding some inputs such as districts of Adana, product types, product specific plant area, average yield per tree, number of fruitless trees, number of fruit trees, total number of trees, population, inflation rate, total fruit area, temperature, average rainfall. The obtained results in which the R^2 values are greater than 0.98 for all datasets show us that the proposed method can predict the production amount accurately regarding the input parameters.

Keywords: Feed forward neural networks, regression analysis, agricultural prediction, citrus fruits production

Yapay Sinir Ağı ve Regresyon Analizi Kullanarak Narenciye Üretim Miktarı Tahmini

Öz

Meyve üretim miktarlarının doğru ve zamanında tahmini, tarım sektöründe önemli rol oynamaktadır. Bu yüzden, narenciye üretimini tahmin etmek çok önemlidir. Bu çalışmada, Türkiye'de Adana ilinde üretilen farklı narenciyelerin üretim miktarlarının tahmin edilmesi amaçlanmaktadır. Narenciye ürünleri olarak portakal, mandalina ve turunç seçilmiştir ve 10 yılın üretim miktarları veri seti olarak kullanılmıştır. Üretim miktarlarının tahmini için yapay sinir ağları ve lineer regresyon analizi uygulanmıştır. İlçeler, ürün tipleri, ürün dikim alanı, ağaç başına ortalama verim, meyve vermeyen ağaç sayısı, meyve veren ağaç sayısı, toplam ağaç sayısı, nüfus, enflasyon oranı, toplam meyve alanı, sıcaklık ve ortalama yağışları girdiler olarak göz önüne alan ileri beslemeli sinir ağı önerilmiştir. R^2 değerinin tüm veriler için 0, 98' den büyük olduğu elde edilen sonuçlar, girdi değerlerini göz önünde bulundurarak önerilen yöntemin üretim miktarını doğruca tahmin edebildiğini göstermektedir.

Anahtar Kelimeler: İleri Beslemeli Sinir Ağı, regresyon analizi, tarımsal tahmin, narenciye üretimi

---

[*] Corresponding Author: E.Gocmen, elifcangocmen@munzur.edu.tr

1. Introduction

Uncertainty of production amounts forces the agriculture firms to predict the amounts accurately. Production amounts are also an important indicator for assessing the agricultural plans. Forecasting the production amounts is critically important based on macroeconomic efficiency. By accurate predictions, the goods produced in the region can meet the demands. Prices of these goods may not increase due to satisfactory production. Thus, consumers do not need the import goods.

Artificial neural networks and regression methods have gained considerable attraction in all fields of science especially in medicine, economics and engineering. Artificial neural network is a processing system that leads to gain skills about analysing, synthesizing and the network is based on the working principles of the human brain imitating the mechanism of the nerves in the brain (Grossberg, 1988).

In this study, we use a feed-forward neural network model and linear regression analysis for prediction of the production amounts based on the twelve inputs including districts of Adana, product types, product specific plant area, average yield per tree, number of fruitless trees, number of fruit trees, total number of trees, population, inflation rate, total fruit area, temperature, average rainfall.

Although there are a lot of studies for forecasting with artificial neural network and regression models in the health, economics, production sectors, there are few studies in the agriculture sector for forecasting with various parameters. Zou et al. (2007) studied forecasting the wheat price with artificial neural network, Autoregressive integrated moving average (ARIMA) and time series models. They

