Health Science Monitor
2022; 1(1): 64-73
Published online (http://hsm.umsu.ac.ir)

Original Article

Modeling the survival of patients with tuberculosis based on the model of artificial neural networks

Mehdi Kazempour Dizaji [1,2*], Majid Marjani[3], Payam Tabarsi [3], Mohammad Varahram [1], Ali Zare [2], Mohammad Ali Emamhadi [4], Rahim Roozbahani [3], Atefe Abedini [5], Parvaneh Baghaei Shiva [3], Afshin Moniri [3], Mohammadreza Madani [5]

[1] *Mycobacteriology Research Center (MRC), National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[2] *Department of Biostatistics, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[3] *Clinical Tuberculosis and Epidemiology Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[4] *Department Forensic Medicine, School of Medicine Shahid Beheshti University of Medical Sciences, Tehran, Iran*

[5] *Chronic Respiratory Diseases Research Center, National Research Institute of Tuberculosis and Lung Disease (NRITLD), Shahid Beheshti University of Medical Sciences, Tehran, Iran*

***Corresponding author:** Mehdi Kazempour Dizaji, **Address**: Mycobacteriology Research Center (MRC) Shahid Beheshti University of Medical Sciences, Tehran, Iran, **Email**: mkazempour@sbmu.ac.ir, **Tel**: +982127122080

Abstract

***Background & Aims*:** The development of treatment methods and increasing the survival of patients with tuberculosis (TB) has led to the complication of relationships between independent and dependent variables associated with this disease. Therefore, it is important to use new methods to model the TB process that can accurately estimate the current situation. This study aimed to model the survival of patients with tuberculosis based on the model of perceptron artificial multilayer neural network (MLP-ANN).

***Materials and Methods*:** In this retrospective cohort study, the data was collected from 2366 TB patients who were treated in Dr. Masih Daneshvari Hospital in Tehran from 2005 to 2015. To model the predictive power of survival in TB patients, an MLP-ANN model consisting of three layers was applied.

***Results*:** The results of this study showed that based on the MLP-ANN model, the correct predictive power of survival in TB patients is 88.4%. In this study, the variables of patients' age and family size as very effective variables also variables of patients' gender, marital status, education, adverse drug effects, exposure to cigarette smoke, imprisonment, pulmonary tuberculosis, and AIDS as effective variables in predicting the survival of patients were diagnosed.

***Conclusion*:** In the model of artificial neural networks, no restrictions are considered for the data structure and the type of relationship between variables. Therefore, these models with their flexibility and high accuracy can be one of the best methods for modeling health data.

***Keywords*:** Perceptron artificial neural network, Survival, Tuberculosis, Modeling

Received 15 July 2022; accepted for publication 24 August 2022

Introduction

Today, due to the advancement of technology and the complexity of modern life, the use of novel methods has replaced many traditional methods that are no longer able to accurately estimate the current situation (1). The complexity of the disease process has also led to nonlinear relationships between disease-related variables, which have made it difficult to study predisposing factors and ultimately decide on treatment methods (2). For instance, in the case of a chronic diseases process such as tuberculosis(TB); the advancement of medical sciences and the application of more appropriate treatment strategies, in addition to improving the quality of life of these patients and increasing their survival, has led to the complication of relationships between independent and dependent variables associated with TB disease (3-4). This issue has somewhat complicated the modeling of the disease process.

The main purpose of modeling methods is to determine the relationships between independent and dependent variables about the desired phenomenon. Choosing the right method for modeling and analyzing health data is very important and sensitive (5). For example, in patients with TB, survival of patients after treatment is very important, and modeling the survival of these patients can guide researchers on the quality of treatment methods. But the choice of modeling method for such data depends on the limitations and assumptions that are considered as statistical model defaults. One of the methods used to study survival modeling in patients with tuberculosis is the Cox regression model (6).

In this model, it is necessary to establish assumptions such as proportional hazards, independence of event times, and linearity of the relationship between independent variables and dependent variables (4,7,8). On the other hand, the sensitivity of these models to missing value, outlier data, and heavy censoring data are other limitations of these methods (9). However, if the data are complex, these assumptions may not be established and may lead to a lack of goodness of fit in modeling and misunderstanding of the TB disease process (5). Therefore, the use of modeling methods that do not depend on these assumptions will be considered.

