Prediction of workers pulmonary disorder exposed to silica dust in stone crushing workshops using logistic regression and artificial neural networks techniques

Maryam Farhadian[1], Hossien Mahjub[2*], Mohsen Aliabadi [3], Saeed Musavi [4], Mehdi Jalali [5]

1,4-Department of Biostatistics, School of public Health, Hamadan University of Medical Sciences, Hamadan, Iran.
2-Department of Biostatistics & Epidemiology and Research Centre for Health Sciences, School of public Health, Hamadan University of Medical Sciences, Hamadan, Iran.
3,5-Department of Occupational Health, School of public Health, Hamadan University of Medical Sciences, Hamadan, Iran.

*Corresponding author:
Hossien Mahjub, Professor Department of Biostatistics & Epidemiology, Centre of Health Researches, School of public Health, Hamadan University of Medical Science, Hamadan, Iran.
Tel: +988118381644
Email: Mahjub@umsha.ac.ir

Abstract

Introduction: The work exposure conditions such as dust concentration, exposure time, use of respiratory protection devices and smoking status are effective to cause pulmonary function disorder. The objective of this study was prediction of pulmonary disorders in workers exposed to silica dust using artificial neural networks and logistic regression.

Methods and Materials: A sample of 117 out of 150 workers employed in the stone crushing workshops placed in Hamadan province, in the west of Iran, was selected based on simple random approach. Information about occupational exposure histories were collected using a questionnaire. To assess the pulmonary disorder status in the workers exposed to silica dust based on the spirometry indices as well as the workers characteristics the prediction models of artificial neural networks and logistic regression were employed using the SPSS software version 16.

Results: Measurements of pulmonary function indices of the studied workers showed that the indices for workers having pulmonary disorder versus the others were statistically significant (P <0.01). The results of the obtained models showed that the artificial neural networks and the logistic regression had a high performance for prediction of pulmonary disorder status. However, the developed neural networks model had a better performance than the logistic regression model in viewpoint of sensitivity, specificity, kappa statistic and the area under ROC curve.

Conclusions: The neural networks prediction model was more accurate compared with the logistic regression. In this regards, the developed prediction model can be used as a helpful tool and guideline by occupational health experts for evaluating workers exposure conditions and determining the health priorities and control measures in the stone crushing workshops.

Keywords: Logistic Regression, Neural Networks, Pulmonary, Prediction, Sensitivity, Specificity.

Introduction

Occupational diseases caused by long term exposure to hazardous agents, are mostly preventable and none-curable (1).Accordingly, in order to implement occupational health programs and interventions in workplaces, the risk of occupational disorders should be anticipated. Silicosis as a most important work-related lung disease is caused by the inhalation of silica dust. Reduction of the lung function capacity will gradually appear mainly affected by silica dust concentration, exposure time, status of using respiratory protective equipment and smoking (1). In the stone crushing process, the workers have exposure to abundant silica-containing dust (2, 3). The study was conducted in Hamadan stone crushing workshops, has shown that the concentration of particles in workplaces air are higher than recommended limit (4).

In several studies respiratory performance test was performed to determine pulmonary disorders and also pulmonary function parameters including forced vital capacity (FVC), forced expiratory volume timed (FEV_T) and forced expiratory volume in one second fraction (FEV_1). To determine the influence of variables such as weight, height, work experience, history of smoking and education level on lung function indices, multiple regression and logistic techniques were used (5, 6, 7, 8, 9). Despite the common use of logistic regression technique, artificial neural network (ANNs) is a new technique for predicting complex phenomena, already widely used in basic science, medicine and engineering (10). These networks can be called electronic models of human brains neural structure, which have innate talent to store and to apply empirical data. The ability to determine the complex relations among different variables and the capacity of using multiple learning algorithms are its features (10).

Despite the increasing use of neural network techniques as a new approach in various fields, few studies about occupational medicine have used them. In the field of occupational medicine, this technique has been used to identify and classify job pneumoconiosis outbreak based on the job history (11). The relations between lung function parameters and blood factors in smoker welders were also analyzed using neural network technique (12). As noted, in the incidence of work-related lung disease, several occupational and none-occupational factors are effective that represent complexity of incidence model of this phenomenon. Due to the importance of preventing work-related pulmonary disorders, use of appropriate techniques for predicting lung disorders in exposed workers is necessary. Accordingly, this study aims to predict the incidence of pulmonary complications in workers exposed to silica dust in terms of effective variables including occupational exposure conditions using artificial neural network. Logistic regression technique was also employed and its results were compared with those of neural networks.

