

# Prediction of mental disorders after Mild Traumatic Brain Injury: principle component Approach

Arash Nademi <sup>1</sup> Elham Shafiei <sup>2</sup> Esmaeil Fakharian <sup>3</sup> Abdollah Omidi <sup>4</sup>

<sup>1</sup> Department of Statistics, Ilam Branch, Islamic Azad University, Ilam, Iran.

<sup>2</sup> Psychosocial Injuries Research Center, Ilam University of Medical Sciences, Ilam, Iran.

<sup>3</sup> Trauma Research Center, Kashan University of Medical Sciences, Kashan, Iran.

<sup>4</sup> Department of Clinical Psychology, Kashan University of Medical Sciences, Kashan, Iran.

Received 8 Oct, 2017

Accepted 4 Mar, 2018

## Original Article

### Abstract

**Introduction:** In Processes Modeling, when there is relatively a high correlation between covariates, multicollinearity is created, and it leads to reduction in model's efficiency. In this study, by using principle component analysis, modification of the effect of multicollinearity in Artificial Neural Network (ANN) and Logistic Regression (LR) has been studied. Also, the effect of multicollinearity on the accuracy of prediction of mental disorders after trauma in patients with Mild Traumatic Brain Injury has been investigated.

**Methods:** In a prospective cohort Study, first, during 6 months period, 100 patients with Mild Traumatic Brain Injury have been selected. Then, by using Primary Covariates and Principle Component Analysis, Logistic Regression and ANN models have been conducted and based on these models prediction have been done. (Receiver Operating Characteristic) ROC curve and Accuracy Rate have been used to compare the strength of model's prediction.

**Results:** The results revealed that Accuracy Rate for ANN before and after applying principle component analysis are 84.22 and 91.23% respectively, and for Logistic Regression models are 72.33% and 74.89% respectively.

**Conclusion:** The study showed that the Accuracy Rate was higher for models based on Principle Component Analysis including primary covariates; hence, when multicollinearity exists, models that use the principle component for prediction of mental disorders are more effective compare to other methods. Also, ANN Models are more effective than Regression models.

#### Correspondence:

Elham Shafiei, PhD.

Psychosocial Injuries Research Center, Ilam University of Medical Sciences.

Ilam, Iran

Tel: +98 9188432410

Email:

eshafiei1524@gmail.com

**Key words:** Traumatic, Brain Injury, Mental disorder, Logistic Regression

**Citation:** Nademi A, Shafiei E, Fakharian E, Omidi A. Prediction of mental disorders after Mild Traumatic Brain Injury: principle component Approach. HMJ 2018;22(1):38-44.

### Introduction:

In the United States, approximately 1.4 million patients with brain injury are treated annually (1).

Also, in terms of mortality in Iran, Brain Injury has the second place. In addition, each year, most of the brain injuries, around two millions, that occur after accident are classified as Mild Traumatic Brain

Injury (MTBI) by using common criteria. In recent years, despite significant breaks through in the field of Head Trauma Surgery and improvement of the hospital services and health care system for patients that suffer from trauma which it leads to increase of their survival and recovery, however, patients with MTBI show set of extensive physical, psychological, neurological problems and signs of mental disorders in days and weeks after injury (2).

Nevertheless, the study about mental disorders followed TBI is still in early stage, and further investigation need to be conducted. Thus, presenting statistical models which can model mental health status is of high importance, and it can provide valuable data on training programs, treatment and the prevention of mental disorders for officials, planners, and managers of organizations or centers which develop services.

The main objectives of modeling are determination of relation between variables, determination of effective variables, and prediction (3). Therefore, presenting methods that based on them prediction has the least error seems useful. In the recent decades, one of these methods which have been used frequently is ANN, and also, it is common in various medical sciences including the prediction of cancer (4), prediction of death after stomach surgery (5), and etc.

