Journal of Computing, Science &Technology
https://focjournal.unidel.edu.ng/
editorjcst@unidel.edu.ng
ISSN:.....Print
ISSN: ..... Online

Assessing the Hybrid of ARIMAX and ANN Models for Predicting Life Expectancy in Nigeria

Christopher Awariefe[1] & Godwin Ekruyota[2]
[1]Department of Statistics, Delta State University of Science & Technology, Ozoro. Nigeria.
[2]Department of Computer Science, Delta State University of Science & Technology, Ozoro. Nigeria.
awariefec@gmail.com[1], go.softsystem@gmail.com[2]

Corresponding Author's Email: awariefec@gmail.com

Article Info

Date Received: 08-02-2024
Date Accepted: 28-02-2024

Keywords:
Life Expectancy, ARIMA, ARIMAX, ANN, Hybrid ARIMAX-ANN.

ABSTRACT

This research investigates the relationship between life expectancy and key variables in Nigeria, employing advanced forecasting models to address a research gap in understanding the collective influence of economic and demographic factors on life expectancy. The study focuses on the Hybrid ARIMAX-ANN method and compares its performance to ARIMA, ARIMAX, and ANN models. Data from 1960 to 2022 are utilized, with auto.arima() for ARIMA model selection, ARIMAX incorporating exogenous variables, and ANN employing a feed-forward neural network. The research emphasizes accurate model selection for reliable forecasting. Results show that the Hybrid ARIMAX-ANN model outperforms other models, demonstrating superior accuracy in predicting Nigerian life expectancy. Performance metrics such as MAPE and RMSE confirm the effectiveness of the hybrid model. Visual comparisons and forecast accuracy evaluations of the train data further support these findings. The study contributes to the predictive modelling discourse for life expectancy in developing nations, providing insights into the interplay of economic and demographic factors in Nigeria. These insights are expected to guide evidence-based policymaking and public health interventions, leading to improvements in population health outcomes in Nigeria and potentially offering strategies for similar contexts globally.

1.0 INTRODUCTION

Life expectancy, a critical metric reflecting a population's overall health and longevity, serves as a pivotal indicator in the domains of public health and demographic studies. The existing literature has explored the relationship between life expectancy (LE) and key variables such as GDP per capita, population growth, and crude death rate, providing valuable insights into the dynamics shaping population health. In Nigeria, where diverse demographics and health challenges exist, accurate life expectancy predictions are crucial for effective public health planning and policy formulation. Studies investigating the impact of economic development, as measured by GDP per capita, on LE have found a positive correlation, emphasizing the role of improved living standards and healthcare access in enhancing overall well-being [2, 6]. Concurrently, investigations into the effects of population growth and the crude death rate on life expectancy have underscored their significance in understanding demographic dynamics and mortality patterns [16, 22]. Research about forecasting employing ARIMA, ARIMAX and ANN methods has been done previously; for example, [25] used the ARIMA model to project trends in Life expectancy (LE) and healthy life expectancy (HALE), and GAP across global regions from 1995 to 2025. [20] used multiple regression and ANN models to evaluate life expectancy in Bangladesh based on gross domestic product and population size, and [26] used ANN and time series models to study the impact of social, economic, and environmental factors on longevity in Turkey. This research considers employing an Artificial Neural Network (ANN), a combination of ARIMAX and ANN called Hybrid ARIMAX-ANN. The combination of these approaches is made to obtain more precise end results [14]. Within Nigeria, research on life expectancy has often centred on the impact of infectious diseases and healthcare infrastructure [1, 8]. These studies provide valuable insights into the unique health challenges facing the country but may not comprehensively capture the broader determinants, including economic and demographic factors.

Globally, [16] highlighted the intricate relationship between population growth and life expectancy, recognizing the potential for demographic shifts to influence mortality trends. The findings underscore the importance of considering population dynamics in understanding life expectancy variations. Despite the wealth of global research on these determinants, there exists a notable research gap concerning their collective influence on life expectancy in the specific context of Nigeria. The intricate socio-economic landscape, coupled with unique demographic challenges, necessitates a focused examination of these variables within the Nigerian framework. This study aims to address this gap by employing Hybrid ARIMAX-ANN models to forecast life expectancy, integrating GDP per capita, population growth, and the crude death rate as key independent variables. Subsequently, a comparison is made between

the hybrid ARIMAX-ANN model and the ARIMA, ARIMAX, and ANN models. Next, the outcomes of the forecast accuracy and performance are examined. By employing advanced forecasting models, the study contributes to the broader discourse on predictive modelling for life expectancy in developing nations. The insights garnered from this research are anticipated to guide evidence-based policymaking and public health interventions, fostering improvements in overall population health outcomes in Nigeria and potentially informing strategies for similar contexts worldwide.