Methods and Materials

In this descriptive cross-sectional study, 117 of 150 permanent employees exposed frequently to silica dust at different workstations in Hamadan stone crashing workshops, in the west of Iran, were selected randomly. The results of lung function test were extracted based on the instruction of the American Thoracic Society using spirometer (Vitalograph, model Spirotrav Iv) (13). A questionnaire was employed to assess the individual characteristics, status of occupational exposure of workers and pattern of behaviors including smoking and using status of respiratory protective equipment.

These questionnaires were completed through interview and observing the file of annual periodic examination of workers in consecutive years. For developing the prediction model of pulmonary disorder, definition of desired event included presence or absence of pulmonary complication in workers was conducted according to the standard criteria based on lung function parameters (14). To predict the status of pulmonary disorders, artificial neural network and logistic regression techniques were employed.

Artificial neural networks consist of a number of interconnected neurons which are positioned in at least three layers, i.e. one input layer of source neurons, at least one hidden layer and an output layer of computational neurons. Each neuron in the hidden layer of the networks can generally be considered as a simple processing element taking one or more input(s) and giving one and more output(s). In each neuron, a related weight is assigned to an input through which the power of each input is adjusted. The neuron can, thereby, combine all the inputs and calculates an output, which is passed on. One can easily propose the basic equations to express the structure of typical artificial neural networks as follows.

$$y_i = F(z_i)$$

(1)

$$z_i = \sum_{i=1}^{n} w_i x_i + b_i$$

(2)

$$\sum_{i=1}^{n} w_i x_i = w_1 x_1 + w_2 x_2 + w_3 x_3 + \ldots\ldots + w_n x_n$$

(3)

Where x_1, x_2, and x_n denote the measured value for input variables, w_1, w_2, and w_n are the weights, b_i is the bias, y_i is output variable while Z_i could be considered as any used activation function.

Nonlinearity can be considered as the main activation function of neural networks, the most typical of which is sigmoid function used as an activating one. Multilayer perceptron, used in the present study, can be regarded as the most widely accepted structure of neural networks, applied to model physical phenomena. Regarding the interconnections within the neural networks structure, feed forward networks, were selected. In this study, training artificial neural networks was performed based on the supervised learning method using error back propagation learning algorithm. Also, each variable was normalized by subtracting from its mean value and then dividing the result by its standard deviation so as to have zero mean value and unity variance for all variables (10). In this study, 70% of data were used for training the network and the remaining 30% for testing validity of the neural networks.

Logistic regression technique in the study has relationship as follows.

$$\log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{p(y=1)}{p(y=0)}\right) = \alpha + \sum_{i=1}^{p} \beta_i x_i$$

Where π, is the probability of observation to the first category of the response variable, X_i, is independent or predictor variable and β_i, is the regression coefficient.

The other advantages of the logistic regression model are possibility to predict every observation in relation to each response variable levels and also feasibility to direct calculation of odds ratios using coefficients of model (15). In the prediction models, response variable was pulmonary disorder status whereas age, weight, height, experience, smoking status, duration of smoking, number of smoked cigarettes and using status of respiratory protective equipment were considered as predictor variables. The neural networks model for predicting the pulmonary disorder was trained and validated using the SSPS 16 software. The activation functions in the hidden and output layer were considered as sigmoid type. Different structures of neural networks were developed and the best

networks were chosen using trial and error in terms of the number of hidden layers and their neurons. Finally, neural networks with two hidden layers along with eleven neurons were considered to be as the best structure of prediction model that is shown in Fig 1. Type of training was selected as online and error function was considered as the error sum of squares. Also, the logistic regression model with the same number of variables for predicting the target variable was fitted using SPSS 16 software.

To compare the performance of logistic regression and neural network models to predict the status of pulmonary disorder, the area under the receiver operating characteristic (ROC) curve was used as an evaluation criterion of predicting models. This curve shows the sensitivity versus one minus the specificity. The area under the curve is between zero and one. Whenever, the value is closer to one, the ability of the prediction model is greater. Note that, the results of the logistic regression model was investigated based on different cutoff point, and the final result has been reported, based on the best specified cutoff point value ($\pi = 0.37$).

Results

The results of descriptive statistics of demographic characteristics and work experience of workers employed in stone crushing workshops, showed that the mean and standard deviation (SD) of workers age was equal to 29.9 ± 9.5 years, weight was equal to 68.5 ± 10.6 kg and length was equal to 174.6 ± 7.2 cm. Also, the mean and standard deviation of work experience was 3.3 ± 2.1 years, smoking history was 2.22 ± 6.35 years and daily number of consumed cigarettes was equal to 1.54 ± 3.73 cigarettes. The results showed that approximately 23% of the studied workers were smokers and 57% of the workers were not using protective equipments. The results of pulmonary function indices of workers based on presence or absence of pulmonary disorders are shown in Table 1. Accordingly, it was observed, that the difference between pulmonary function indices in workers with pulmonary disorders compared to healthy workers was statistically significant. The results of logistic regression model were presented in Table 2. The results of pulmonary disorder prediction among workers using neural networks and logistic regression models were presented in Table 3 and the Results of the performance of neural network compared to logistic regression techniques in the terms of sensitivity, specificity, area under the ROC curve and Kappa statistics criteria were presented in Table 4.

