ARTIFICIAL INTELLIGENCE STUDIES
YAPAY ZEKA ÇALIŞMALARI
2018, 1(1):1-7

www.aistudies.org

artificial intelligence studies

Gazi AKADEMİK YAYINCILIK

Yaşam Beklentisinin Tahmin Edilmesinde YSA Kullanımı: Türkiye Örneği

Ayhan AYDIN[*,a], Ümit ATİLA[b] Serpil AYDIN[c]

[a,*] *Ondokuz Mayıs Üniversitesi, SAMSUN 55200, TÜRKİYE*
[b] *Karabük Üniversitesi Bilgisayar Mühendisliği Bölümü, KARABÜK 78050, TÜRKİYE*
[c] *Ondokuz Mayıs Üniversitesi İstatistik Bölümü, SAMSUN 55200, TÜRKİYE*

MAKALE BİLGİSİ

Alınma: 10.04.2018
Kabul: 05.06.2018

***Anahtar Kelimeler*:**
YSA, Hayat Beklentisi, Zaman serileri, ADF

Sorumlu Yazar:
e-posta: ayhan@ ayhanaydin.net

ÖZET

Yaşam beklentisi, refah, sağlık düzeyi ve ülkelerin gelişimi gibi parametrelerin karşılaştırılması için kullanılan belirleyici bir ölçektir. Bu göstergenin yükselen trend izlemesi, ancak bu belirleyicilerin olumlu ve olumsuz etkilerinin belirlenmesi ve bu yönde inisiyatiflerin alınmasıyla başarılabilir. Çalışmamızda, ülkemizdeki yaşam beklentisini, sosyal, ekonomik ve çevresel faktörler açısından, yaşam beklentisini etkileyen zaman serisi modelleri ve Yapay Sinir Ağı (YSA) kullanarak modelledik. Çalışmada 1960-2016 yılları arasında OECD ve WORLDBANK'dan derlenen veriler kullanılarak gerçekleştirilen iki tahminci modelinin karşılaştırılması yapılmaktadır. Sosyal bilimler verileri üzerinde yapılan uygulamada, anlamlı göstergeler YSA yönteminin başarısı ile birlikte yorumlanmıştır. Çalışmanın sonucunda, ülkenin refah düzeyi için belirleyici bir kriter olan doğumdan beklenen yaşam beklentisini olumlu yönde artırmak için bir dizi öneri ve gelişim önerisi sunulmuştur.

DOI: 10.30855/AIS.2018.01.01.01

Use of ANN in Predicting Life Expectancy: The Case of Turkey

ARTICLE INFO

Received: 2018
Accepted: 2018

Keywords:
ANN,Life Expectancy, Time Series, ADF

Corresponding Authors
e-mail: ayhan@ ayhanaydin.net

ABSTRACT

Life expectancy is a good measure for comparing parameters such as welfare, health level and development of countries. The high value for this indicator can only be achieved by identifying the positive and negative effects of these determinants and by making initiatives in this direction. In our study, we observed life expectancy estimation by using time series models and Artificial Neural Network (ANN) in terms of social, economic and environmental factors, that affect the life expectancy. Comparison of two estimator models performed using data compiled from OECD and WORLDBANK of Turkey between 1960-2016. In the application performed on social sciences data, meaningful indicators were interpreted together with the success of the ANN method. As a result of the study, a number of suggestions and development recommendations are presented in order to increase the life expectancy from birth, which is a decisive criterion for the country's level of prosperity, in a positive way.

DOI: 10.30855/AIS.2018.01.01.01

1. INTRODUCTION *(GİRİŞ)*

Preparing a future where more people live longer is one of the biggest challenges of society and the country's governments. In addition to the long-term social and economic impacts on health and care services, effective indicators in the planning of insurance and social security are interpreted and interpreted. It will need to be understood in the best possible way to use this information to try to narrow the range of uncertainty about what is happening and future trends.