2.0 METHODOLOGY

2.1 Data collection and pre-processing

The research analyzes data on Nigerian life expectancy (LE), population growth (POPG), GDP per capita (GDPPC), and crude death rate (CDR) from the World Bank Development Indicators website, covering the period from 1960 to 2022. The data will be divided into training and testing sets using an 80:20 percent ratio.

ARIMA Modelling

This study focuses on the efficacy of the auto.arima() function in automating ARIMA model selection for univariate time series data. The function efficiently identifies optimal ARIMA models by employing automated parameter selection guided by either the Akaike Information Criterion or the Bayesian Information Criterion. Drawing on insights from [11], the study underscores the significance of accurate model selection in time series analysis. The auto.arima() function's ability to detect seasonality aligns with recommendations by [3]. After selection, the study emphasizes the importance of diagnostic checks on residuals to validate the fitted ARIMA model. The ARIMA model is applied to model and forecast Nigerian life expectancy, integrating Autoregressive (AR), Integrated (I), and Moving Average (MA) components. The model's mathematical representation is as follows:

$$L_t = c + a_1 L_{t-1} + \ldots + a_p L_{t-p} + w_t - T_1 w_{t-1} - \ldots - T_q w_{t-q} \quad (1)$$

where:

L_t= is the life expectancy at time 't'

c is a constant term

Autoregressive (AR) Component (a_1): Captures the correlation between current and past observations.

Integrated (I) Component (d): Represents differencing to achieve stationarity in the time series.

Moving Average (MA) Component (T_i): Captures the relationship between the current observation and past residual terms (w_t).

These parameters (a_1, T_i, and d) are determined during the model fitting. The auto.arima () function automates this process, aligning with the principles of automated forecasting emphasized by [11]. The foundational work of Box and Jenkins (1976) provides key insights into time series analysis, supporting the robustness of the ARIMA model in capturing temporal patterns and aiding accurate forecasting in diverse contexts.

ARIMAX Model

In this research endeavour, the ARIMAX model is employed to comprehensively model and forecast Nigerian life expectancy, incorporating exogenous variables of GDP per capita, population growth, and crude death rate. The ARIMAX model is expressed as:

$$L_t = a_1 L_{t-1} + aL_{t-2} + \ldots + a_p L_{t-p} + m_1 R_{t-1} + m_p R_{t-p} + C_1 Z_1 +, \ldots . + C_p Z_{t-p} + w_t \quad (2)$$

where L_t= is the life expectancy at time 't', a_i and m_i are autoregressive and moving average parameters, C_i represent the coefficients for exogenous variables, Z_i denotes the values of the exogenous variables at time t, such as GDP per person, population growth, and crude death rate and w_t is the white noise error term.

This study aligns with the principles of incorporating exogenous factors into time series modelling, as suggested by [7]. Additionally, Hyndman and [10] emphasize the significance of considering external variables in forecasting. The application of the ARIMAX model, as demonstrated by [5], enables a more nuanced and accurate prediction of life expectancy, addressing potential confounding effects of economic and demographic factors.

ANN Model

A feed-forward artificial neural network (FANN) is a computational model based on biological neural networks that processing information unidirectionally through hidden layers and applying activation functions to generate output. A FANN, as elucidated by [12], processes information unidirectionally through interconnected nodes and adjustable weights. The back-propagation algorithm, introduced by [19], enables training by adjusting weights based on prediction errors. Bishop [4] highlights the application of feed-forward neural networks in pattern recognition. Widely used in machine learning, these networks demonstrate versatility in tasks such as image recognition and financial forecasting, showcasing their ability to capture intricate patterns with adaptability and proficiency. The mathematical representation of a basic feed-forward neural network, a common type of ANN, can be described as:

$$L_k = f\left(\sum_{i=1}^{n} g_{ki} \, . \, z_i + \pi_k\right) \quad (3)$$

where:

L_k is the output (predicted life expectancy at time 't') of neuron k in the output layer.

z_i represents the input (exogenous variables such as GDP per capita, population growth, and crude death rate) from neuron i in the previous layer.

g_{ki}denotes the weight associated with each exogenous variable wish is the connection between neuron i and neuron k.

