Predicting the Future Evolution of TB in Malaysia: Arttificial Neural Networks Approach

[1]**Dr. Smartson. P. NYONI**, [2]**Thabani NYONI**

[1]ZICHIRe Project, University of Zimbabwe, Harare, Zimbabwe
[2]Department of Economics, University of Zimbabwe, Harare, Zimbabwe

Abstract - **In this paper, the ANN approach was applied to analyze TB incidence in Malaysia. The employed annual data covers the period 2000-2018 and the out-of-sample period ranges over the period 2019-2023. The residuals and forecast evaluation criteria (Error, MSE and MAE) of the applied model indicate that the model is stable in forecasting TB incidence in Malaysia. The model suggests that the incidence will drop slightly over the period 2019-2023. In order to contribute meaningfully to the national control strategy of a TB-free Malaysia, the government should, among other things, intensify TB surveillance and control programs.**

Keywords: ANN, Forecasting, TB incidence.

I. INTRODUCTION

Tuberculosis (TB) is the commonest infectious cause of morbidity and mortality globally (Goroh et al, 2020). In 2018, 44% of TB cases were in South East Asia (WHO, 2019) and TB remains a serious threat to public health in developing countries. Malaysia is a South East Asian nation with a high TB disease burden(WHO, 2017).The burden of TB worldwide is aggravated by poverty and malnutrition(WHO, 2018).TB incidence in Malaysia was estimated to be 80 cases per 100 000 population(WHO, 2016).The Malaysian government is committed to the global goal of ending TB by 2035 .The aim of the National TB program is to identify and screen at least 10 contacts for every TB infected patient with a goal to detect sputum positive cases and treat them early to minimize community spread. Latent TB infection treatment is recommended and is offered to under 5 children and immune suppressed patients although implementation is limited (Goroh et al, 2020).The government recommends tracing of TB contacts and appropriate screening tests must be performed which include symptom screening, Tuberculin skin test (TST),with or without chest x-ray and sputum collection from TB contacts (Malaysia Ministry of Health, 2002).The aim of this paper is to model and forecast the annual TB incidence in Malaysia using the artificial neural network, multilayer perceptron (MLP) type ANN(9,12,1) model. The findings of this paper will assist the government of Malaysia to understand the future trends of the TB incidence in the country. This will stimulate an early and appropriate response to the TB epidemic.

II. LITERATURE REVIEW

Abdullah et al (2012) modelled and forecasted TB cases using data covering the period 2003-2010 in Kalantan and forecasted TB cases over the next 2years.The study applied 4 models: Naïve Method, Average forecast, Exponential smoothing and Box Jenkins method. The study noted an increasing trend in TB cases. The double exponential smoothing technique was found to be the best time series model comparing to single exponential smoothing, Holt's Method, ARRES and Holt winter's trend. Another study by Kilicman & Rosian (2020) analyzed TB cases from Jabatan Kesihatan Negeri Terengganu by applying linear trend, quadratic trend, simple moving average, simple exponential smoothing and Holt's trend corrected exponential smoothing. A forecast of TB cases for 2009-2013 suggested that the numbers of TB cases are expected to increase. People aged between 35-44 years old, male, Malay, unemployed, are high risk group of being infected with TB. Goroh et al (2020) investigated the epidemiology of TB in Sabah state, Malaysia for the period 2012-2018.They conducted a retrospective review of TB cases notified in Sabah during the period 2012-2018. The study revealed that between 2012-2018, TB notifications in Sabah state equated to 20 % of Malaysia's total TB notifications. Hotspots of TB were noted in urban population hubs and points of migration as well as evidence of late presentation and diagnosis. Tok et al (2020) assessed the TB treatment outcomes among patients registered in the National TB surveillance database in Malaysia from 2014 to 2017and identified factors associated with unsuccessful treatment outcomes and all-cause mortality. A retrospective cohort study was conducted. TB patients' sociodemographic, clinical data and treatment outcomes at 1 year surveillance were extracted from the database and analyzed using logistic regression. The findings of the research revealed that unsuccessful TB treatment outcomes were associated with older age, males, foreign nationality, urban dwellers, lower education levels, passive detection of TB cases, absence of a BCG

scar, underlying diabetes mellitus, smoking, extra-pulmonary TB, history of previous TB treatment, advanced chest radiological findings, and HIV. Factors associated with all-cause mortality were similar except for nationality(higher for Malaysians) and place of residence(higher among rural dwellers),while smoking and previous TB treatment were not found to be associated with all-cause mortality. Liew et al (2018) conducted a study to determine TB incidence and factors associated with mortality among Healthcare workers in Malaysia registered in 2012 to 2014 with the National TB surveillance Registry. Regression analysis was used. The findings of the study revealed that healthcare workers have double the risk of contracting TB but with lower rates of mortality compared with the general population. Older age, diabetes, HIV infection and not receiving DOTS were associated with TB mortality.