Many of the predictions of recent developments in life expectancy have been wrong. Despite the assumption that the rate of recovery in life expectancy will begin to fall, the opposite is true. The rates of recovery in recent years have been historically high. Life expectancy in Turkey has increased over the years. In recent years it has been higher improvements than average. It has many aspects of public policy and its effects for the nature of society [1].

Numbers and developments differ according to socioeconomic group and sex [2]. However, many of the factors that influence diversity have continued to expand [3]. The developments in the field of technology and medicine as well as the increase in social and health investments are compared with the existing subject competencies of ANN and time series methods in our study which will be evaluated in adverse effects that develop over time [4].

2. LITERATURE ANALYSIS *(LİTERATÜR TARAMASI)*

There are some studies in the literature on the explanation of the life expectancy. Discigil et al. analyzed the explanations of health indicators of 176 economic indicators by multiple regression analysis and found that the expected life span and infant mortality rate were more influential than the income level of health expenditures and increase in income level was unreliable [5, 6].

Stewart et al. conducted a panel data analysis of the effects of smoking cessation and obesity on the life expectancy of the United Nations (1971-2006 National Health and Nutrition Examination Survey) and showed the negative effects of these factors [7]. Arpacıoğlu discussed the analysis of poverty in the world and in our country, not only the material aspects of poverty, which is a global problem, but also the effects on prosperity and life expectancy [8].

Teker et al. analyzed the demographic and economic factors such as the national income rate of health expenditures, the ratio of the elderly population to the working population, the number of beds per thousand persons, the number of doctors per thousand hospitals, and the life expectancy of men and women between 1975 and 2009. He used vector correction model to find out factors affecting the life span of men and women. As a result of these analyzes, it was determined that each of the factors examined had a significant effect on the life span of men and women [9].

James et al. analyzed the life expectancy of the population between years 2010-2012 living in UK using Bayesian spatiotemporal forecasting [2]. Marius et al. attempted to predict people's life expectancies by using the average male and female life span and the longest life span (record) of countries using Lee-Carter approach and the Cairns-Blake-Downd method [10].

Mikael et al. analyzed relationships between life expectancy at birth public and private health expenditures with econometric panel time series methods for 34 OECD countries using panel VAR models and impulse response analysis. Results showed importance of the positive relationship between public health expenditures and life expectancy [11].

3. MATERIALS AND METHOD *(MATERYALLER VE METOD)*

WORLDBANK and OECD data of our country are used between 1960-2016 years (Figure 1) [12]. The descriptive variables used were birth rate, mortality rate, mortality rate under 5 years, health expenditures, number of people over 65, urban population, rural population [13]. In the analysis of time series (unit root tests, modelling) EViews 9 and for ANN model Matlab R15 software were used.