Table 1: The results of pulmonary function indices in workers employed in stone crushing workshops

Index	with pulmonary disorder			Without pulmonary disorder			P-value
	Mean and SD(%)	Min(%)	Max(%)	Mean and SD(%)	Min(%)	Max(%)	
FVC	80.84±8.26	64.3	96.5	97.61±11	81.3	137.2	< 0.01
FEV$_1$	71.78±6.12	59.6	88	94.38±10.24	80.5	123.5	< 0.01
FEV$_1$%	89±10.13	71.6	119	96.82±7.44	78.4	114	< 0.01

Table 2: The results of logistic regression model

Variables	β	*Exp(β)*	*S.E*	*P-value*
Age	-0.089	0.915	0.058	0.126
Weight	0.001	1.001	0.03	0.96
Length	0.034	1.034	0.042	0.429
Work experience	0.11	1.117	0.139	0.429
Smoking status (Yes-No[*])	1.406	4.08	0.61	0.021
Daily number of consumed cigarettes	0.055	1.05	0.084	0.51
Smoking history(Years)	0.14	1.15	0.088	0.113
Using protective equipments(Yes-No[*])	-1.259	0.275	0.575	0.029

* Reference category

Table 3: Classification of workers based on neural networks and logistic regression predictions

Type of prediction model		Pulmonary disorder in real conditions	
		With pulmonary disorder	Without pulmonary disorder
Logistic regression	With pulmonary disorder	21	3
	Without pulmonary disorder	16	77
Neural networks	With pulmonary disorder	24	1
	Without pulmonary disorder	13	79

Table 4: The criteria of sensitivity, specificity, area under the ROC curve and Kappa statistics for neural networks and logistic regression models

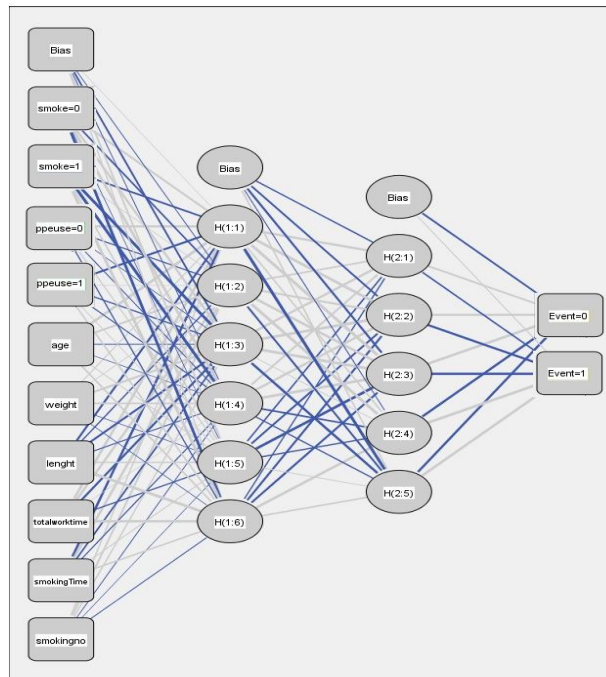| Type of model	Sensitivity	Specificity	Kappa statistics	Area under the ROC curve
Logistic regression	0.962	0.567	0.585	0.765
Neural networks	0.99	0.65	0.697	0.818

Fig1: The structure of the developed neural network for predicting pulmonary disorder

Discussion

More accurate predictions model for occupational disorders can be considered as a valid basis for planning and implementing healthy programs in order to treat workers exposure conditions. Therefore, highly accurate prediction models are necessary. The result confirmed that the neural networks and the logistic regression have acceptable performance to predict the complex phenomena such as the lung disorders in the workers exposed to harmful pollutants. The developed neural networks model has more appropriate performance compared with the logistic regression to predict pulmonary disorder in terms of the sensitivity, specificity, Kappa statistic and area under the ROC curve. Similar results was found in relation to classification of occupational pneumoconiosis based on the job descriptions and using neural networks compared with logistic regression (11). Studies performed in the medical field using neural network technique, with similar nature of phenomena, have been recommended using the neural networks technique compared with statistical techniques included predicting metabolic syndrome by Sadahi et al and predicting unwanted pregnancies by Hashemi et al (16,17,18). The limitations of this study in relation to the artificial neural networks included the need to user skill, spending long time for development of model structure and trial and error functions for improvement of prediction model. However, one of the main features of the artificial neural networks compared with classical models is no need for specific function as default. Moreover, artificial neural networks have the ability to identify complex and non-linear relationships between predictor and response variables through the hidden layers (10). It is noted that for achieving more credible models based on empirical data – driven techniques such as neural networks, is necessary to