III. METHOD

The Artificial Neural Network (ANN), which we intend to use in this paper; is a data processing system consisting of a large number of simple and highly interconnected processing elements resembling a biological neural system. It has the capability of learning from an experimental or real data set to describe the nonlinear and interaction effects with great accuracy. ANN-based curve fitting technique is one of the extensively applied artificial intelligence methods that are used for forecasting and prediction purpose. It consists of basically three layers i.e., input layer, hidden layer, and output layer, the present work includes the number of years as input layer and the annual TB incidence in Malaysia as output data for the network. In this piece of work, our ANN is based on the hyperbolic tangent function.

Data Issues

This study is based on TB incidences (referred to as M series in this study) in Malaysia. The annual data covers the period 2000-2018 while the out-of-sample forecast covers the period 2019-2023. All the data employed in this research paper was gathered from the World Bank online database.

IV. FINDINGS OF THE STUDY

DESCRIPTIVE STATISTICS

Table 1: Descriptive statistics

Mean	Median	Minimum	Maximum
79.105	75.000	68.000	94.000
Std. Dev.	C.V.	Skewness	Ex. kurtosis
10.187	0.12877	0.41796	-1.5433
5% Perc.	95% Perc.	IQ range	Missing obs.
Undefined	94.000	21.000	0

ANN MODEL SUMMARY FOR TB INCIDENCE (new cases per 100 000 population/year) IN MALAYSIA

Table 2: ANN model summary

Variable	M
Observations	10 (After Adjusting Endpoints)
Neural Network Architecture:	
Input Layer Neurons	9
Hidden Layer Neurons	12
Output Layer Neurons	1
Activation Function	Hyperbolic Tangent Function
Back Propagation Learning:	
Learning Rate	0.005
Momentum	0.05

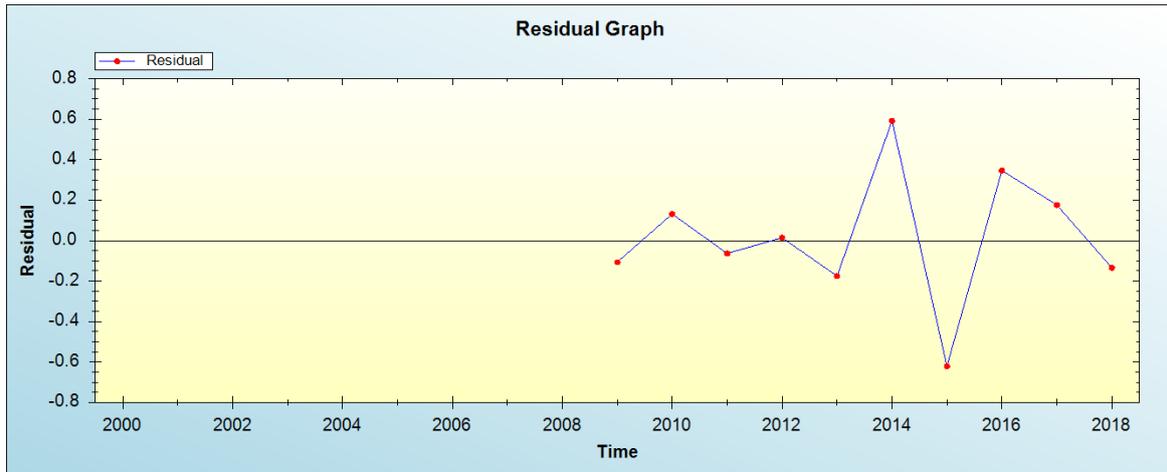*Residual Analysis for the ANN model*

Figure 1: Residual analysis
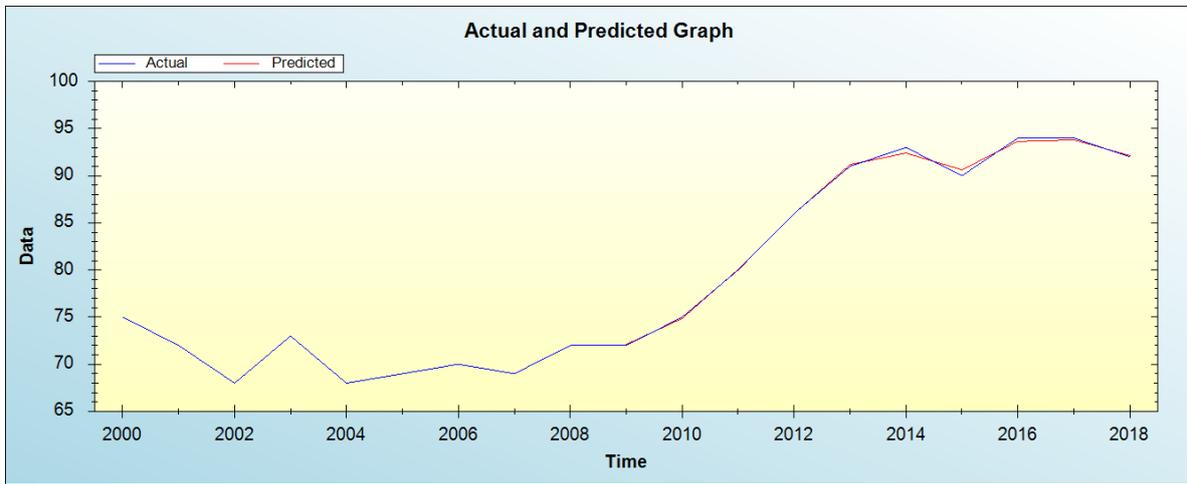
In-sample Forecast for M

Figure 2: In-sample forecast for the M series

Figure 2 shows the in-sample forecast for M series.