One of the problems which cause reduction in accuracy of model's prediction in both methods-of Logistic Regression and ANN is multicollinearity (6). It happens when input variables have relatively high correlation with each other. Existence of multicollinearity in Logistic Regression leads to increase of error in estimation of regression coefficients, and it may cause predictions that are outside of the expected range (7). In ANN model, multicollinearity causes reduction in accuracy of prediction, and estimation of weights in different layers of the network encounters a lot of changes for any iterations of training algorithm, and this issue may leads to divergence in ANN (8). This problem is more likely, especially when the numbers of input variables are high. One of the ways of tackling this problem is use of principle component analysis for processing of input variables, elimination of correlation between input variables, and reduction of number of input variables which it leads to improvement of results

of regression and ANN models (9). Principle Component Analysis is one of the multivariate statistical methods which can be used for reduction of number of variables and for better interpretation of data (10). By using this method, primary input variables change into new components that have no correlation, so that created components are linear combination of input variables. Principle Component Analysis is one of the most practical methods for dimension reduction in multivariate models (11). Principle components based on their characteristics are used for dealing with multicollinearity and dimension reduction of data. In this method, first, by using Eigen value Matrix, principle components are made from linear combination of the initial variables, and then, these principle components are used instead of initial variables in data analysis. This study has been conducted to examine the effect of principle components on the accuracy of prediction of mental disorder after Mild Traumatic Brain Injury by using both Logistic Regression and ANN Models.

### Methods:

This prospective cohort study had been carried out on 100 patients with Mild Traumatic Brain Injury, with GCS of between 13 and 15, that were hospitalized in pediatric neurosurgery ward. By using non-probability sampling method, samples were selected from patients of both sexes with the age of between 15 and 65, and after approve of ethics committee of the hospital and consent of patients.

In the first phase, hospitalized patients were examined by the neurosurgeon, and based on the defined criteria, if they are eligible for the study, their data, both demographic and clinical data, was entered the questionnaire. Then, after 6 months, they were asked to attend advising and neurological evaluation center to complete BSI questionnaire. In order to meet consistency in completion of the tests, questions were uttered one by one, then, their oral answers were recorded in the related option of the questionnaire by a psychology expert (MA). Also, some patients were excluded from this study including patients with the history of psychotic diseases, patients who are unwilling to take part in this study, patients that had vegetative state or

severe impairment of consciousness, patients with any evidence of Spinal Cord Injury, and also, patients that had any types of neurological disease before TBI or brain injury with non-traumatic sources including brain tumors, stroke, arterial dilatation and other brain accidents.

In this study, Instrument that has been used was Brief Symptom Inventory (BSI) questionnaire which is the short form of the SCI-90-R questionnaire. This questionnaire includes 53 questions which evaluate psychological symptoms, and it can be used to separate healthy individuals from the patients. Dragutis et al. in 1973 introduced this questionnaire which consists of 9 dimensions as following: somatization, obsessive-compulsive, interpersonal sensitivity, depression, anxiety, anger, anxiety phobia, paranoia, psychoticism. After completion of the questionnaires, the score for each questionnaire is calculated, and if the value of T-score was above 60 in any subscales, they were considered as a patient.

Input variables of this study were as following: demographic variables such as age, gender, job, level of education, marital status, financial status, history of mental disorders among family members, history of hospitalization in neurosurgery ward, history of having trauma, history of taking psychological medicine, and alcohol usage. Neurological variables including duration of being hospitalized in neurosurgery ward, location of head trauma, having multi trauma.

In order to process Logistic Regression and ANN models, data were divided randomly into two groups the first half have been used for processing of models, and the other half for investigating the accuracy of the prediction and validation. Stages of random fragmentation of data were repeated 300 times, and in each stage of Logistic Regression Model, stepwise method with Forward selection approach with entry significant level of 0.05 and exit significant level of 0.10 have been used. The best regression model was obtained by using Akaike Information Criterion.