π_k is the bias term for neuron k.

f is the activation function (AF), typically a nonlinear function such as the hyperbolic tangent applied to the weighted sum.

Activation Function (AF): Hyperbolic Tangent (tanh) - a widely adopted AF in neural NNs that introduces non-linearity to the model, facilitating the capture of complex relationships.

The tanh AF is defined as:

$$\tanh(l) = \frac{2^l - 1}{2^l + 1} \quad (4)$$

It constricts the result to a range of -1 to 1, introducing non-linearity to the model. The tanh function is suitable for NNs in time series forecasting tasks as it allows capturing intricate patterns and connections within the data.

The Hybrid ARIMAX-ANN Method

Hybrid models combine nonlinear and linear components, with ARIMAX excelling for linear data and ANN for nonlinear data. These methodologies are useful for understanding data features, because they capture fundamental patterns. Zhang [24] proposed a time-series model with both components. Considering time-series as a combination of linear and nonlinear features, given that Z_t is a time-series, it can be represented as

$$Z_t = E_t + R_t \quad (5)$$

From (5), both E_t and R_t which consist of linear and non-linear components, have to be estimated from the Z_t time series. The fitted ARIMAX model will include linear components of the time series; the residuals obtained from the ARIMAX model will comprise ~~of~~ only the nonlinear features. Let w_t be the residuals from the estimated ARIMAX, and then

$$w_t = Z_t - \hat{F}_t \quad (6)$$

The forecasted value $\hat{F}_t$ and the residual w_t are obtained from the first step, in next step the residuals from the ANN model becomes

$$w_t = Z_t - \hat{N}_t \quad (7)$$

The forecasted values $\hat{F}_t$ and $\hat{N}_t$ from both ARIMAX and ANN models are combined to improve the performance of the time series, the new estimated new series can then be defined as

$$\hat{Z}_t = \hat{F}_t + \hat{N}_t \quad (8)$$

The residual value from the ARIMAX output (prediction) and the exogenous values are processed into an ANN method input to perform the Hybrid ARIMAX-ANN method. The proposed hybrid uses the distinct characteristics of both the ARIMAX and ANN models to establish distinct patterns [27]. Therefore, as inferred by Zhang [24], it will be beneficial to combine the linear and nonlinear properties of different models to improve the performance of forecast accuracy.

3.0 DATA ANALYSIS AND DISCUSSION OF RESULTS

3.1 Data Analysis

In order to obtain the model and predict the data, the data was partitioned into two parts, which are the train subset sample and the test subset sample. The data used for modelling is the train data. Meanwhile the data used for the prediction is the test data. We used 80% of the train data to develop the model. 20% of the remaining data is used as the test data. We computed the prediction accuracy using Mean Percentage Error (MAPE) and Root Mean Square Error (RMSE). MAPE and RMSE were calculated using the following equations:

$$\text{MAPE} = \frac{1}{sn}\sum_i^{sn}\left(\frac{|Z_i - \hat{Z}_i|}{|Z_i|}\right) \quad (9)$$

where:
sn is sample size
Z_i represent the actual values of the sample data
$\hat{Z}_i$ denotes the predicted values of the sample data

$$\text{RMSE} = \sqrt{\frac{1}{sn}\sum_{i=1}^{n}\left(Z_i - \hat{Z}_i\right)^2} \quad (10)$$

4.0 RESULTS AND DISCUSSION

In this study, the Box-Jenkins procedure was applied to construct ARIMA and ARIMAX models for predicting Nigerian LE using the R program and the auto. arima () function from the forecast package. These models were compared with ANN models, including the hybrid version of ARIMAX-ANN. The ANN models incorporated residuals from ARIMAX as inputs. We used the neuralnet library in R to vary the number of nodes in the hidden layer. The optimal pure ANN model had 4 nodes in each hidden layer, while the best hybrid model had 3 nodes. We conducted the performance evaluation of the models using metrics such as MAPE and RMSE. Table 1 summarizes the results