One of the modeling methods designed to solve such problems is the model of artificial neural networks (ANNs), the use of which has increased in recent decades. The ANNs models, because they do not require any assumptions and also have a lot of flexibility, are widely used in various fields of medical sciences, including disease prediction, diagnosis, and modeling of factors affecting events related to chronic diseases. An ANN model in a way that is inspired by the process of learning and processing information, has the ability of very complex analyzes (1-2). The smallest processing unit of an ANN is the neuron. These neurons are present in different layers of the artificial neural network, such as the input, hidden, and output layers, and are connected by synaptic weights (1-2,10-12).

One of the most widely used artificial neural networks is the multilayer perceptron. This network is a kind of ANN that is based on a computational unit called perceptron. The most common type of this artificial neural network is the three-layer perceptron, which is shown in Figure 1.

Artificial neural networks are implemented in two phases training and testing. In the training phase, the network learns and evaluates the relationship between independent and dependent variables by receiving a random sample of data, which is usually more than 50% of the information. After completing the training phase, the network performance in the training phase and its predictive power will be tested with the remaining observations. The ANNs model can be one of the best methods for modeling health data that not only do not impose an initial assumption on the data but also does not place any restrictions on the type of relationship between independent and dependent variables (10,11.13). Therefore, this study was designed to model the survival of patients with TB based on the ANN model as a flexible model and implemented among TB patients who were referred to Dr. Masih Daneshvari Hospital in Tehran.

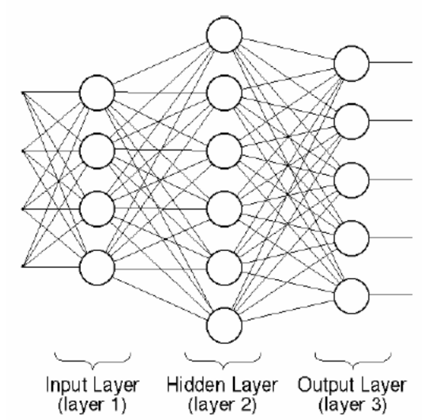

Fig 1. Schematic of a three-layer artificial perceptron neural network

Materials & Methods

A retrospective cohort study was carried out in Dr. Masih Daneshvari Hospital, Tehran, Iran. Data for this study were obtained from the medical records of 2366 TB patients registered in the national research institute of tuberculosis and lung disease of this hospital from 2005 to 2015. Specific death from TB was considered for those patients whose cause of death was TB. For all patients of this study, variables of gender, age, marital status, educational level, residency area, nationality, family size, adverse drug effects, smoking, passives smoker, history of drug use, exposure to a patient with TB, imprisonment, pulmonary TB, extra-pulmonary TB, diabetic mellitus, HIV positive and comorbidities as independent variables and patients' survival time and death status were studied as dependent variables in the ANN model. For dead patients, the cause of death and time of death was determined using information from their medical records and through telephone interviews with close relatives of the patients. Those patients whose status is alive until the end of the study and after that no information is available about them, as well as patients whose information is complete and does not indicate death until the end of the study, were considered censored.

A three-layer perceptron ANN was employed to model the survival of TB patients. In this 3- layer ANN model, the first layer is the input layer, the second is a hidden layer, and the third is the output layer. According to the independent variables, 39 neurons were considered for the input layer and according to the dependent variables, 3 neurons were considered for the output layer. The appropriate number of hidden or middle layer neurons was also determined based on the best structure for the neural network when analyzing the data.

The effect of variables on predicting patient survival in the ANN model was presented by the importance index and normalized importance. All analyses were performed using SPSS 19 software.

Results

According to the results of this study, out of 2366 TB patients, death occurred for 723 patients (30.6%) as the desired event after treatment and (69.4%) 1643 patients were considered as censored (being alive at the end of the study or lack of information about their survival status). To model the survival of TB patients, a set of demographic and clinical information of these patients were examined as independent variables and survival time and status of death as dependent variables (Table 1).

