Letter to the Editor

What is the better model in burn patients?

Dear Sir

We read with interest the research article by Yeong et al. [1], and praise their implementation to construct an artificial neural network for predicting burn healing time based on the data from reflectance spectrometer.

Artificial neural networks are a computational model composed of non-linear processing elements arranged in highly interconnected layers with a configuration that simulates a biological nervous system. It has been widely investigated as a predictive model in clinical medicine with the assistance of advances in computer-aided analysis. Therefore, we must ask whether the selected artificial neural network is valuable as a useful model in our daily practice. To evaluate the quality for classification models, *discrimination* and *calibration* are simultaneously required in model assessment [2]. Discrimination is a measure of how well a model recognizes subjects correctly as two different classes; calibration, on the other hand, evaluates the degree of correspondence between the estimated probabilities produced by a model and the actual observation.

Common assessments used in discrimination for predictive classification include sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratios for positive and negative tests, and area under the receiver operating characteristic curve. Although Yeong et al. documented the sensitivity and specificity of their artificial neural network model, they did not report area under the receiver operating characteristic curve, which provides a superior index for the performance of predictive models.

Although area under the receiver operating characteristic curve with the best concurrent sensitivity and specificity can be used to determine discriminatory power of a model, poor calibration can occur in a superior discriminating model when classifier outputs are transformed monotonically. To alleviate this problem, one can use misclassification rate, Pearson Chi-square, or Hosmer–Lemeshow statistic [3] as measurements of calibration.

We would like to emphasize the importance of calculating discrimination and calibration concurrently in the field of predictive model construction and comparison because calibration information was ignored in 75% of papers using artificial neural network to build their classification models [4]. Application of any new model should be critically appraised since inappropriate evaluations may lead to wrong decisions and can, in turn, affect the therapeutic plans and outcome for patients.

Reference

[1] Yeong EK, Hsiao TC, Chiang HK, Lin CW. Prediction of burn healing time using artificial neural networks and reflectance spectrometer. Burns 2005;31(4):415–20.

[2] Li YC, Liu L, Chiu WT, Jian WS. Neural network modeling for surgical decisions on traumatic brain injury patients. Int J Med Inform 2000;57(1):1–9.

[3] Lemeshow S, Hosmer DW. A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 1982;115(1):92–106.

[4] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform 2002;5(5):52–359.

Jainn-Shiun Chiu
Chao-Shun Lin
Fu-Chiu Yu
Yu-Chuan Li*
Graduate Institute of Medical Informatics
Taipei Medical University, No. 250
Wusing St., Sinyi District, Taipei City, Taiwan

*Corresponding author. Tel. +886 2 23776730
fax: +886 2 27339049
E-mail address: jack@tmu.edu.tw