Table 1: MAPE and RMSE Estimates

Model	MAPE (%)	RMSE
ARIMA	51.29	0.334
ARIMAX	8.26	0.043
ANN	1.69	0.023
ARIMAX-ANN	1.42	0.019

Source: Prepared by the researchers based on R software outputs

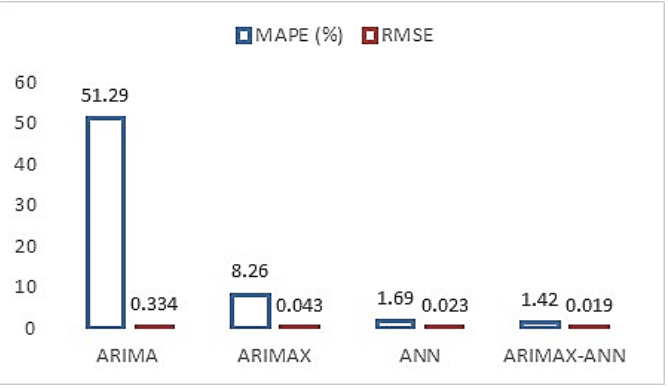

Figure 1: Comparative MAPE and RMSE plot of the models (Source: Prepared by the researchers based on R software outputs).

The test data showed that ARIMA had a high MAPE (52.29%) and RMSE (0.334), ARIMAX did better (MAPE

8.26%, RMSE 0.043), and pure ANN did better than both (MAPE 1.42%, RMSE 0.019). Hybrid models, ARIMA-ANN and ARIMAX-ANN, showed competitive performance. Additionally, we have created a bar graph representing the applied models to analyse their relative performances, depicted in Figure 1. Upon examining figure 1, it becomes evident that the ANN and ARIMAX-ANN models exhibit the lowest Root Mean Squared Error (RMSE) values when compared to the other models.

This study is consistent with existing literature highlighting the efficacy of hybrid models, combining statistical and machine learning approaches for time series forecasting [21, 23]. Results suggest the proposed models improve predictions of Nigerian Life Expectancy, aligning with similar works in the field [13,15].

Figure 2 illustrates that the predicted values of ARIMA align with the actual Life Expectancy pattern. Therefore, the ARIMA model will be appropriate for forecasting.

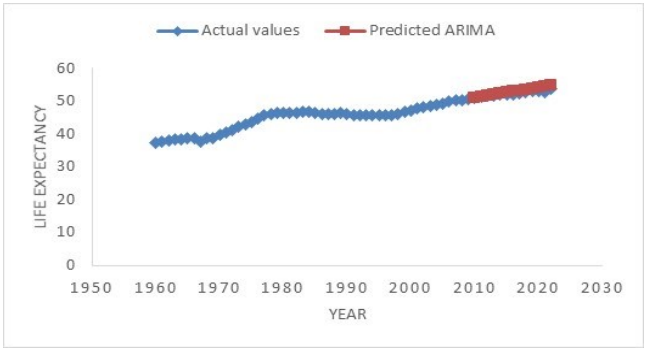

Figure 2: Predicted values of ARIMA vs Actual Life Expectancy (Source: Prepared by the researchers based on R software outputs)

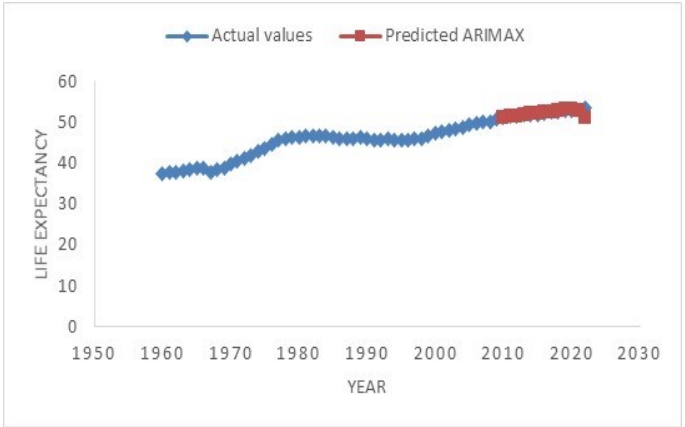

Figure 3: Predicted values of ARIMAX vs Actual Life Expectancy (Source: Prepared by the researchers based on R software outputs)