In order to use ANN Model, the stages of the random selection of data was conducted similar to Regression Model, and in each stage, Perceptron Artificial Neural Network with 3 layers have been used. The first layer, or the input layer, includes input variables. The second layer is middle layer, or

the hidden layer, and the third layer is determined by response variable which in this study, due to the type of response variable of mental disorder, T-scorer > 60 = patient, T-scorer < 60 = healthy, output layer of the network with 1 neurons was considered. In the training stage of used ANN, 3 layers network with 14 input neurons accordance with the number of variables, 1 output nodes with back propagation learning algorithm, Sigmoid Transfer Function, learning rates of 0.01 to 0.40, and momentum of was 0.8 to 0.95 have been used. The number of neurons in the middle layer was considered 6 to 11 which based on least Root-Mean-Square Deviation (RMSD), appropriate number of neurons in the middle layer was determined. The number of neurons in the middle layer is important, because if it is low, the network will face lack of learning sources for non-linear and complex matters, and if it is high, it will lead to 2 problems: first, the time of training of network will increase, and second, the network may learn errors in data, and as a result, it may work poorly for prediction. Finally, after selection of best structure, the network was tested and validated using data that did not participate in modeling. To compare of predictions resulted from Logistic Regression and ANN Model, ROC analysis and Accuracy Rate have been used.

For investigation of multicollinearity in the observations, Correlation Coefficient Matrix and Bartlett's Test have been used, and for investigation of sampling adequacy, Kaiser Meyer Olkin (KMO) has been used. Bartlett's Test studies whether Correlation Matrix is the same as the identity matrix or not. If Correlation Matrix is identity matrix, independent input variables will disintegrate, and if the test became meaningful ( $\text{sig} < 0.05$ ), it will show correlation among variables, and therefore, use of Principle Component Analysis can be more effective. KMO Index measures the size of partial correlation among variables, and it determines whether the variance of the research variables is under the influence of shared variance of some basic and hidden factors or not. The index is between 0 and 1, and values of near 1 illustrate sampling adequacy. By combine of 14 input variables with Principle Component Analysis, 4 principle components have been obtained which have been used as predictive variables in Logistic

Regression and ANN models. Finally, after selection of the best models, ROC curve Analysis and Accuracy Rate have been used to compare predictions resulted from Logistic Regression and ANN, before and after use of Principle Component Analysis.

## Results:

Processing Modeling in ANN was carried out by using training data set. By processing different models of 3-layers ANN for 6 structures including 6 to 11 neurons in the middle layer, a model with 14 input neurons, 8 middle neurons and 1 output neurons, with learning rate of 0.06, and momentum of 0.98, and with back propagation learning algorithm as an appropriate model for prediction of data in which mean square error was 0.1164 and Accuracy Rate was 84.22% (Table 2 and 3).

**Table 1. The value of defined variance by 4 main components**

Principle components	Percentage of the variance defined	The cumulative percentage of variance
PC1	23.487	23.487
PC2	21.248	44.735
PC3	19.508	64.243

**Table 2. Selection of best models of Neural Networks for psychiatric disorder data before and after Principle Component Analysis**

	Architecture of models (input/middle/output layers)	Mean square error	Percentage of false prediction
	(14/6/1)	0.1214	16.97
<b>Before</b>	(14/7/1)	0.1203	16.84
<b>Principle Component Analysis</b>	(14/8/1)*	0.1164	15.78
	(14/9/1)	0.1186	16.02
	(14/10/1)	0.1194	16.35
	(14/11/1)	0.1198	16.48
<b>After principle Component Analysis</b>	(4/2/1)*	0.1011	8.77
	(4/3/1)	0.1129	9.12
	(4/4/1)	0.1168	9.64

\*The best selected architecture after training the network including least mean square error

As it continues, data have been processed by Logistic Regression Model. In this model, by using stepwise method with Forward selection approach, an appropriate model based on 14 predictive variables were selected so that in the final model, Akaike Information Criterion of 124.01 have been obtained, and accuracy was 72.33% (Table 3).