compared the three methods for the food grain price. The study showed that ANN is better than the other methods for turning point and profit. However, a combined method including ANN and ARIMA were more effective for forecasting performance in errors. Elizondo et al. (1994) used artificial neural network to forecast the times of germination of soybean and ripening of them by minimizing the errors. Tamari et al. (1996) proposed a layered network to predict the soil water permeability. The results of the artificial neural network were compared with the linear regression method. Kohzadi et al. (1996) examined the performances of ANN and ARIMA model for forecasting cattle and wheat prices. ANN forecast gave lower error measures than the ARIMA in this study. Parmar (1997) developed a neural network including 4 input data and 8 hidden layers to examine the earth contaminated with alpha toxin. Yarar (2004) discussed the water level changes in Beysehir Lake by using the neural network method. The inputs were precipitation, evaporation, level measurement values. These level values were also estimated with artificial neural network. Kaul et al. (2005) compare the regression and ANN results to predict the corn and soybean yields. ANN produced more effective prediction than the regression related with the adjusting ANN parameters. Bayraktar (2006) determined the factors affecting the shelf life and the observations about the factors are predicted with artificial neural network. In the network, sigmoid activation functions at layers and feedback network at training were used. The output parameter was shelf life. The neural network performance gave better results. Movagharnejad and Nikzad (2007) used both of the artificial neural network and

empirical mathematical equations for modelling of the data obtained from the experimental studies including air flow rate, temperature etc. values of drying the tomatoes. Ji et al. (2007) compared ANN and regression models to predict rice yield for typical climatic conditions. ANN produced more effective prediction than the regression related with the adjusting ANN parameters. Çakır et al. (2014) used ANN and regression method to predict the wheat yield. Multi-Layer Perceptron (MLP) neural network model is conducted. Results showed that MLP is better than the regression method. Matsumura et al. (2015) conducted a forecasting problem for the maize yield production related with climate conditions and fertilizer using linear regression and non-linear ANN models. ANN gives better results than the regression in this study. Identification of citrus fruits using ANN (Fiona et al., 2019), prediction of kiwifruit using ANN and multiple linear regressions (Torkashvand et al., 2017), identification a red dragon fruit based by back propagation approach (Prasetyo, and Bimantaka, 2018) are conducted in the literature. Abrougui et al. (2019) addressed the prediction of potato crop yield using ANN and multiple linear regression methods. Results demonstrated that regression method gives better performance to predict, however it gives lower effective performance than ANN. Boukelia et al. (2020) developed ANN and regression models to analyse cooling performance of the different solar power plants. ANN gives more accurate results than the regression models for predicting hourly cooling performances of the plants. Tušek et al. (2020) used multivariate regression and ANN to predict of aqueous extracts properties. Results demonstrated that regression models could

be used to predict, however, ANN gave more accurate results to predict both properties tan the regression models. Hosseinzadeh et al. (2020) used ANN and multiple linear regression models to predict the nutrient recovery of solid waste. Three layered ANN model were used to choose the effective prediction model. ANN is better than the regression model based on the statistical analysis.

The aim of the paper is to provide that the proposed method can predict the production amount accurately based on the input parameters and to obtain high R^2 values to show the power of our approach.

2. Material and Methods

In this paper, a feed-forward neural network methodology is applied to determine the citrus fruits production amount regarding different citrus fruit types which are called as orange, mandarin and bitter orange. The dataset is acquired from the different sources. The dataset includes temperature and rain values are obtained from Turkish State Meteorological Service (TSMS). The other inputs are obtained from database of Turkish Statistical Institute (TSI, 2017). The dataset has twelve different input attributes and the output is the production amount (ton). Input attributes are: (i) districts of Adana, (ii) product types, (iii) product specific plant area, (iv) average yield per tree, (v) number of fruitless trees, (vi) number of fruit trees, (vii) total number of trees, (viii) population, (ix) inflation rate, (x) total fruit area, (xi) average temperature, (xii) average rainfall. The inputs are chosen since they are mostly related with the citrus fruits production. We believe that their effects on the production amount is critical. Besides,

availability of the data for these inputs is important to chose these inputs. The statistical analysis of the data set is summarized in Table 1.