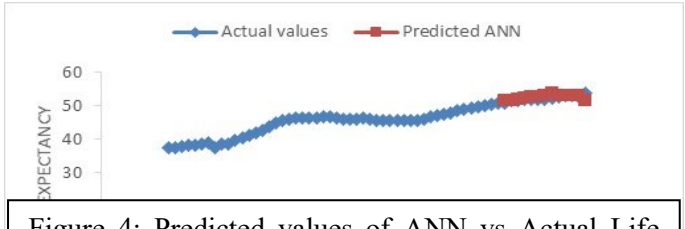

Figure 4: Predicted values of ANN vs Actual Life Expectancy (Source: Prepared by the researchers based on R software outputs)

Figure 3 shows that the ARIMAX prediction values from 2010 to 2022 roughly follow the real life expectancy. It is a suitable model that can be applied. The Life Expectancy (LE) predictions by ANN and the actual LE, which are somewhat comparable, are shown in Figure 4.

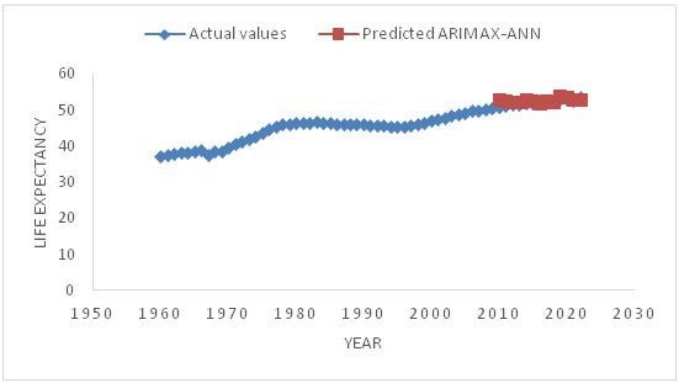

Figure 5: Predicted values of Actual Life Expectancy vs the Hybrid model (Source: Prepared by the researchers using results from the R software)

Figure 5 depicts how ARIMAX-ANN predicted outcomes closely approximate actual LE values. This model may be considered for the efficient forecasting model.

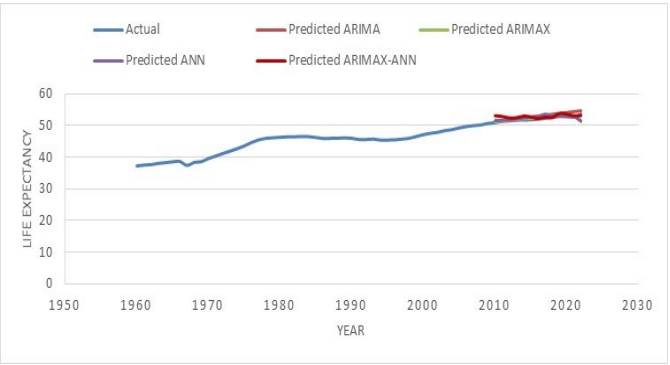

Figure 6: Predicted values comparison between ARIMA, ARIMAX, ANN, and Hybrid ARIMAX-ANN (Source: Prepared by the researchers based on R software outputs)

Figure 6 illustrates how the patterns of the actual data and the predictions made by the four models match up. The ARIMA model does not, however, perform as well as the other four models. The hybrid model (HM) and the ANN model do better in forecasting. The hybrid forecast model provided the lowest MAPE of all the forecasts, as shown in Table 1. The results indicate that both ARIMAX and ANN models' forecasting accuracy is greatly improved by the HM. In this case, its capacity to accurately capture patterns that display linearity as well as those that deviate from linearity proved useful.

5.0 CONCLUSIONS AND FUTURE WORKS

In this article, a forecasting framework was developed and used to assess the hybrid ARIMAX and ANN models for predicting the life expectancy of Nigerians. The World Bank Development Indicators website provided the data utilized in the analysis. ANN, hybridization between

ARIMAX and ANN, and ARIMAX are the models that are used. As a conclusion to this comparison analysis, performance measure data show that Nigerian life expectancy can be successfully predicted by the ARIMA, ARIMAX, and ANN models. Their predictions can represent the actual data's pattern. With the lowest MAPE and RMSE of these three models, the hybrid model emerges as the top model. It demonstrates how the hybrid approach could greatly improve the ARIMAX and ANN models' forecast accuracy. We believe that this approach holds great promise for enhancing forecast performance, particularly with regard to Nigerian life expectancy. In addition, there are numerous ways to combine linear and nonlinear models in future studies.