Results of Bartlett's Test, Chi-Square=12.34, Sig<0.05, and KMO=0.735 show that the data satisfy the criterion for apply of Principle Component Analysis. In sum, 78.68 percent of the total variance has been covered by four main components (PC1, PC2, PC3, PC4) that have been used in the analysis. So, in the next stage, 4 principle components have been used as inputs for ANN and Logistic Regression Models (Table1).

Processes Modeling in ANN was similar to modeling in the previous stage with the following differences: in the middle layer, 2 to 4 neurons have been used in which final model had 4 input neurons, 2 middle neurons and 2 output neurons, with learning rate of 0.05, momentum of 0.9, and back propagation learning algorithm as a good model for prediction of data so that its mean square error was 0.1011 and accurate predictions of the model was 91.23% (Table 2 and 3).

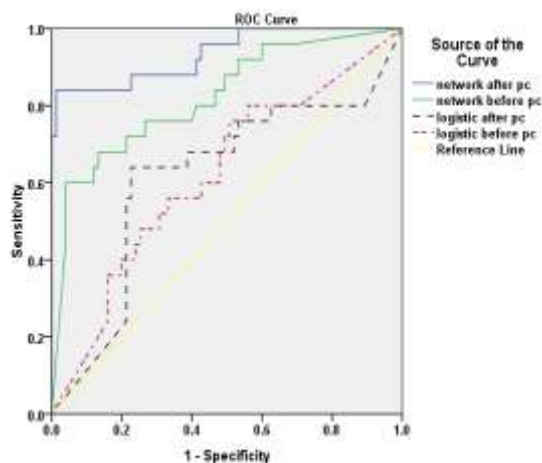
In addition, in process of Logistic Regression Model, the appropriate model has been selected based on 4 predictive variables (PC1- PC4) so that Akaike Information Criterion and Accuracy Rate obtained for the final model were 110.16 and 74.89%, respectively.

One of diagnostic criteria for determination of performance of models is the area under ROC curve so that value of 0-0.5 show random rating and value of 0.5-1 show overall diagnostic capability of models. According to Table 3, the area under ROC curve in the test for ANN before and after Principle Component Analysis were 82.1 and 93.4 respectively, and for logistic regression models were 73.2 and 76.8% respectively. Figure 1 presents ROC Curve for models.

**Table 3. Compared Results of 300 pairs of ANN and Logistic Regression Model before and after a Principle Component Analysis**

	LR before PC	LR after PC	ANN before PC	ANN after PC
Area under ROC curve	07.32 (0.621-0.789)*	0.768 (0.653-0.883)	0.821 (0.795-0.945)	0.934 (0.872-0.997)
Accuracy Rate	72.33 (71.12-74.53)	74.89 (75.23-76.35)	84.22 (83.20-85.16)	91.23 (91.01-91.86)

\*Confidence interval of 95%

**Figure 1. ROC curve based on ANN and Logistic Regression before and after a Principle Component Analysis****Conclusion:**

Mental disorders and signs like depression, anxiety, and obsession along with Traumatic Brain Injury cause reduction in social activities, and finally impose a huge burden on the health care system, society and their families. Therefore, this study aimed to find the best model to predict the probability of mental disorder by BSI. Therefore, the results of Logistic Regression Models and with Artificial Neural Network Models before and after Principle Component Analysis have been used for prediction of mental disorder. So far, Studies regarding statistical modeling which have been carried out about mental disorder after trauma, mainly aimed at investigation of effective factors on mental disorder by using Logistic Regression, but the current study seeks for investigation of accuracy of prediction with Logistic Regression and ANN Models which have not been investigated yet. Here, in order to decrease the effect of multicollinearity, Principle Component Analysis has been used to increase accuracy of prediction in Logistic Regression and ANN Models. The idea of using principle component instead of all variables of the

study with the purpose of reducing variables dimensions was resulted from this fact that the covariate variables can merge based on the relationship between observations.

