

Clinical Applications of Artificial Intelligence to Forecast Target Range of Radioimmuno-metric Intact Parathyroid Hormone in Hemodialysis Patients

Jainn-Shiun Chiu¹, Shih-Hsueh Huang², Tsung-Ming Hu³,
Yen-Yu Chen⁴, Yu-Chuan Li⁵, Yuh-Feng Wang¹

¹*Department of Nuclear Medicine, Buddhist Dalin Tzu Chi General Hospital, Chiayi, and Department of Medicine, Tzu Chi University, Hualien, Taiwan*

²*Department of Management Information System, Buddhist Dalin Tzu Chi General Hospital, Dalin, Chiayi, Taiwan*

³*Department of Adult Psychiatry, Yuli Hospital, Department of Health, Executive Yuan, Hualien, Taiwan*

⁴*Department of Neurology, Changhua Christian Hospital, Changhua, Taiwan*

⁵*Institute of Biomedical Informatics, National Yang Ming University, Taipei, Taiwan*

Background: Application of artificial intelligence to predict and explore potential relationship between predictors and outcome in biologic nature has been increasingly used in many clinical scenarios. The purpose of this study was to apply and validate artificial neural network (ANN) and naive Bayes classifier (NBC), two models of artificial intelligence, in predicting the target range of plasma intact parathyroid hormone (iPTH) concentration for hemodialysis patients.

Methods: The study population included 130 stable hemodialysis patients. The predictors consisted of demographic characteristics (gender, age), associated diseases (diabetes, hypertension), and blood biochemistries (hemoglobin, protein, albumin, calcium, phosphorus, alkaline phosphatase, and ferritin), calcium-phosphorus product, and transferrin saturation values. Plasma iPTH concentration measured by radioimmuno-metric assay was the dichotomous outcome variable, either target group ($150 \text{ ng/L} \leq \text{iPTH} \leq 300 \text{ ng/L}$) or non-target group ($\text{iPTH} < 150 \text{ ng/L}$ or $\text{iPTH} > 300$

ng/L) on the basis of Kidney Disease Outcomes Quality Initiative guidelines. The leave-one-out cross validation was employed to surmount the generalization problem caused by a small amount of study population. To compare the performance of the ANN and NBC models, discrimination was evaluated using the area under the receiver operating characteristic curve (AUC) and calibration was estimated using the Hosmer-Lemeshow goodness-of-fit statistic (H-statistic).

Results: Pairwise comparison of each AUC showed that the ANN model significantly outperformed the NBC model (AUC = 0.90 ± 0.06 vs. 0.62 ± 0.08 , $P < 0.01$). The H-statistic values of the ANN and NBC models were 6.88 ($P = 0.08$) and 6.97 ($P = 0.07$), respectively. The ANN model with a lower H-statistic and a higher P value than the NBC model was associated with a better fit.

Conclusion: The ANN model could serve as a promising tool to forecast the target range of plasma iPTH concentration in hemodialysis patients.

Key words: artificial intelligence, hemodialysis, naive Bayes classifier, neural network, parathyroid hormone

Received 5/16/2006; revised 6/27/2006; accepted 6/28/2006.

For correspondence or reprints contact: Yuh-Feng Wang, M.D., Department of Nuclear Medicine, Buddhist Dalin Tzu Chi General Hospital. Tel: (886)5-2648000 ext. 5700; Fax: (886)5-2648508, E-mail: nment@ms7.hinet.net

Ann Nucl Med Sci 2006;19:149-159



Introduction

Renal osteodystrophy is an influential factor of morbidity in patients with chronic kidney disease. It is generally classified as a high bone turnover disease (e.g., osteitis fibrosa) or low bone turnover disease (e.g., osteomalacia or adynamic bone disease) based on the plasma level of parathyroid hormone (PTH) and representative findings of bone histomorphometry. However, bone biopsy is an invasive and painful procedure which is rarely performed in clinical practice. Therefore, determination of plasma PTH level is pivotal for optimal intervention for the prevention and treatment of renal osteodystrophy since dosages of phosphate binders, vitamin D analogues, or calcimimetic drugs are established upon plasma calcium, inorganic phosphorus, and especially PTH concentrations. Suggested by Kidney Disease Outcomes Quality Initiative (K/DOQI) *Clinical Practice Guidelines for Bone Metabolism and Disease in chronic kidney disease*, plasma intact PTH (iPTH) concentration between 150 and 300 pg/mL is currently recommended as target range and must be monitored every three months in dialysis patients [1].