Table 1. Statistical analysis of the dataset

Parameter	Type	Minimum value	Maximum value	Mean value	Standard deviation
Product specific plant area	decare	1,03	420	53,2698	92,46823
Production	ton	1,011	850	114,9245	198,692
average yield per tree	kg	35	212	103,8553	34,44747
number of fruit trees	unit	1,1	700	154,7863	194,6534
number of fruitless trees	unit	0	630	36,22446	124,6606
total number of trees	unit	1,1	865	164,787	213,3866
population	person	4840	990073	200869,3	266875,7
inflation rate	%	134,85	319,69	214,4605	58,03387
total orchard area	decare	10,294	218608	17263,34	45844,77
average temperature	degrees	18.0250	20.9333	19.6023	0.588
average rainfall	kg/m^2	20.35	64.83	64.8294	21.2629

Artificial neural networks are widely used for variety of applications such as prediction, machine learning. In this study, a feed forward neural network methodology is implemented to predict the production amount of the three different citrus products in Adana, Turkey. In a typical feed forward neural network approach which is demonstrated in Figure 1 for this study, the network consists of neurons on the different layers and their connections. The first layers include the input neurons and similarly the last layer is comprised of the outputs of the system. During the training phase, each of these layers is connected to directly next layers and contributes only them; therefore, this type of networks is called as feed forward neural networks (Priddy and Keller, 2005).
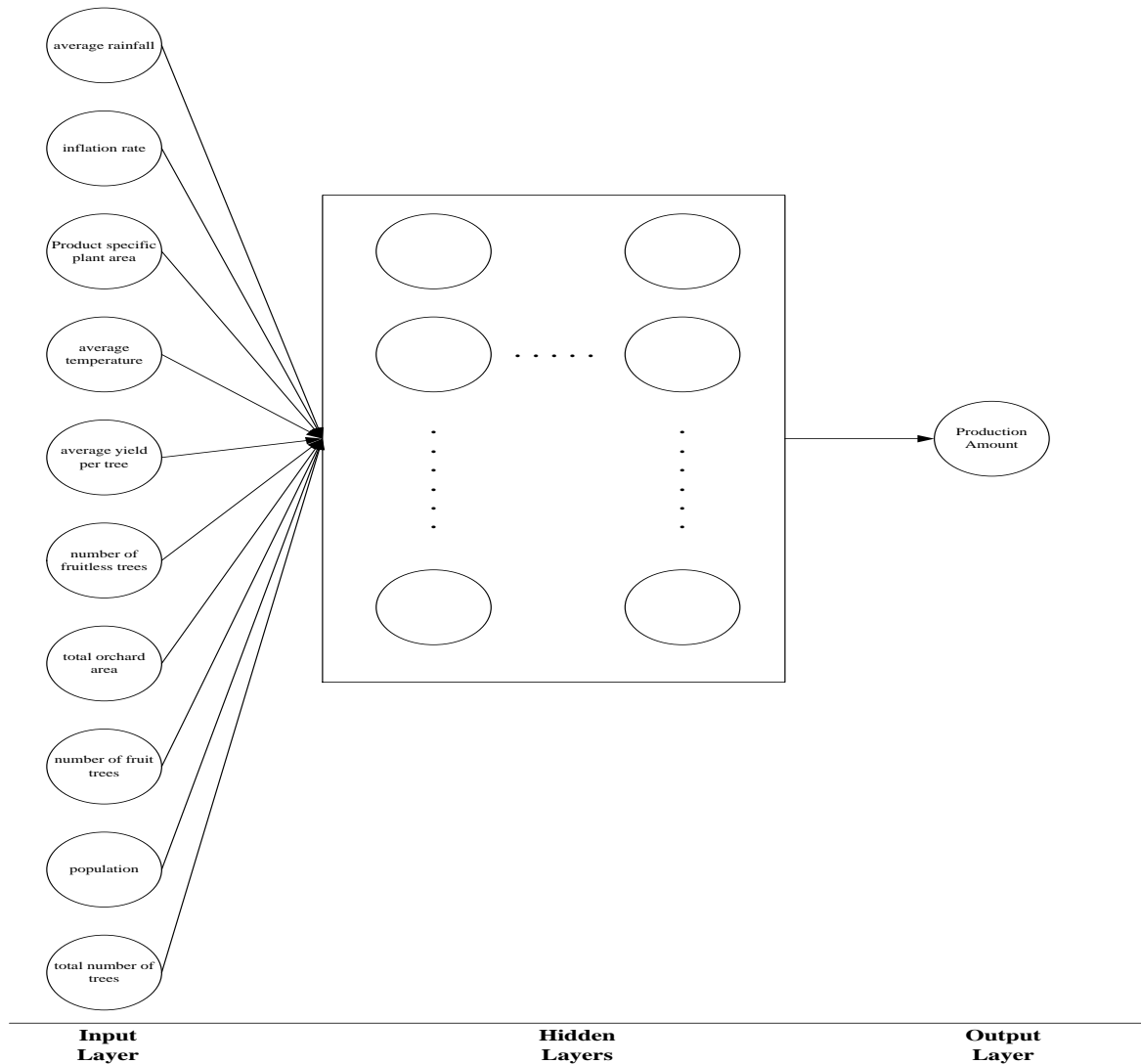
Figure 1. The network architecture proposed by the authors