Neural network is a good way to predict the probability of having mental disorder in Traumatic Patients. Also, this study shows that the accuracy of prediction of mental disorder in Logistic Regression and ANN based on principle component is higher than those models that are based on primary collinear variables. Hence, when covariates have correlation, Principle Component Analysis can be effective on increase of the accuracy of prediction for Logistic Regression and ANN Models.

So, the prediction resulting from this method can be applied in classification of patients. Additionally, considering that nearly all predicting models use linearity and logistic for analyzing, using modified nonlinear neural network can be applied in designing more effective plans for screening individuals susceptible to mental disorder.

**Acknowledgments:**

This paper is a part of research design approved by Medical University of Kashan (the project number=92115), Iran. Through this, we also sincerely thank Research Assistant of Medical University of Kashan.

**References:**

1. Naghavi M, Abolhassani F, Pourmalek F, Lakeh M, Jafari N, Vaseghi S, et al. The burden of disease and injury in Iran 2003. *Population Health Metrics*. 2009;7:9.
2. Dobryakova E, Boukrina O, Wylie GR. Investigation of information flow during a novel working memory task in individuals with traumatic brain injury. *Brain Connect*. 2014;5(7):433-441.

3. Anastasijevic A, Coja D, Neskovic N, Neskovic A, Budimir D. Joint power amplifier and I/Q modulator impairments modelling and compensation for LTE transmitters using artificial neural networks. *Int J Electron Commun.* 2015;69(2):529-538.
4. Lv DJ, Zhang Y, Wang XY, Guo XM, Wang CY. [Application of artificial neural network to diagnosis of prostate cancer]. *Beijing da xue xue bao Yi xue ban.* 2009;41(4):469-473.
5. Amiri Z, Mohammad K, Mahmoudi M, Parsaeian M, Zeraati H. Assessing the effect of quantitative and qualitative predictors on gastric cancer individuals survival using hierarchical artificial neural network models. *Iran Red Crescent Med J.* 2013;15(1):42-48.
6. Zhang WG, Bai XJ, Sun XF, Cai GY, Bai XY, Zhu SY, et al. Construction of an integral formula of biological age for a healthy Chinese population using principle component analysis. *J Nutr, Health & Aging.* 2014;18(2):137-142.
7. Cruces S, Duran-Diaz I. The minimum risk principle that underlies the criteria of bounded component analysis. *IEEE Trans Neural Netw Learn Syst.* 2015;26(5):964-981.
8. Huang TL, Cho YT, Su H, Shiea J. Principle component analysis combined with matrix-assisted laser desorption ionization mass spectrometry for rapid diagnosing the sera of patients with major depression. *Clin Chim Acta.* 2013;424:175-181.
9. Lei J, Liu S, Wang X, Liu Q. An image reconstruction algorithm for electrical capacitance tomography based on robust principle component analysis. *Sensors (Basel).* 2013;13(2):2076-2092.
10. Taspinar F. Improving artificial neural network model predictions of daily average PM10 concentrations by applying principle component analysis and implementing seasonal models. *J Air Waste Manag Assoc.* 2015;65(7):800-809.
11. Vural C, Yildiz M. Determination of sleep stage separation ability of features extracted from EEG signals using principle component analysis. *J Med Syst.* 2010;34(1):83-89.

## پیش‌بینی اختلالات روانی بعد از تروما در بیماران دچار آسیب مغزی خفیف: رویکرد تحلیل مولفه‌های اصلی

آرش نادمی<sup>۱</sup> الهام شفیعی<sup>۲</sup> اسماعیل فخاریان<sup>۳</sup> عبدالله امیدی<sup>۴</sup>

<sup>۱</sup> گروه آمار، واحد ایلام، دانشگاه آزاد اسلامی، ایلام، ایران.

<sup>۲</sup> مرکز تحقیقات آسیب‌های روانی - اجتماعی، دانشگاه علوم پزشکی ایلام، ایلام، ایران.