In the field of artificial intelligence, artificial neural network (ANN) is a revolutionary application in clinical medicine. It is a computational model composed of nonlinear processing elements ("neurons") arranged in highly interconnected layers with a configuration that simulates a biological nervous system [2]. Every processing element is interconnected through a set of weighted signals similar to synaptic connections used in memory and learning [3]. The ANN model has the advantage of recognizing relationships between input variables (data from cases) and output variables (known outcomes) that may not be apparent when using traditional statistical techniques [4]. Furthermore, ANN can improve its accuracy through a learning algorithm, and it has been successfully used as a decision supporting tool to solve clinically complicated problems [5]. For nuclear medicine physicians, some applications of ANN based on scintigraphy of particular interest are diagnosis of coronary artery disease [6-9], assessment of cardiac function [10], detection of pulmonary embolism [11-14], prediction of parathyroid adenoma [15], nodal staging in lung cancer [16], generation of regions of interest [17], classification of brain

diseases [18-21], stratification of chronic hepatitis and cirrhosis [22], and issue of renal transplants [23]. On the other hand, naive Bayes classifier (NBC) is based on the Bayesian theorem primarily formulated for performing classification tasks [24]. The NBC assumes that the independent variables are statistically independent and is a particularly appropriate easy-to-use classification tool when the number of dimensions of the input variables is high. The NBC often outperforms more complicated classification methods. Applications of the NBC in biomedicine include recruitment of patients with congestive heart failure in a clinical trial [25], differential gene expression profiling in mantle cell lymphoma [26], diagnoses of dementia [27], and prediction of prostate cancer recurrence [28]. Herein, the ANN and NBC models represent two different applications of artificial intelligence, which one is complex and the other is simple.

In addition to the growing prevalence of low bone turnover diseases in uremic patients, overzealous treatment of high bone turnover disease patients with vitamin D analogues could result in the occurrence of low bone turnover diseases [29]. Therefore, frequent monitoring of plasma iPTH concentration by the K/DOQI-recommended radioimmunoassay is reasonable [1]. However, frequent monitoring might result in expense stress of financial affairs and burden of medical insurances. Actually, many dialysis institutions monitor plasma iPTH concentrations for dialysis patients every six months or longer. In an attempt to ensure dialytic quality without additional cost of frequent measurements, clinical applications of artificial intelligence could be a feasible tool to solve these clinical problems. In a previous paper, we utilized ANN modeling to predict plasma iPTH concentration using biochemical covariates from continuous ambulatory peritoneal dialysis patients, and showed the performance was comparable with the radioimmunoassay measurement [30]. In this report, the ability to identify hemodialysis patients whose plasma iPTH concentration will be within the target range from those for whom it will not, is investigated using ANN and NBC models.

Materials and Methods

The patients were selected from an independent dialysis unit of a teaching hospital in western Taiwan (Buddhist

Dalin Tzu Chi General Hospital, Chiayi County) on the basis of the following criteria: (A) age more than 18 years; (B) clinically stable status with end-stage renal disease on maintenance hemodialysis for more than 6 months; (C) absence of heart disease, liver disease, infection, and other major diseases, as determined by clinical history and physical examination; (D) absence of admission within 6 months. The final study population consisted of 130 stable hemodialysis patients. All patients received 4-h hemodialysis session thrice a week using a cellulose acetate hollow fiber dialyzer with blood flow 250-400 ml/min, and dialysate flow 500 ml/min. The Ethics Committee on Human Studies in our hospital approved the study and informed consent for enrolled patients was not required for clinical data collection from medical records according to the protocol of our institutional review board. To preserve patient confidentiality, direct patient identifiers were not collected. Data were reported only in aggregate form.