Learning algorithms are used to train the neural network model. In this study, the Levenberg-Marquardt learning algorithm is applied for training the ANN model (Kisi, 2004). In order to divide the dataset, tenfold cross validation methodology is used for determining the training and test data sets. The MATLAB software package is used for the modelling the ANN.

The number of hidden neurons can be influenced the prediction accuracy of the neural network model. For this reason, different numbers of hidden neurons are evaluated in order to find the optimal prediction methodology. Figure 2 shows the results of using different number of hidden neurons on the hidden layer of the neural network model. The best performance is acquired when the number of hidden neuron is equal to 6. Therefore, it is set to 6. In addition, the hidden layer number is 1 for ANN and 1000 epoch is set with the Levenberg-marquadt algorithm.
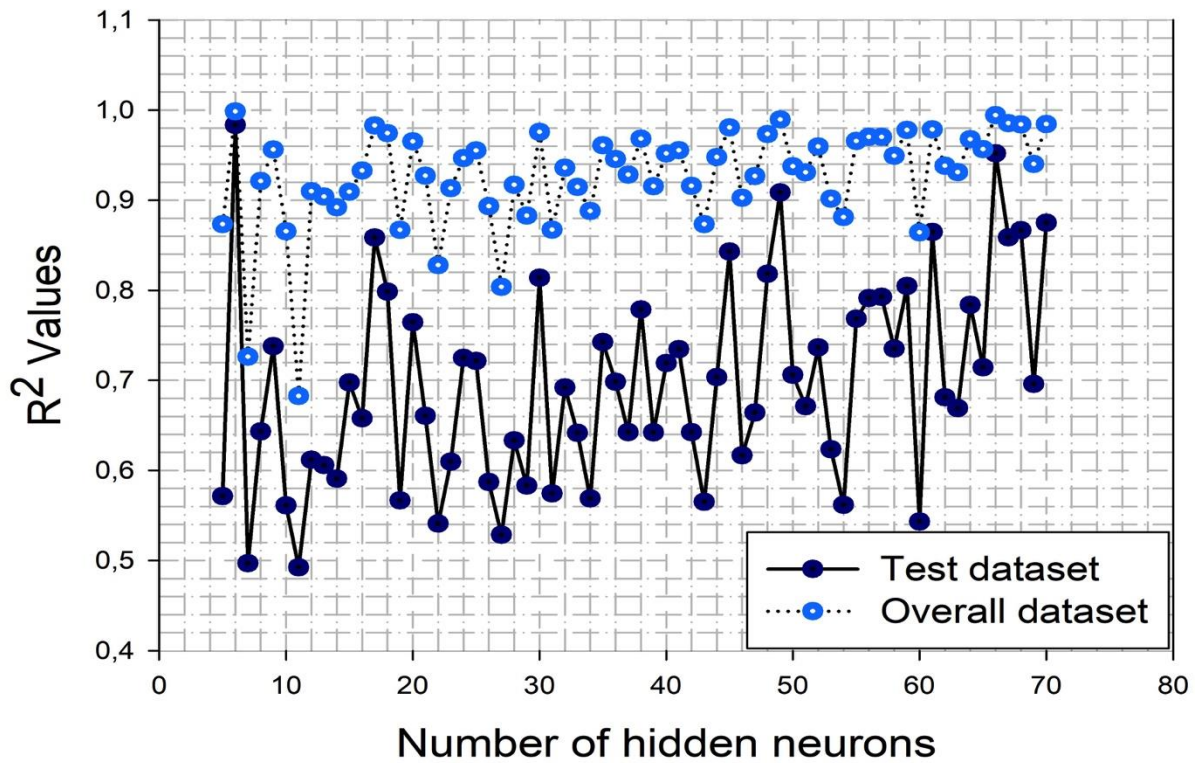
3. Results and Discussion