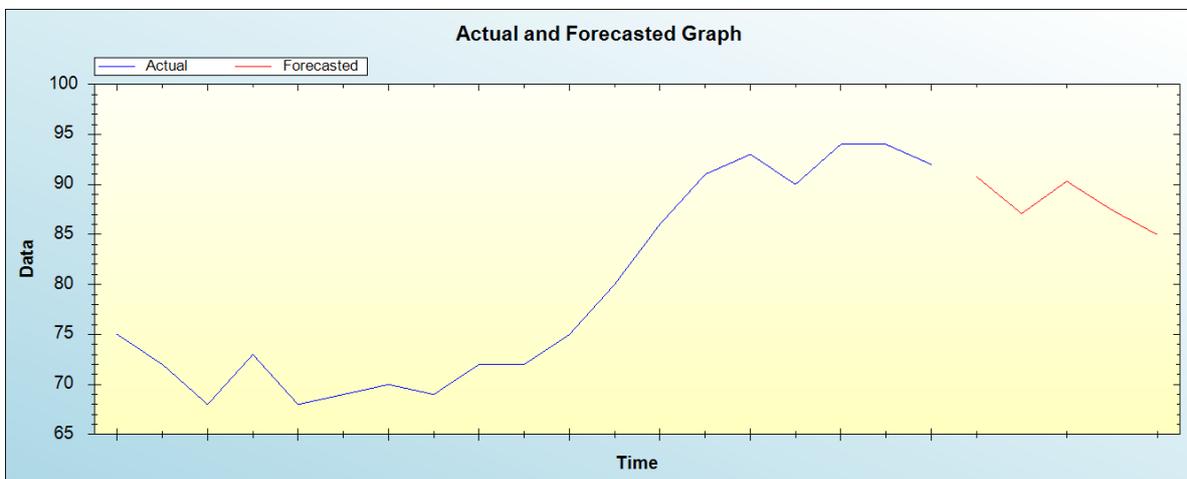*Out-of-Sample Forecast for M: Actual and Forecasted Graph*

Figure 3: Out-of-sample forecast for M: actual and forecasted graph

Out-of-Sample Forecast for M: Forecasts only

Table 3: Tabulated out-of-sample forecasts

Year	Forecasts
2019	90.7777
2020	87.0802
2021	90.3034
2022	87.4316
2023	84.9709

Over the study period 2000-18 the incidence of TB has been rising gradually. Table 1 shows that the minimum and maximum incidence was 68 and 94 cases per 100 000 population/year respectively. The average incidence was 79 cases per 100 000 population per year. The applied data collected from the World Bank online database is positively skewed with an excess kurtosis of -1.5433, meaning that the data is not normally distributed. The residual graph and evaluation statistics (ERROR, MSE, MAE) indicate that the applied ANN (9,12,1) model is stable and suitable for forecasting TB incidence in Malaysia. In-sample forecasts reveal that the model simulates the observed data very well. In the out of sample period 2019-2023 the model predicts that TB incidence will decline from around 91(1n 2019) to approximately 85 cases per 100 000 population/year in 2023.

V. CONCLUSION & RECOMMENDATIONS

Malaysia recorded a general increase in the incidence of TB over the period 2000-2018.This essentially means that the government has to do a lot of ground work to reverse this undesirable trend. The model predicts a small decline in the incidence of TB over the period 2019-2023.The results mean that the government has to intensify TB surveillance and control programs. The health authorities should advocate for increased funding towards TB/HIV programs in order to significantly reduce the incidence of TB.

REFERENCES

[1] Michelle May D. Goroh., Christel H.A. van den Boogaard., Mohd Yusof Ibrahim., Naing Oo Tha., Swe., Freddie Robinson., Khamisah Awang Lukman., Mohammad Saree Jeree., Timothy William., Anna P& Ralph (2020) Factors Affecting Continued Participation in Tuberculosis Contact Investigation in a Low-Income, High-Burden Setting. Trop. Med. Infect. Dis, 5, 124; doi:10.3390/tropicalmed5030124

[2] Ministry of Health Malaysia (2002). Manual System Maklumat Tibi Kebangsaan. In National Tuberculosis Information System [TBIS]Manual; National Information System Development Committee Dcd, Ed.; Ministry of Health Malaysia: Putrajaya, Malaysia, 20

[3] WHO (19). Global Tuberculosis Report 2019; WHO Press: Geneva, Switzerland, 2017.

[4] WHO (2016). Global Tuberculosis Report 2016; WHO Press: Geneva, Switzerland.

[5] WHO (2017). Global Tuberculosis Control Report 2017.Geneva Switzerland

[6] WHO (218). Global Tuberculosis Report 2018; WHO Press: Geneva, Switzerland, 2018.

****