The demographic characteristics (gender, age), associated diseases (diabetes, hypertension), blood biochemistries (hemoglobin, protein, albumin, calcium, phosphorus, alkaline phosphatase, and ferritin), calcium-phosphorus product, and transferrin saturation values - monthly information typically available to the physicians - were entered as continuous or nominal input variables as predictors into the ANN and NBC models, respectively. On the basis of K/DOQI guidelines [1], the target group included patients having plasma iPTH concentration within the range $150 \text{ pg/mL} \leq \text{iPTH} \leq 300 \text{ pg/mL}$ and the non-target group included patients having $\text{iPTH} < 150 \text{ pg/mL}$ or $\text{iPTH} > 300 \text{ pg/mL}$, and "target range" was entered as the dichotomous output variable. Blood hemoglobin was measured by hematology analyzer (Sysmex XE 2100, Kobe, Japan); blood protein, albumin, calcium, phosphorus, alkaline phosphatase, and ferritin concentrations were measured by automatic biochemistry analyzer (Hitachi 7170; Hitachi Co., Tokyo, Japan) and plasma iPTH concentration was measured by radioimmunoassay (Active I-PTH DSL-8000; Diagnostic Systems Laboratories, Inc., Webster, Texas, USA). All data of input and output variables were simultaneously collected.

To solve the statistical problem stemming from a small amount of patients in our study, the leave-one-out cross vali-

dation was employed in both models. This technique requires a substantial number of experiments, but it can compute an average score over different partitions with an accurate estimate of generalization performance which is destined to avoid possible bias introduced by depending on any one particular subset into test and training parts [31]. The leave-one-out cross validation involves removing one case from the training data; training is performed on the basis of remaining data and then testing is executed on this removed case. In this manner, if the training data consist of 130 cases, then 130 models are produced using each of the cases as a test set while using the other cases as the training data. Although this is the most extreme test of cross validation, it is the most accurate way to estimate the performance of a method when the amount of training data is small [32].

STATISTICA 7.0 (StatSoft, Inc., Tulsa, Oklahoma, USA) was used to construct the ANN and NBC models. No any protocol is standardized to determine the configuration of the ANN model including numbers of hidden layers, numbers of neurons in each hidden layer, or activation functions. Initially, the formation is created by the designer and the best practice seems to be based on trial and error. Hence, the popular multilayer perceptron network was selected and a built-in automatic network designer was employed to decide an appropriate architecture, using a combination of heuristic and optimal algorithms [33]. The network designer conducted a large number of tests to decide the best architecture and automatically chose the smoothing factor and the number of units for multilayer perceptron network. To compare the performance of networks with different predictors, the network designer balanced error against type and diversity as criteria, in which case it conserved networks with a range of performance/complexity trade-offs. If the dossier is full and the new model is inferior to the candidate for substitution, the network set will be intensified in maximal size to comply the new networks. After the network was allowed to run and a prediction was made, the predicted outcome was correlated with the observed outcome; and if the network predicted the outcome incorrectly, by a process of back propagation, hidden weights within the network were remodified until the predicted outcome was accurate. At last, the intelligent problem solver retained the best network, architecture, and the

optimum set of input variables.

For building the NBC model, all input and output variables were unconditionally entered into the model since the NBC model can deal with an unrestricted number of independent variables whether continuous or categorical. If there is a set of predictors $X = \{x_1, x_2, x_3, \dots, x_n\}$, we may construct the posterior probability for the categorical levels C_j among a set of possible dependent variables $C = \{c_1, c_2, c_3, \dots, c_n\}$. Using Bayes' rule: $p(C_j | x_1, x_2, x_3, \dots, x_n) \propto p(x_1, x_2, x_3, \dots, x_n | C_j)$, where $p(C_j / x_1, x_2, x_3, \dots, x_n)$ is the posterior probability of class membership, i.e., the probability that X belongs to C_j . Since the NBC model supposes that the conditional probabilities of the independent variables are statistically independent, we can disintegrate the likelihood to a product of terms: $p(X | C_j) \propto \prod_{k=1}^n p(x_k | C_j)$ and revise the posterior as:

$p(C_j | X) \propto p(C_j) \prod_{k=1}^n p(x_k | C_j)$. Using Bayes' rule above, we label a new case X with a class level C_j that achieves the highest posterior probability. In order to specify the conditional distributions of the independent variables that are utilized to estimate the conditional posterior probabilities for making forecasting, the distributions of continuous and categorical variables were selected as normal and discrete, respectively. When the frequency of a categorical level of an independent variable is zero concerning one of the categorical levels of a dependent variable, the likelihood of the NBC model will always produce zero, which is not profound. Hence, the threshold parameter was set as 0.0001 to specify a lower bound when such frequencies are encountered. After the predictions were made by the classifier, its performance was compared with the ANN model.