gather the adequate and reliable data on the features affecting the phenomenon. In this regard, application domain of the developed model can be increased. If the new observation for prediction isn't in the domain of trained data used for model development, the target feature won't predict with an expected accuracy. The developed neural network model could accurately predict pulmonary disorder in workers exposed to silica dust based on job and behavior conditions of workers like smoking and using status of pulmonary protective equipments. Therefore, this model can be used by occupational health experts for evaluating exposure conditions of different workers in the stone crushing workplaces. Due to the frequent non curability of pulmonary disorders, results of predictions can be used for developing or improving preventive health programs.

Conclusions

The neural networks prediction model was more accurate compared to the logistic regression. In this regards, the developed prediction model can be used as a helpful tool for occupational health experts to evaluate the workers health status.

Acknowledgement

References

1-World Health Organization. Monographs on the evaluation of the carcinogenic risk of chemicals to humans: silica and some silicates. International Agency for Research on Cancer(IARC). Lyon, France;1997.

2-Fulekar MH. Occupational exposure to dust in quartz manufacturing industry. Ann Occup Hyg1999;43(4):269-73.

3-Bahrami AR, Mahjub H. Comparative study of lung function in Iranian factory workers exposed to silica dust. East Mediterr Health J 2003;9(3):390–8.

4-Aliabadi M, Bahrami AR, Mahjub H, Ghorbani F, Barghi MA, Golbabaie F. Evaluation of free silica emission in ambient air of stone crushing in Azandarian area of Hamadan state by using of X-Ray diffraction method. Sci J Hamadan Univ Med Sci 2007;14(3):67-73. [In Persian]

5-Mohebbi I, Abdi Rad I. Secondary spontaneous pneumothorax in rapidly progressive forms of silicosis: characterization of pulmonary function measurements and clinical patterns. Toxicol Ind Health 2007;23(3):125–32.

6-Neghab M, Choubineh AR. The Relationship between occupational exposure to cement dust and prevalence of respiratory symptoms and disorders. Kermanshah Univ Med Sci J 2007;11(2):215-26. [In Persian]

7-Neghab M, HassanZadeh J. Evaluation of respiratory disorders associated with occupational inhalation exposure to raw materials used in ceramic production. Iran J Epidemiol 2009;4(3,4):27-33.[In Persian]

8-Nwibo A, Ugwuja EI, Nwambeke NO, Emelumadu OF, Ogbonnaya LU. Pulmonary problems among quarry workers of stone crushing industrial site at Umuoghara, Ebonyi state, Nigeria. Int J Occup Environ Med 2012;3(4):178-85.

9-Ozkurt S, Kargi BA, Kavas M, Evyapan F, Kiter G, Baser S. Respiratory symptoms and pulmonary functions of workers employed in Turkish textile dyeing factories. Int J Environ Res Public Health 2012;9(4):1068-76.

10-Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 2000;43(1):3–31.

11-Liu H, Tang Z, Yang Y, Weng D, Sun G, Duan Z, et al. Identification and classification of high risk groups for Coal Workers' Pneumoconiosis using an artificial neural network based on occupational histories: a retrospective cohort study.BMC Public Health 2009;9:366-9.

12-Kolarzyk E, Stepniewski M, MendykcA, Kitlinski M, Pietrzycka A. The usefulness of artificial neural networks in the evaluation of pulmonary efficiency and antioxidant capacity of welders. Int J HygEnviron Health 2006;209(4):385–92.

13-Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. Eur Respir J 2005;26(2):319–38.

14-Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J 2005;26(5):948–68.

15-Agresti A. Categorical data analysis. 2nd ed. NewYork:John Wiley& Son;2002.

16-Sedehi M, Mehrabi Y, Kazemnejad A.[Comparison of artificial neural network, logistic regression and discriminant analysis methods in prediction of metabolic syndrome]. Iran J Endocrinol Metab2009;11(6):638-46.[In Persian]

17-Hashemi M.[Comparison of artificial neural network, logistic regression and discriminant analysis methods in prediction of pregnancy types]. IranJ Basic Med Sci2004;7(1):25-31. [In Persian]

18-Green M, Björk J, Forberg J,Ekelund U, Edenbrandt L, Ohlsson M.Comparison between neural networks and multiple logistic regression to predict acute coronary syndrome in the emergency room. Artif Intell Med 2006;38(3):305-18.