Table 1. Independent and dependent variables of tuberculosis patients and number of neuron for each variable in artificial neural networks model

Role of variables	Variables	Category	Number of neuron
Dependent	Survival time	-	1
Dependent	Death status	Death Censored	2
Independent	Gender	Female Male	2
Independent	Age	-	1
Independent	Marital Status	Single Married Widow Divorced	4
Independent	Education	Illiterate Primary Secondary High School Higher education	5
Independent	Residency area	Rural Urban	2
Independent	Nationality	Iranian Non-Iranian	2
Independent	Family size	-	1
Independent	Adverse effect	No Yes	2
Independent	Smoker	No Yes	2
Independent	Passives smoker	No Yes	2
Independent	Drug user	No Yes	2
Independent	TB contact	No Yes	2
Independent	Imprisoned	No Yes	2
Independent	Pulmonary TB	No Yes	2
Independent	Extra-Pulmonary	No Yes	2

Independent	Diabetic Mellitus	No Yes	2
Independent	HIV Positive	No Yes	2
Independent	Co morbidities	No Yes	2

For modeling based on 3- layer perceptron ANN, 70% of patients were randomly included in the training phase and 30% of patients in the testing phase. Also, to determine the best structure of the perceptron neural network model and to specify the number of neurons suitable for the hidden layer, a set of perceptron neural network models with 6 to 25 neurons for the hidden layer was examined based on the training phase (Table 2). These models were compared using indices such as the sum of square error, percent incorrect predictions, and average overall relative error. Perceptron neural network model with 39 neurons for the input layer, 20 neurons for the hidden layer, and 3 neurons for the output layer with the least sum of square error, average overall relative error, and percent incorrect predictions among other models as the best structure for the network neural perceptron was selected.

Table 2. Best structure for Artificial Neural networks model based on varing the number of neruns in hidden layer

Artificial neural networks structure (Input/Hidden/Output)	Sum of squares error	Percent incorrect predictions	Average overall relative error
39/6/3	48.00	13.1	0.90
39/7/3	49.31	13.4	0.93
39/8/3	47.63	12.9	0.89
39/9/3	46.28	12.6	0.87
39/10/3	48.59	13.4	0.91
39/11/3	48.73	12.6	0.91
39/12/3	47.81	12.4	0.90
39/13/3	45.38	12.1	0.85
39/14/3	46.14	12.9	0.86
39/15/3	45.25	12.1	0.85
39/16/3	46.05	11.9	0.86
39/17/3	43.98	12.1	0.824
39/18/3	46.62	12.9	0.87
39/19/3	44.08	11.9	0.83
39/20/3*	43.76	11.6	0.82
39/21/3	45.90	11.9	0.86
39/22/3	47.56	13.4	0.89
39/23/3	45.75	12.6	0.86
39/24/3	47.66	13.1	0.89
39/25/3	47.97	13.1	0.90

Results are based on training sample.

Based on the results of this model, the correct predictive power of survival in TB patients for both the training and testing phase was the same and equal to 88.4%. This model had only an 11.6% error in predicting patient survival based on the training and testing phase, in other words, it can correctly predict the survival of 7 out of every 8 patients with tuberculosis. A scale based on the normalized importance index was used to evaluate the importance of independent variables in predicting the survival of TB patients. In this study, variables with a normalized importance index above 60% as highly effective variables, variables with a normalized importance index between 20 to 60% as effective variables, and variables with normalized importance less than 20% as low-impact variables were recognized.

In this study, the results of the ANN model in selecting important variables in predicting patient survival showed that the variables of patient age and family size were very effective, while patient gender, marital status, educational level, adverse drug effects, smoke exposure, imprisonment, pulmonary TB and AIDS were as effective variables and patient's nationality variables, place of residence, smoking, history of drug use, history of contact with a person with TB, extrapulmonary TB, diabetes, and comorbidities had the little effect on predicting the survival of patients with TB (Table 3). Also, to better understand the order of importance of the variables from the most to the least effect in predicting the survival of TB patients based on normalized importance and importance indices, it has been graphically marked (Figure 2).

Table 3. Importance of independent variables based on artificial neural networks model

Variables	Importance	Normalized Importance%
Gender	0.04	22.6
Age	0.20	100
Marital Status	0.06	30.4
Education	0.07	36
Residency area	0.02	9
Nationality	0.02	9
Family size	0.17	88.5
Adverse effect	0.09	48.1
Smoker	0.02	9.4
Passives smoker	0.02	22.4
Drug user	1.07	8.9
TB contact	0.02	12.8
Imprisoned	0.05	27.5
Pulmonary TB	0.06	31
Extra-Pulmonary	0.02	12.4
Diabetic Mellitus	0.03	13.4
HIV Positive	0.04	20
Co morbidities	0.02	7.6