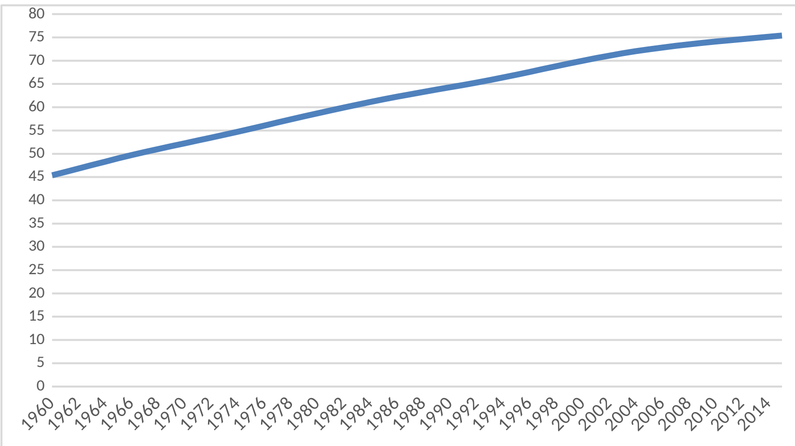

Figure 1. Life expectancy in Turkey

Time series analysis examines the data with processes that have their own characteristics in a certain time interval. This characteristic is that the data is obtained by the result of a stationary probability process. Due to the fact that the time series obtained as a result of the random walk which is a stochastic process, the characteristic cannot be determined. Thus, the time series analysis of the data obtained by this method may not include an estimate, which is often fictional. Therefore, t and F tests related to the results obtained in this way are significant but they cannot give accurate results [14]. Econometrically, the average variance over time does not change and the covariance between the two periods is not the return in which this covariance is calculated. This is just one stop for a possible process that depends on the distance between two periods. Such stasis is also defined as the weak stagnant probability process, and for such analyzes, such stability is sufficient [15].

A. Augmented Dickey-Fuller (ADF) Unit Root Test

For solving the problem of autocorrelation, the ADF test is applied to the right of the equation to which the delay of the dependent variable is added and the new model is applied [16]. This test is based on the article by Said and Dickey [17]. Authors have developed this test by issuing problems from autoregressive time series in their articles. Here, the Akaike and Schwarz criteria are used when determining the appropriate lag rank of the lagged variable [18].

$$\Delta Yt = \beta 1 + \beta 2 + \delta Y \left(t-1 \right) + \alpha i + ut$$

It is a form that recognizes the dependent variable to prevent the model error term from being consecutively dependent.

B. Philips – Perron (PP) Unit Root Test

Although the ADF test is the most common use in unit root tests, there are some deficiencies that the test includes, and these deficiencies are addressed with helper tests. Phillips and Peron have developed unit root tests which are more popular in financial time series in their articles [19]. This test differs from the ADF in that it deals with the serial correlation and the varying variance problem that are leading to errors. Instead of adding delayed values in order to prevent autocorrelation in the ADF equation, the authors rearrange the t statistics by estimating the DF equation. This test is stronger to reject an incorrect H0 hypothesis. Below are the hypothesis tests and statistics used by this test.

$$Yt = \delta Y \left(t-1 \right) + ut \qquad \text{[Constant]}$$

$$Yt = \beta 1 + \delta Y \left(t-1 \right) + ut \qquad \text{[Constant and Tendency]}$$

$$Yt = \beta 1 + \delta Y \left(t-1 \right) + \beta 2 \left(t - \frac{T}{2} \right) + ut \qquad \text{[Coefficient]}$$

The models shown above show the constant term trend coefficient and the number of observations. In all models, the error rate average is equal to zero, may be successive dependent, or may violate heteroskedasticity covariance hypothesis. Therefore, the PP test is not dependent on the assumptions of the DF or the ADF test. Hypothesis testing is done by testing the hypothesis H0 by equality as tested in the DF test, and the rejection of H0 shows us that the series does not have a unit root, that is, it has a static probability processing characteristic.

C. Kwiatowski-Phillips-Schimidt-Shin (KPSS) Unit Root Test

The common point in the DF, ADF, and PP tests mentioned earlier that the hypothesis H0 does not show that the series contains unit root, that is, the steady-state probability characteristic of the data. The opposite is true for the KPSS test. This time the H0 hypothesis will be established suggesting that the review does not include the unit root of the series. The hypothesis of the KPSS test is as follows:

H0: $\sigma u2 = 0$ unit root is not exist / data is stationary.
H1: $\sigma u2 \neq 0$ unit root exist / data is not stationary.

According to the article, Kwiatkowski et al. [20] the above basic hypotheses are the following equations for testing:

$$yt = \xi t + rt + \varepsilon t$$

Here, the deterministic tendency coefficient showing tendency (trend) indicates random tendency and disturbance [21]. The random term is delayed by a value in a relationship as follows:

$$rt = r\left(t-1\right) + ut$$

Where u is the error term between the lagged value of the random term and itself. This error is an error term in which specific assumptions for the KPSS test are established, so the error term is equal to the consecutive non-dependent and homoskedasticity principles (σu2). The fact that the variance of this error term is equal to zero fulfills the condition that allows rt to be stationary. The test statistical threshold values were determined by the Lagrange multiplier and published in study of Kwiatkowski et al. [20].