<sup>۳</sup> مرکز تحقیقات تروما، دانشگاه علوم پزشکی کاشان، کاشان، ایران.

<sup>۴</sup> گروه روانشناسی بالینی، دانشگاه علوم پزشکی کاشان، کاشان، ایران.

مجله پزشکی هرمزگان سال بیست و یکم شماره پنجم ۹۶ صفحات ۴۴-۳۸

### چکیده

**مقدمه:** در فرآیندهای مدل‌سازی، زمانی که بین متغیرهای کمکی همبستگی‌های نسبتاً قوی وجود داشته باشد، هم خطی چندگانه ایجاد شده و باعث کاهش کارایی مدل می‌گردد. هدف از این مطالعه، استفاده از تحلیل مولفه‌های اصلی برای تعدیل اثر هم خطی چندگانه در مدل‌های رگرسیون لجستیک و شبکه عصبی مصنوعی و بررسی تأثیر آن بر صحت و دقت پیش‌بینی اختلالات روانی بعد از تروما در بیماران دچار آسیب مغزی خفیف بود.

**روش کار:** در یک مطالعه کوهورت آینده نگر، ۱۰۰ نفر بیمار ترومایی طی مدت ۶ ماه انتخاب شدند. سپس، مدل‌های رگرسیون لجستیک و شبکه عصبی مصنوعی با استفاده از متغیرهای اولیه و سپس با استفاده از روش مولفه‌های اصلی به داده‌ها برازش گردید و پیش‌بینی بر اساس این مدل‌ها انجام شد. از تحلیل سطح زیرمنحنی راک (ROC) و صحت کلاسه‌بندی برای مقایسه قدرت پیش‌بینی مدل‌ها استفاده گردید.

**نتایج:** نتایج این مطالعه نشان داد، شاخص درصد پیش‌بینی درست برای مدل‌های شبکه‌های عصبی مصنوعی قبل و بعد از استفاده از روش مولفه‌های اصلی به ترتیب برابر ۸۴/۲۲ و ۹۱/۲۳ درصد و برای مدل‌های رگرسیون لجستیک به ترتیب برابر ۷۲/۳۳ و ۷۴/۸۹ درصد می‌باشد.

**نتیجه‌گیری:** تحقیق نشان داد که صحت پیش‌بینی مدل‌ها بر اساس مولفه‌های اصلی از مدل مبتنی بر متغیرهای اولیه بیشتر بوده و بنابراین در هنگام وجود هم خطی چندگانه، مدل‌های مبتنی بر مولفه‌های اصلی برای پیش‌بینی اختلال روانی کاراتر هستند. همچنین مدل‌های شبکه عصبی مصنوعی کارایی بیشتری را نسبت به مدل‌های رگرسیونی دارند.

**کلیدواژه‌ها:** تروما، آسیب مغزی، اختلال روانی، رگرسیون لجستیک

نویسنده مسئول:

دکتر الهام شفیعی

مرکز تحقیقات آسیب‌های روانی -

اجتماعی، دانشگاه علوم پزشکی ایلام

ایلام - ایران

تلفن: +۹۸ ۹۱۸۸۴۲۲۴۱۰

پست الکترونیکی:

eshafie1524@gmail.com

**نوع مقاله:** پژوهشی

دریافت مقاله: ۹۶/۷/۱۶ اصلاح نهایی: ۹۶/۱۱/۲۳ پذیرش مقاله: ۹۶/۱۲/۱۳

**ارجاع:** نادمی آرش، شفیعی الهام، فخاریان اسماعیل، امیدی عبدالله. پیش‌بینی اختلالات روانی بعد از تروما در بیماران دچار آسیب مغزی خفیف: رویکرد تحلیل مولفه‌های اصلی. مجله پزشکی هرمزگان ۱۳۹۶؛ ۲۱(۵): ۴۴-۳۸.