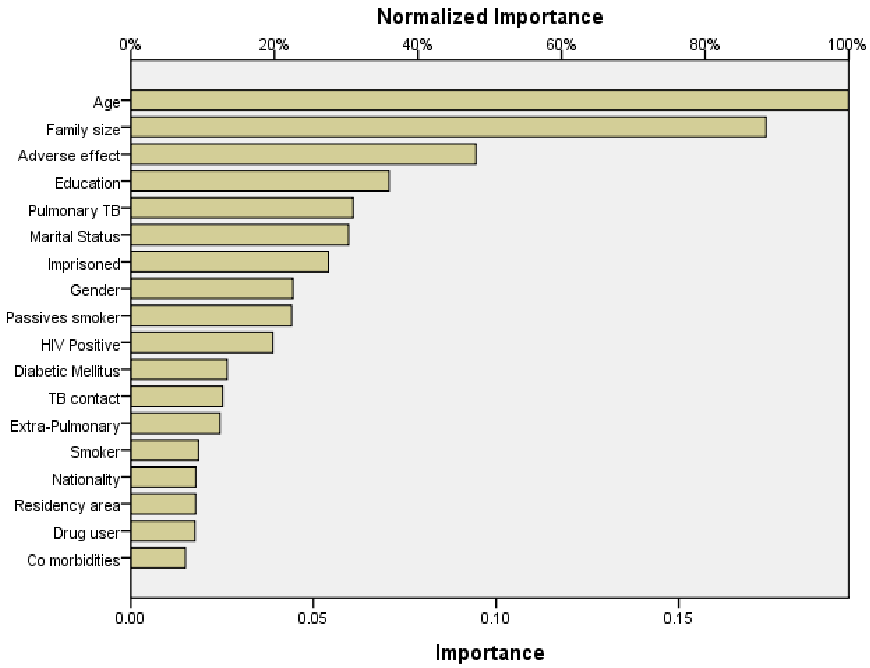

Fig 2. The effect of independent variables based on normalized importance and importance index.

Discussion

In chronic diseases such as tuberculosis, the survival study of these patients is very important and in some cases sensitive. Because by increasing the survival of patients with tuberculosis, the success of treatment methods can be optimistic and the quality of these methods can be evaluated. This requires accurate identification of factors affecting the survival of TB patients and modeling the survival of these patients.

Numerous studies have been performed to model the survival of patients with tuberculosis, but these models used methods that require basic assumptions for both the statistical method and the data structure. The inconsistency of these hypotheses as well as the complexity of the data structure will deprive researchers of the possibility of accurately identifying the factors affecting the survival of patients with tuberculosis as well as proper modeling of the survival of these patients. This can lead to incorrect decisions about the quality of treatment and incorrect prediction of patient survival (7,9).

Overall, in studies on TB, both in the diagnosis and treatment of the disease, also in evaluating the effectiveness of treatment strategies, and in modeling the events that occur during treatment for TB patients, such as multidrug-resistant tuberculosis (MDR-TB), disease relapse, and death, all were designed and performed based on traditional statistical methods. In most of these studies, no attention had been paid to the appropriateness of data structure and establishment of statistical model assumptions. Therefore, there is a need for a model with high flexibility and accuracy in predicting the data structure in the study of the tuberculosis process (4,5, 14-27).

Hence, this study has been designed and implemented to consider new methods for analyzing TB data (7,9.28,29). The present study was designed and performed with a three-layer perceptron neural network model with 39 neurons in the input layer, 20 neurons in the hidden layer, and 3 neurons in the output layer to model patient survival. Based on the results of this model, the accuracy of predicting the survival of TB patients was estimated to be 88.4%, which was relatively high accuracy in predicting the survival of patients. To better understand the predictive power of this model, it should be noted that out of every 8 TB patients, this model was able to accurately predict the survival of 7 patients. Various studies have also shown that neural networks have higher predictive power compared to a range of statistical methods (1,2,10,11,28,30).

In the current study, in addition to determining the effect of independent variables on predicting the survival of TB patients, their importance order was also graphically examined based on normalized importance and importance indices. Presenting the order of significance of effective variables on the survival of TB patients is one of the advantages of using ANN in the study of patient survival. This is especially important in policy-making for health promotion and making prevention (12,13,28,31). In this study, the variables of patient age and family size with normalized importance index of 100 and 88.5%, respectively, showed the greatest effect in predicting the survival of TB patients. Patient gender, marital status, educational level, adverse drug effects, exposure to secondhand smoke, imprisonment, pulmonary TB and AIDS with normalized importance index between 20 to 60% were diagnosed as effective variables in predicting patient the survival of TB patients. These results are consistent with other studies on the survival of TB patients (4,5,32).