D. Artificial Neural Networks

ANN is an attempt to create a new system by imitating the human brain [22]. These approaches constitute experiences from the information obtained from the examples and then give similar decisions on similar issues. Multi-Layer Perceptron (MLP) is the most famous ANN structure used for solving non-linear regression and classification problems (Figure 2). In an MLP;

Input Layer: Receives input from the outside and sends it to the hidden layer. There is no information processing in this layer. Each incoming information goes directly to the next layer.
Hidden Layer: Number of hidden layer can be one or more. Hidden layer neurons process the information coming from the input layer or other hidden layers and sends it to the next layer.
Output Layer: The information from the hidden layer is processed and sent out as network response.

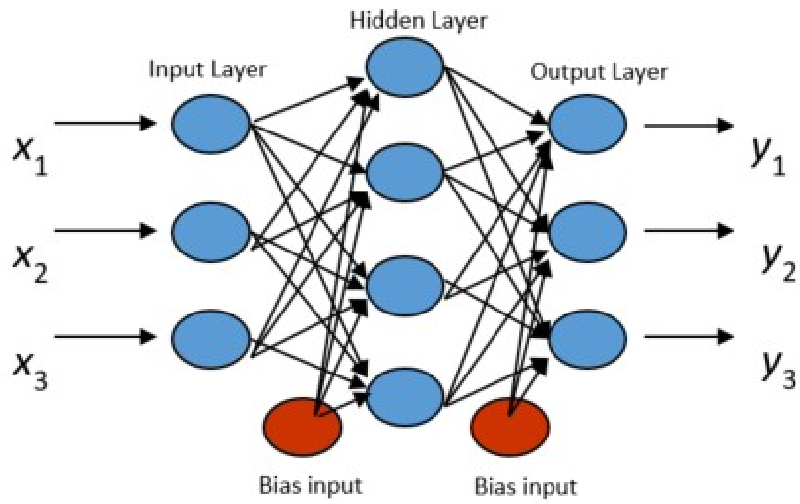

Figure 2. Multi-Layer Perceptron Structure.

Most famous learning algorithm used for training MLP is Backpropagation Algorithm. It was first developed by Werbos and later developed by Parker, Rummelhart and McClelland [23]. The Standard Backpropagation Algorithm, which performs operations in two stages, namely Propagate and Adapt, is a multi-layered, forward-feed and instructor-trained ANN model with a complete connection between the layers. Standard Backpropagation Algorithm; received this name because its tried to reduce the mistakes backwards from entry to entry. Backpropagation learning is used to recalculate the weights of each layer according to the current error level at the network output. In a back-propagation network model, it is possible to increase the number of hidden layers according to the characteristics of the problem.

4.APPLICATION ***(UYGULAMA)***

First, unit root tests and structural fracture controls were performed on time series using EViews program. After the control of the series is made and the stability is obtained in the second difference, time series model acquired presented in Table 1.

As a result of the time series analysis, we have obtained a model with over 80% explanatory power and suffered data loss due to the second difference being captured. The fact that the entire data considered as input cannot be understood in the analysis and the failure correction model does not work, can also be counted as the weaknesses of the model.