Figure 2. Determining the best neural network architecture

In order to compare the performance of the ANN, the linear regression model is built for the predicting the citrus production amount. Both linear regression and the ANN model training are done by the same training dataset for a fair comparison. The linear regression results are given in Table 2. The results shows that the proposed linear regression model can be used for this manner (significance value 0.000<0.05). The linear regression equation is given in Equation (1).

Production rate = 158.883 −19.884 * district −5.781* product type + 1.711 * product specific plant area +0.586 * average yield per tree − 0.511 * number of fruit trees − 0.059 * number of fruitless trees + 0.422 * total number of trees − 0.00009* population +0.340 * inflation rate -0.0003 * total fruit area -7.934 * average temperature + 0.194 * average rainfall   (1)

Table 2. Linear regression results of predicting the citrus production amount

Model	Sum of Squares	df	Mean Square	F	Sig.
Regression	4183234	12	348602,8	36,07827	1E-34
Residual	1198138	124	9662,403		
Total	5381372	136			

The results are summarized in Table 3. According to the results, the ANN model has a better convergence to training data set with less root mean squared error (RMSE) error. The same results can be obtained from the test and overall dataset. It is more important to predict test data accurately than training; because, the test data is new for the model and it shows the actual performance of the method. This is obvious that the ANN model has a great potential to predict the test instances. Figure 4 shows the fit of the both ANN and linear regression on each data sample. It seems that ANN predicts more accurately than linear regression predictions.

Table 3. Comparison of the ANN and the linear regression methods based on root mean squared error (RMSE) error