Some studies have identified the variable of contact with a TB patient as a variable affecting patient survival, but in the present study, this variable with a normalized importance index of less than 20% was considered one of the variables with less effect than other variables. However, due to the high accuracy of the ANN model in predicting the survival of TB patients and the probability that the assumptions of the statistical models used were not met in other studies, these results can be justified. In the study of many chronic diseases such as tuberculosis, ANN models can be a good alternative to a wide range of statistical methods common due to their high predictive accuracy and great flexibility in data structure.

Conclusion

The ANN model is more flexible than other TB data analysis methods and due to its insensitivity to data structure has a lot of high predictive power, even with complex data structures such as missing data, outlier data, and heavy censoring data. For that reason, the use of these models in health and medical research is important and can illuminate many dark angles for researchers.

Acknowledgments

We appreciate the close cooperation of the staff and officials in this study.

Conflict of interest

The authors have no conflict of interest in this study.

References

1. Sargent DJ. Comparison of artificial neural networks with other statistical approaches: Results from medical data sets. Cancer. 2001 Apr 15;91(S8):1636–42.
2. Ahmed FE. Artificial neural networks for diagnosis and survival prediction in colon cancer. Mol Cancer. 2005 Dec;4(1):29.
3. Miller TL, Wilson FA, Pang JW, Beavers S, Hoger S, Sharnprapai S, Pagaoa M, Katz DJ, Weis SE. Mortality hazard and survival after tuberculosis treatment. American journal of public health. 2015 May;105(5):930-7.
4. Dizaji MK, Kazemnejad A, Tabarsi P, Zayeri F. Risk factors associated with survival of pulmonary tuberculosis. Iranian Journal of Public Health. 2018 Jul;47(7):980.
5. Kazempour-Dizaji M, Kazemnejad A, Tabarsi P, Zayeri F. Estimation of Ten-Year survival of patients with pulmonary tuberculosis based on the competing risks model in Iran. Tanaffos. 2016;15(1):37.
6. Kleinbaum D, Klein M. Survival Analysis: A self-learning text, 2005. New York, NY: Spring Science+ Business Media. 2011.