Table 1. Life Expectancy Time Series Model TURKEY.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
2.difdeathrt	-1.563.520	0.193991	8.059.737	0.0000
2.difbirthrt	-0.194781	0.048328	4.030.425	0.0001
2.difover65p	0.273248	0.087290	3.130.345	0.0025
2.difHEALTH_EXP_GDP	-0.008899	0.000919	9.683.906	0.0000
C	0.045527	0.004583	9.933.097	0.0000
R-squared	0.869410	Mean dependent var		-0.014901
Adjusted R-squared	0.871253	S.D. dependent var		0.018567
S.E. of regression	0.008655	Akaike info crit.		-6.173.248
Sum squared resid	0.003200	Schwarz criterion		-6.924.315
Log likelihood	180.7680	Hannan-Quinn cri.		-6.124.654
F-statistic	34.74419	Durbin-Watsonstat		1.086.953
Prob(F-statistic)	0.000000			

For training process with ANN, after normalizing training dataset, the data of 5 inputs and 55 rows were subjected to MATLAB NNTool and the following model shown in Figure 3 was created.

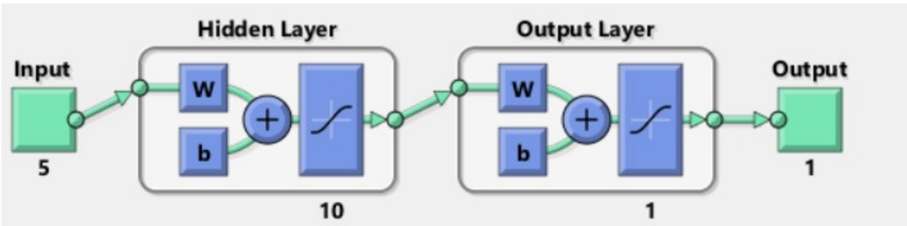

Figure 3. The ANN Model

Estimated values by ANN model, desired outputs and errors are shown for each year in Table 2.

Table 2. ANN estimation and error values.

Yea	Estimation	Real Value	Error	% Error
1974	54,85981613	54,73365854	0,126157598	0,230493632
1975	55,26903859	55,38521951	-0,116180922	-0,209768821
1976	55,98907380	56,04975610	-0,060682300	-0,108265056
1977	56,72963773	56,71868293	0,010954804	0,019314277
1978	57,40611400	57,38500000	0,021114002	0,036793591
1979	58,04244246	58,04426829	-0,001825838	-0,003145595
1980	58,67629010	58,69158537	-0,015295270	-0,026060414
1981	59,31778691	59,32612195	-0,008335043	-0,014049533
1982	59,95060088	59,94646341	0,004137462	0,006901929
1983	60,56129427	60,55265854	0,008635729	0,014261519
1984	61,14904014	61,14319512	0,005845015	0,009559551
1985	61,71301155	61,71443902	-0,001427471	-0,002313026
1986	62,26127116	62,26226829	-0,000997137	-0,001601510
1987	62,78885732	62,78751220	0,001345125	0,002142344

1988	63,29763065	63,29453659	0,003094069	0,004888366
1989	63,79213932	63,79031707	0,001822242	0,002856613
1990	64,27796513	64,28234146	-0,004376331	-0,006807983
1991	64,78301725	64,78309756	-8,03134E-05	-0,000123973
1992	65,29985322	65,30021951	-0,000366289	-0,000560931
1993	65,83095387	65,83790244	-0,006948567	-0,010554054
1994	66,38789632	66,39873171	-0,010835389	-0,016318668
1995	66,97543546	66,98192683	-0,006491374	-0,009691231
1996	67,59150394	67,58621951	0,005284426	0,007818793
1997	68,20715895	68,19968293	0,007476026	0,010961965
1998	68,8170254	68,81226829	0,004757111	0,006913173
1999	69,41242045	69,41539024	-0,002969797	-0,004278297
2000	69,99065654	69,99934146	-0,008684927	-0,012407155
2001	70,54284409	70,55580488	-0,012960785	-0,018369552
2002	71,07849290	71,08197561	-0,003482712	-0,004899571
2003	71,58003053	71,57419512	0,005835407	0,008152948
2004	72,04352903	72,02987805	0,013650983	0,018951834
2005	72,46423352	72,44812195	0,016111564	0,022238760
2006	72,84111031	72,82912195	0,011988358	0,016460939
2007	73,17767390	73,17812195	-0,000448055	-0,000612280
2008	73,48736201	73,50282927	-0,015467256	-0,021043075
2009	73,78381913	73,80631707	-0,022497945	-0,030482411
2010	74,08172338	74,09404878	-0,012325396	-0,016634799
2011	74,36929092	74,36941463	-0,000123713	-0,000166349
2012	74,65561805	74,63629268	0,019325369	0,025892723
2013	74,93119499	74,89604878	0,035146213	0,046926659
2014	75,16501159	75,15214634	0,012865248	0,017118936
2015	75,32047585	75,40709756	-0,086621710	-0,114872092

Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) of the ANN model is 0.001084 and 0.032924 respectively. While MSE of the ANN model was obtained very close to zero performances of ANN and time series model is compared in terms of R-Squared value which is a statistical measure of how close the data are to the fitted regression line. In other words, R-Squared value is the percentage of the estimated value variation that is explained by the model. %0 indicates that model explains none of the variability of the estimated values around its mean while 100% indicates that model explains all the variability of the estimated values around its mean. R-Squared values of the ANN model and time series model applied in this study are as shown in Table 3. This shows that ANN model which is constructed using all the decision-making units has better explanatory power than time series analysis for estimating life expectancy.

Table 3. R-Squared Values of Applied Models

	ANN Model	Time Series Models
R-Squared	0.9995	0.8694

5. RESULTS AND RECOMMENDATIONS *(SONUÇLAR VE ÖNERİLER)*

Although different criteria are taken into consideration in the literature it has been revealed that the welfare levels and socio-economic status of countries are the main determinants of life expectancy at birth [24]. Therefore, it is important to correctly estimate the life expectancy in the interpretation of the social economic development and the trends of the countries. In this study life expectation estimation capacity of time series and ANN models were examined. As a result, ANN model provided better fit with 99% to the data compiled from OECD and WORLDBANK of Turkey that belongs to years 1960-2016.

The fact that most of the child deaths are below 5 years of age and that only 2-5% of the deaths in developed countries are below the age of 5 indicates that child health should be prioritized in health investments in developing countries [25]. Secondly, how a society maintains older people is important to understand the place of the elderly in society. At the same time, it should be determined how many of the elders receive a regular pension, how many of them are cared for by their relatives, and how long they live alone.

By using the methods with high success like ANN in performing life expectancy estimations, the right choices can be made to overcome the current and future problems of different age groups and to determine the proportion of health investments in the relevant area. For different age groups such as infants, children under 5 years, adults and individuals over 65 years of age, ANN can effectively be used to shape future health investments and functions.

KAYNAKLAR (REFERENCES)

[1] Q. Li, S. Ma, D. Bishai, and A. A. Hyder, "Potential gains in life expectancy by improving road safety in China," *Public health,* vol. 144, pp. S57-S61, 2017.

[2] J. E. Bennett, G. Li, K. Foreman, N. Best, V. Kontis, C. Pearson*, et al.*, "The future of life expectancy and life expectancy inequalities in England and Wales: Bayesian spatiotemporal forecasting," *The Lancet,* vol. 386, pp. 163-170, 2015.

[3] S. Erdoğan and H. Bozkurt, "Türkiye'de yaşam beklentisi-ekonomik büyüme ilişkisi: ARDL modeli ile bir analiz," *Bilgi Ekonomisi ve Yönetimi Dergisi,* vol. 3, 2008.

[4] J. V. Tu, "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes," *Journal of clinical epidemiology,* vol. 49, pp. 1225-1231, 1996.

[5] G. Dişcigil, N. Şensoy, N. Tekin, and A. Söylemez, "Breast health: Knowledge, behaviour and performance in a group of women living in aegean region," 2007.