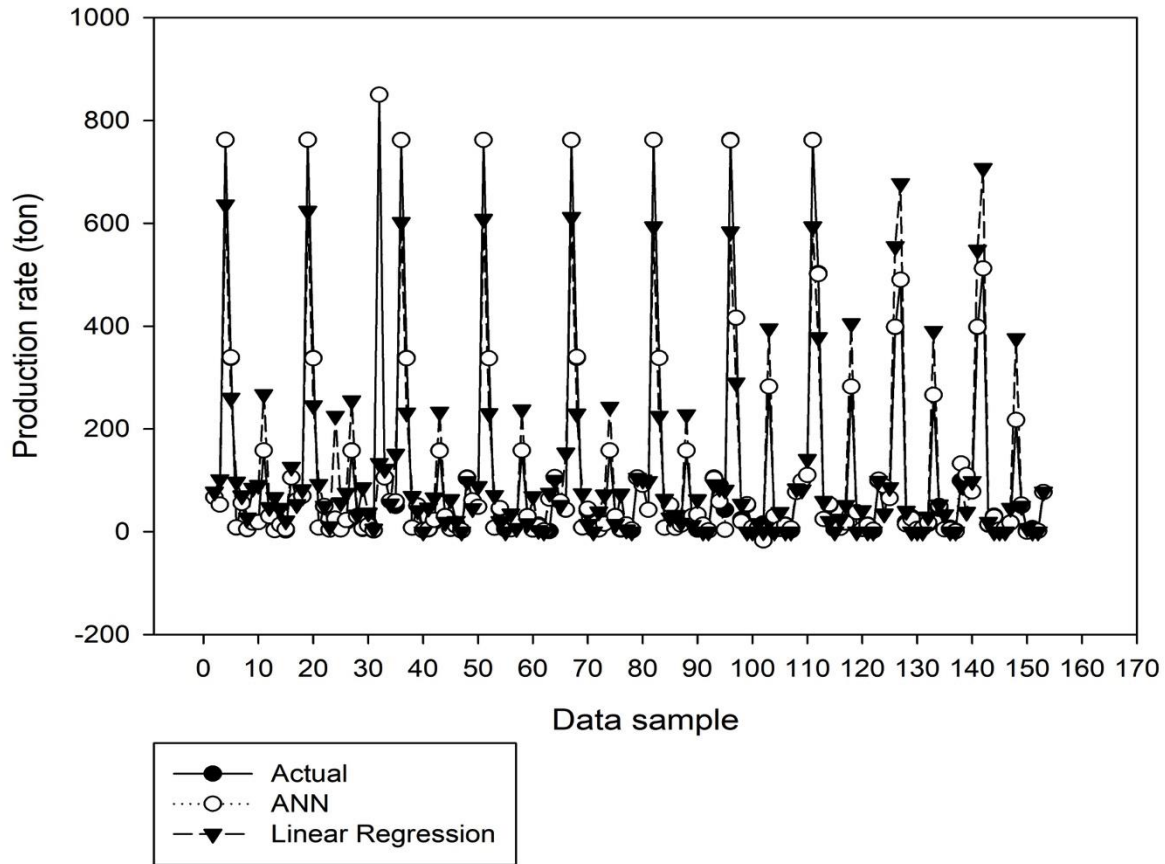| Method	Training Dataset		Test Dataset		Overall	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
The ANN model	0.999	1.5048	0.983	25.42	0.998	8.1123
The Linear regression model	0.776	92.3424	0.8358	65.198	0.779	90.0284

Figure 4. Comparison of the ANN and the linear regression methods based on production rate on each data sample

There are twelve different predictors as aforementioned before. The most effective predictors or ineffective predictors should be determined in order to decide the best prediction model. For this reason, the effects of the input parameters are evaluated by omitting each parameter on the model. Test performances are summarized in Table 4 for omitting each input parameters. In such condition, it can be concluded from the effect of the parameter on the prediction accuracy. According to results, the most effective predictors are product specific plant area and total fruit area. The model's prediction accuracy is affected about %49 percentages regarding the test R^2 values

when the most effective predictors are removed from the model. Other influenced factors are district, average yield per tree, number of fruit trees, average rainfall and product types, respectively. The least effective parameter is total number of trees. Omitting the total number of trees parameter does not change the test performance well. It means the total number of trees can be omitted in this prediction approach.

Table 4. Effects of parameters on R^2 values

Parameter	R^2	% decrease on R^2	RMSE	% increase on RMSE
District	0,557	43,564	349,188	1273,674
Product types	0,545	44,783	317,284	1148,166
Product specific plant area	0,501	49,293	421,987	1560,058
Average yield per tree	0,543	45,015	180,832	611,375
Number of fruit trees	0,559	43,370	280,901	1005,039
Number of fruitless trees	0,599	39,377	225,169	785,795
Total number of trees	0,983	0,433	25,701	1,105
Population	0,689	30,177	159,542	527,625
Inflation rate	0,600	39,250	227,858	796,372
Total fruit area	0,508	48,574	435,613	1613,663
Average temperature	0,629	36,318	195,105	667,526
Average rainfall	0,558	43,510	276,916	989,361

4. Conclusion

In this study, citrus fruits production amount of different districts of Adana is predicted by the artificial neural network model and regression analysis. The main purpose of the study is to build a generic prediction approach for different citrus fruit products in the same model. For this reason, effect of different attributes which are districts of Adana, product types, product specific plant area, average yield per tree, number of fruitless trees, number of fruit trees, total number of trees, population, inflation rate, total fruit area, temperature, and average rainfall on the production rate are evaluated. The ANN model is compared with the linear regression model in order to find the best prediction approach. ANN model has a good prediction method which provides less errors and more accurate prediction of production rate.