7. Kleinbaum DG, Klein M. Evaluating the Proportional Hazards Assumption. In: Survival Analysis [Internet]. New York, NY: Springer New York; 2012 [cited 2022 Sep 4]. p. 161–200. (Statistics for Biology and Health). Available from: http://link.springer.com/10.1007/978-1-4419-6646-9_4
8. Collett D. Modelling survival data in medical research. CRC press; 2015 May 4.
9. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. New York: Springer; 2003 Feb.
10. Akl A, Ismail AM, Ghoneim M. Prediction of Graft Survival of Living-Donor Kidney Transplantation: Nomograms or Artificial Neural Networks? Transplantation. 2008 Nov 27;86(10):1401–6.
11. Chi CL, Street WN, Wolberg WH. Application of artificial neural network-based survival analysis on two breast cancer datasets. InAMIA annual symposium proceedings 2007 (Vol. 2007, p. 130). American Medical Informatics Association.
12. LaFaro RJ, Pothula S, Kubal KP, Inchiosa ME, Pothula VM, Yuan SC, et al. Neural Network Prediction of ICU Length of Stay Following Cardiac Surgery Based on Pre-Incision Variables. PLoS ONE. 2015 Dec 28;10(12):e0145395.
13. Yegnanarayana B. Artificial neural networks. 2009: PHI Learning Pvt.
14. Glaziou P, Sismanidis C, Floyd K, Raviglione M. Global Epidemiology of Tuberculosis. Cold Spring Harbor Perspectives in Medicine. 2015 Feb 1;5(2):a017798–a017798.
15. Tabarsi P, Mohammadtaheri Z, Chitsaz E, Farokhzad B, Hadavand F, Gachkar L, et al. Performance of QuantiFERON-TB Gold test compared to tuberculin skin test in detecting latent tuberculosis infection in HIV-positive individuals in Iran. Ann Thorac Med. 2010;5(1):43.
16. Tabarsi P, Chitsaz E, Moradi A, Baghaei P, Farnia P, Marjani M, et al. First-line antituberculosis drug resistance prevalence and its pattern among HIV-infected patients in the national referral tuberculosis centre, Iran. Int J STD AIDS. 2009 Aug;20(8):566–70.
17. Mirsaeidi SM, Tabarsi P, Khoshnood K, Pooramiri MV, Rowhani-Rahbar A, Mansoori SD, et al. Treatment of multiple drug-resistant tuberculosis (MDR-TB) in Iran. International Journal of Infectious Diseases. 2005 Nov;9(6):317–22.
18. Tabarsi P, Chitsaz E, Moradi A, Baghaei P, Farnia P, Marjani M, et al. Treatment outcome, mortality and their predictors among HIV-associated tuberculosis patients. Int J STD AIDS. 2012 Sep;23(9):e1–4.
19. Albuquerque M de FPM de, Batista J d'Arc L, Ximenes RA de A, Carvalho MS, Diniz GTN, Rodrigues LC. Risk factors associated with death in patients who initiate treatment for tuberculosis after two different follow-up periods. Rev bras epidemiol. 2009 Dec;12(4):513–22.
20. Saraceni V, Durovni B, Cavalcante SC, Cohn S, Pacheco AG, Moulton LH, et al. Survival of HIV patients with tuberculosis started on simultaneous or deferred HAART in the THRio cohort, Rio de Janeiro, Brazil. The Brazilian Journal of Infectious Diseases. 2014 Sep;18(5):491–5.
21. Oursler KK, Moore RD, Bishai WR, Harrington SM, Pope DS, Chaisson RE. Survival of Patients with Pulmonary Tuberculosis: Clinical and Molecular Epidemiologic Factors. Clinical Infectious Diseases. 2002 Mar 15;34(6):752–9.
22. Leprosy A.K.S.O. Tuberculosis situation in Iran 2015, Infectious Diseases Management Center - Ministry of Health, Treatment and Medical Education.
23. Vasantha M, Gopi PG, Subramani R. Survival of tuberculosis patients treated under DOTS in a rural Tuberculosis Unit (TU), south India. Indian Journal of Tuberculosis. 2008;55(2):64-9.
24. Akessa GM, Tadesse M, Abebe G. Survival Analysis of Loss to Follow-Up Treatment among Tuberculosis Patients at Jimma University Specialized Hospital, Jimma, Southwest Ethiopia. International Journal of Statistical Mechanics. 2015 Dec 31;2015:1–7.
25. Wu YC, Lo HY, Yang SL, Chou P. Factors correlated with tuberculosis reported after death. int j tuberc lung dis. 2014 Dec 1;18(12):1485–90.
26. Wu YC, Lo HY, Yang SL, Chu DC, Chou P. Comparing the Factors Correlated with Tuberculosis-Specific and Non-Tuberculosis-Specific Deaths in Different Age Groups among Tuberculosis-Related Deaths in Taiwan.

Hozbor DF, editor. PLoS ONE. 2015 Mar 3;10(3):e0118929

27. Farley JE, Ram M, Pan W, Waldman S, Cassell GH, Chaisson RE, et al. Outcomes of Multi-Drug Resistant Tuberculosis (MDR-TB) among a Cohort of South African Patients with High HIV Prevalence. Goletti D, editor. PLoS ONE. 2011 Jul 22;6(7):e20436.
28. Kay JW, Titterington DM. Statistics and neural networks: advances at the interface. Oxford University Press on Demand; 1999.
29. Kwon YS, Kim YH, Song JU, Jeon K, Song J, Ryu YJ, et al. Risk Factors for Death during Pulmonary Tuberculosis Treatment in Korea: A Multicenter Retrospective Cohort Study. J Korean Med Sci. 2014;29(9):1226.
30. Warner B, Misra M. Understanding Neural Networks as Statistical Tools. The American Statistician. 1996 Nov;50(4):284–93.
31.. Tsai PF (Jennifer), Chen PC, Chen YY, Song HY, Lin HM, Lin FM, et al. Length of Hospital Stay Prediction at the Admission Stage for Cardiology Patients Using Artificial Neural Network. Journal of Healthcare Engineering. 2016;2016:1–11.
32. Organization, W.H., Tuberculosis Fact sheet N°104. 2016.