REFERENCES

[1] B. A. Ahonsi (1993), *Factors affecting infant and child mortality in Ondo State, Nigeria*. University of London, London School of Economics (United Kingdom).

[2] D. E. Bloom, D. Canning & J. Sevilla (2004), The effect of health on economic growth: A production function approach. World Development, 32(1), 1-13.

[3] G. E. P. Box & G. M. Jenkins (1976). Time series analysis: Forecasting and control. San Francisco, CA: Holden-Day.

[4] C. M. Bishop (1995), *Neural networks for pattern recognition*. Oxford university press.

[5] C. Chen & L. M. Liu (1993), Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, *88*(421), 284-297.

[6] D. M. Cutler, A. Deaton & A. Lleras-Muney (2006), The determinants of mortality. Journal of Economic Perspectives, 20(3), 97-120.

[7] W. Enders (2012), Applied econometric time series. *Privredna kretanja i ekonomska politika*, *132*, 93.

[8] A. F. Fagbamigbe & E.S. Idemudia (2016), Survival analysis and prognostic factors of timing of first childbirth among women in Nigeria. *BMC pregnancy and childbirth*, *16*(1), 1-12.

[10] R. J. Hyndman & G. Athanasopoulos (2018), Forecasting: principles and practice (2nd ed.). OTexts.

[11] R. J. Hyndman & Y. Khandakar (2008), Automatic time series forecasting: the forecast package for R. Journal of statistical software, 27, 1-22. https://www.jstatsoft.org/article/view/v027I03.

[12] S. Haykin (1998). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

[13] W. H. Hong, J. H. Yap, G. Selvachandran, P H. Thong & L. H. Son (2021), Forecasting mortality rates using hybrid Lee–Carter model, artificial neural network and random forest. *Complex & Intelligent Systems*, *7*, 163-189.

[14] L. A. Díaz-Robles, J. C. Ortega, J. S. Fu, G. D. Reed, J. C. Chow, J. G. Watson & J. A. Moncada-Herrera (2008), A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmospheric Environment*, *42*(35), 8331-8340.

[15] S. Levantesi, A. Nigri & G. Piscopo (2022), Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. *International Journal of Approximate Reasoning*, *140*, 282-297.

[16] W. Lutz, W. P. Butz & K. C. Samir (Eds.), (2017). World Population and Human Capital in the Twenty-First Century. Oxford University Press.

[17] S. H. Preston (1975), The Changing Relation between Mortality and Level of Economic Development. Population Studies, 29(2), 231-248.

[19] D. E. Rumelhart, G. E. Hinton & R. J. Williams (1986), Learning representations by back-propagating errors. *nature*, *323*(6088), 533-536.

[20] M. A. Rubi, H. I. Bijoy and A. K. Bitto (2021), Life Expectancy Prediction Based on GDP and Population Size of Bangladesh using Multiple Linear Regression and ANN Model. *12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kharagpur, India. pp. 1-6.

[21] T. H. Shakiru, X. Liu & Q. Liu (2023), A hybrid Modelling and Forecasting of Carbon dioxide Emissions in Tanzania. *General Letters in Mathematics (GLM)*, *13*(1).

[22] J. Vallin & F. Meslé (2004), Convergences and divergences in mortality: A new approach of health transition. Demographic Research, Special Collection, 2, 11-44.

[23] Q. Wang & L. Zhang (2020), Hybridizing Time Series Models with Machine Learning for Improved Forecasting. International Journal of Data Science and Analytics, 8(2), 143-158.

[24] G. P. Zhang (2003), Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159-175.

[25] X. Cao, Y. Hou, X. Zhang, C. Xu, P. Jia, X. Sun, L. Sun, Y. Gao, H. Yang, Z. Cui, Y. Wang and Y. Wang (2020), A comparative, correlate analysis and projection of global and regional life expectancy, healthy life expectancy, and their GAP: 1995-2025. Journal of global health. 10 (2).

[26] A. AYDIN, U. ATİLA, & S. AYDIN (2018), Use of ANN in Predicting Life Expectancy: The Case of Turkey. *Artificial Intelligence Studies*, *1*(1), 1-

[27] Khashei, M., & Bijari, M. (2011), A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied soft computing*, *11*(2), 2664-2675.