For the future works, other parameters such as other cities or fertilizer sales can be taken into consideration or other artificial neural network approaches can be tried.

5. References

Abrougui, K., Gabsi, K., Mercatoris, B., Khemis, C., Amami, R., and Chehaibi, S. 2019. "Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR)", Soil and Tillage Research, 190, 202-208.

Alp, M. and Cığızoğlu, K. 2004. "Farklı yapay sinir ağı metotları ile yağış-akış ilişkisinin modellenmesi". *İTÜ Dergisi/ D Mühendislik* 3(1), 80-88.

Baş, N. 2006. "Yapay sinir ağları yaklaşımı ve bir uygulama." Mimar sinan güzel sanatlar üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstanbul.

Boukelia, T. E., Ghellab, A., Laouafi, A., Bouraoui, A., and Kabar, Y. 2020. "Cooling performances time series of CSP plants: Calculation and analysis using regression and ANN models", Renewable Energy.

Çakır, Y., Kırcı, M., and Güneş, E. O. 2014. "Yield prediction of wheat in south-east region of Turkey by using artificial neural networks", In 2014 The Third International Conference on Agro-Geoinformatics (pp. 1-4). IEEE.

Çolak, C., Çolak, C.M., and Atıcı, M.A. 2005. "Ateroskloroz'un tahmini için bir yapay sinir ağı", Ankara Üniversitesi Tıp Fakültesi Mecmuası , 58, 159-162.

Elizondo, D.A., McClendon, R.W., Hoongenboom, G. 1994. "Neural network models for predicting flowering and physiological maturity of soybean". Transactions of the ASABE 37, 981–988

Fiona, M. R., Thomas, S., Maria, I. J., and Hannah, B. 2019. "Identification Of Ripe And Unripe Citrus Fruits Using Artificial Neural Network". In Journal of Physics: Conference Series, 1362(1), 012033.

Grossberg, S. 1988. "Nonlinear neural networks: principles. Mechanisms, and architectures", Neural Networks 1(1), 17-61.

Ho, S.L., Xie, M., and Goh, T.N. 2002. "A comparative study of neural network and box-jenkins arima modelling in time series prediction", Comput. Ind. Eng. 42, 371-375.

Hosseinzadeh, A., Baziar, M., Alidadi, H., Zhou, J. L., Altaee, A., Najafpoor, A. A., and Jafarpour, S. 2020. "Application of artificial neural network and multiple linear regression in modeling nutrient

recovery in vermicompost under different conditions", Bioresource Technology, 303, 122926.

Ji, B., Sun, Y., Yang, S., and Wan, J. 2007. "Artificial neural networks for rice yield prediction in mountainous regions", The Journal of Agricultural Science, 145(3), 249-261.

Kaul, M., Hill, R. L., and Walthall, C. 2005. "Artificial neural networks for corn and soybean yield prediction", Agricultural Systems, 85(1), 1-18.

Kisi, Ö. 2004. "Multi-layer perceptrons with Levenberg-Marquardt training algorithm for suspended sediment concentration prediction and estimation / Prévision et estimation de la concentration en matières en suspension avec des perceptrons multi-couches et l'algorithme d'apprentissage de Levenberg-Marquardt," Hydrological Sciences Journal, 49(6),1025-1040.

Kohzadi, N., Boyd, M.S., Kermanshahi, B., and Kaastra, I. 1996. "A comparison of artificial neural network and time series model for forecasting commodity price", Neurocomputing 10, 169-181.

Luxhoj, J.T, Riis, J.O, and Stensballe, B. 1996. "Neural network models for forecasting univariate time series", Neural Networks World 6, 747-772.

Matsumura, K., Gaitan, C. F., Sugimoto, K., Cannon, A. J., and Hsieh, W. W. 2015. "Maize yield forecasting by linear regression and artificial neural networks in Jilin, China", The Journal of Agricultural Science, 153(3), 399-410.