[6] G. Discigil, N. Tekin, and A. Soylemez, "Obesity in Turkish children and adolescents: prevalence and non-nutritional correlates in an urban sample," *Child: care, health and development,* vol. 35, pp. 153-158, 2009.

[7] S. T. Stewart, D. M. Cutler, and A. B. Rosen, "Forecasting the effects of obesity and smoking on US life expectancy," *New England Journal of Medicine,* vol. 361, pp. 2252-2260, 2009.

[8] Ö. A. M. YILDIRIM, "Dünyada ve Türkiye'de yoksulluğun analizi," *Ömer Halisdemir Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* vol. 4, p. 60, 2011.

[9] D. Teker, S. Teker, and M. Sönmez, "Ekonomik değişkenlerin kadın ve erkeğin yaşam süresine etkisi," *İşletme Araştırmaları Dergisi,* vol. 4, pp. 118-126, 2012.

[10] M. D. Pascariu, V. Canudas-Romo, and J. W. Vaupel, "The double-gap life expectancy forecasting model," *Insurance: Mathematics and Economics,* vol. 78, pp. 339-350, 2018.

[11] M. Linden and D. Ray, "Life expectancy effects of public and private health expenditures in OECD countries 1970–2012: Panel time series approach," *Economic Analysis and Policy,* vol. 56, pp. 101-113, 2017.

[12] T. W. B. Group. (20.10.2018). Available: https://data.worldbank.org/country/turkey?view=chart

[13] S. Demir, *Birleşmiş Milletler Kalkınma Programı İnsani Gelişme Endeksi ve Türkiye Açısından Değerlendirme*: DPT, 2006.

[14] N. Gujarati Damodar, "Sangeetha.(2007). Basic Econometrics," ed: The McGraw-Hill Hardell, L., Nasman, A., Pahlson, A., Hallquist, A., & Hansson Mild, K, 1999.

[15] P. Egüez-Guevara and F. C. D. Andrade, "Gender differences in life expectancy with and without disability among older adults in Ecuador," *Archives of gerontology and geriatrics,* vol. 61, pp. 472-479, 2015.

[16] L. Iliadis and K. Kitikidou, "Combination of Survival Analysis and Neural Networks to Relate Life Expectancy at Birth to Lifestyle, Environment, and Health Care Resources Indicators," in *Engineering Applications of Neural Networks*, ed: Springer, 2011, pp. 491-498.

[17] S. E. Said and D. A. Dickey, "Testing for unit roots in autoregressive-moving average models of unknown order," *Biometrika,* vol. 71, pp. 599-607, 1984.

[18] W. Enders and B. Falk, "Threshold-autoregressive, median-unbiased, and cointegration tests of purchasing power parity," *International Journal of Forecasting,* vol. 14, pp. 171-186, 1998.

[19] P. C. Phillips and P. Perron, "Testing for a unit root in time series regression," *Biometrika,* vol. 75, pp. 335-346, 1988.

[20] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?," *Journal of econometrics,* vol. 54, pp. 159-178, 1992.

[21] O. Peker, "Türkiye'deki Cari Açık Sürdürülebilir mi? Ekonometrik Bir Analiz," *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi,* pp. 164-174, 2009.

[22] A. Abraham, "Artificial neural networks," *handbook of measuring system design,* 2005.

[23] D. Rummelhart, J. McClelland, and P. R. Group, "Parallel Distributed Processing, vol. 1, 2," ed: MIT Press, Cambridge, MA, 1986.

[24] A. Golan, G. G. Judge, and D. Miller, *Maximum entropy econometrics: Robust estimation with limited data*: Wiley New York, 1996.

[25] T. Kabaş, "Gelişmekte olan ülkelerde yoksulluğun nedenleri ve yoksullukla mücadele yolları," *Doktora Tezi, TC Çukurova Üniversitesi, Sosyal Bilimler Enstitüsü, İktisat Anabilim Dalı, Adana,* 2009.