Movagharnejad, K. and Nikzad, M. 2007. "Modelling of tomato drying using artificial neural network", Journal Computers and Electronics in Agriculture, 59(1), 2.

Parmar, R.S. 1997. "All estimation of aflatoxin contamination in preharvest peanuts using neural networks", Transactions of the ASAE, 40(3), 809-813.

Priddy, K. L., and Keller, P. E. 2005. "Artificial neural networks: an introduction" SPIE press, 68.

Prasetyo, D., and Bimantaka, R. D. M. 2018. "Identification of Red Dragon Fruit Using Backpropagation Method Based on Android". International Journal of Applied Business and Information Systems, 2(2), 40-45.

Prybutok, V.R, Yi, J, and Mitchell, D. 2000. "Comparison of neural network models with arima and regression models for prediction of Houston's daily maximum ozone concentrations", Eur. J. Oper. Res. 122, 31-40.

Şahin, C., Oğulata, S.N., Kırım, S., and Koçak, M. 2004. "Troid bezi bozukluklarının yapay sinir ağları ile teşhisi," YA/EM'2004. Yöneylem Araştırması/ Endüstri Mühendisliği XXIV Ulusal Kongresi, Gaziantap- Adana.

Sofu, A.B. 2006. "Yoğurtların depolama esnasında mikrobiyal ve kimyasal değişimlerinin bilgisayarlı görüntüleme sistemiyle belirlenmesi ve elde edilen verilerin yapay sinir ağlarıyla değerlendirilmesi", Gıda Mühendisliği Anabilim Dalı, Isparta.

Tamari, S., Ruiz-Sudrez, J.C., and Wösten, J.H.M. 1996. "Testing an artificial neural network for predicting soil hydraulic conductivity", Proceedings of 6th Intern. Conf. On Computers In Agriculture, Mexico: 912-919.

Torkashvand, A. M., Ahmadi, A., and Nikravesh, N. L. 2017. "Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear

regressions (MLR)", Journal of integrative agriculture, 16(7), 1634-1644.

Tsenga, F.M., Yu, H.C., and Tzeng, G.H. 2002. "Combining neural network model with seasonal time series arima model", Technol. Forecasting Soc. Change 69, 71-87.

TSI, 2017. "Dataset" http://www.tuik.gov.tr, (15.12.2019)

Turkish State Meteorological Service (TSMS), 2018, "Rain and temperature data", (15.12.2019)

Tušek, A. J., Jurina, T., Benković, M., Valinger, D., Belščak-Cvitanović, A., and Kljusurić, J. G. 2020. "Application of multivariate regression and artificial neural network modelling for prediction of physical and chemical properties of medicinal plants aqueous extracts", Journal of Applied Research on Medicinal and Aromatic Plants, 16, 100229.

Wang, J.H. and Leu, J.Y. 1996. "Stock market trend prediction using Arima-based neural networks", IEEE Int. Conf. Neural Networks 4, 2160-2165.

Yarar, A. 2004. "Beyşehir su gölü su seviyesi değişimlerinin yapay sinir ağları ile belirlenmesi," Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, İnşaat Mühendisliği Ana Bilim Dalı, Konya.

Yu, L., Wang, S.Y., and Laic, K.K. 2005. "A novel nonlinear ensemble forecasting model incorporating glar and Ann for foreign exchange rates", Comput. Oper. Res. 32, 2523-2541.

Yurtoğlu, H. 2005. "Yapay sinir ağları metodolojisi ile öngörü modellemesi: bazı makroekonomik değişkenler için Türkiye örneği", Uzmanlık Tezi, Ekonomik Modeller Ve Stratejik Araştırmalar Genel Müdürlüğü, Ankara.

Zou, H.F, Xia, G.P, Yang, F.T, and Wang, H.Y. 2007. "An investigation and comparison of artificial neural network and time series models for Chinese food grain price forecasting". Neurocomputing 70, 2913-292.