Data were analyzed using MedCalc 8.1 (MedCalc Software Inc., Mariakerke, Belgium) and expressed as mean \pm standard error (SE). To assess the quality of classification models in clinical investigation, discrimination and calibration should be calculated concurrently [34]. 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. To compare the discriminatory ability of the ANN and NBC models, the receiver operating characteristics (ROC)

curves were plotted as a graph and the area under ROC curves (AUC) were calculated as a measure of a model's discriminatory power with pairwise comparison (a statistical level was defined as p value less than 0.05) [35]. An AUC of 1.0 infers perfect discrimination, whereas an AUC of 0.5 is equivalent to a random model. An AUC between 0.7 and 0.8 was classified as "acceptable" and between 0.8 and 1.0 as "excellent" discrimination [36]. On the other hand, calibration was assessed using the Hosmer-Lemeshow goodness-of-fit statistic (H-statistic) which divides subjects into deciles based on predicted probabilities and then computes a chi-square from observed and expected frequencies [37]. Lower H-statistic value and higher p value is associated with a better fit (a statistical fit was defined as p value more than 0.05).

Results

The characteristics of input variable are presented in Table 1. The mean age of all hemodialysis patients was 59.72 ± 1.23 years and the male to female ratio was 0.91:1. The mean plasma iPTH concentration measured by radioimmunoassay was 118.49 ± 13.62 pg/mL. The patient numbers in the target and non-target groups were 13 (10%) and 117 (90%), respectively. Figure 1 depicts the schematic diagram of final best ANN, which was created to predict the target range of radioimmunoassay plasma iPTH concentration in hemodialysis patients. The figure shows the number of artificial neurons in each of the four layers, and it illustrates the fact that the network was fully connected in that each artificial neuron in a given layer was connected to every

Table 1. The characteristics of the input variables

Input variables	Mean \pm standard error or percentage
Gender (male, %)	47.69
Age (years)	59.72 ± 1.23
Diabetes (%)	39.23
Hypertension (%)	57.69
Hemoglobin (g/dL)	9.47 ± 0.11
Protein (g/dL)	7.34 ± 0.05
Albumin (g/dL)	3.90 ± 0.03
Calcium (mg/dL)	9.35 ± 0.06
Phosphorus (mg/dL)	4.29 ± 0.13
Alkaline phosphatase (IU/L)	133.52 ± 4.72
Ferritin (ng/mL)	678.89 ± 55.01
Calcium-phosphorus product	40.18 ± 1.20
Transferrin saturation (%)	37.17 ± 1.51

artificial neuron in the adjacent layer. No input variable was pruned after training processes and all were adopted as significant predictors. The fact that the network was fully interconnected meant that 351 weights had to be modified following the processing of each record in the training period.

Figure 2 depicts the ROC curves for the ANN and NBC models. The AUCs of the ANN and NBC models were 0.90 ± 0.06 and 0.62 ± 0.08 , respectively. Discriminatory power of the ANN model's AUC was excellent, while that of the NBC model's AUC was not acceptable. Pairwise comparison of the AUCs showed a difference between the areas of the ANN and NBC models of 0.28 ± 0.09 ($P < 0.01$), suggesting that the ANN model was significantly better than the NBC model to distinguish between the target and non-target groups (i.e., groups having plasma iPTH concentration inside and outside the target range, respectively). The H-statistic values were 6.88 ($P = 0.08$) in the ANN model and 6.97 ($P =$

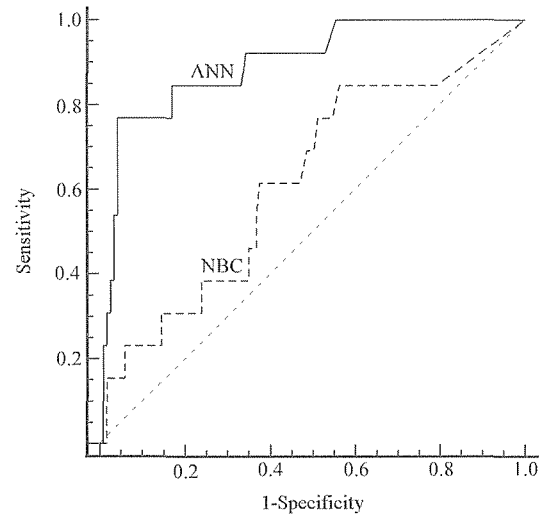


Figure 2. The ROC curves for the ANN and NMC models

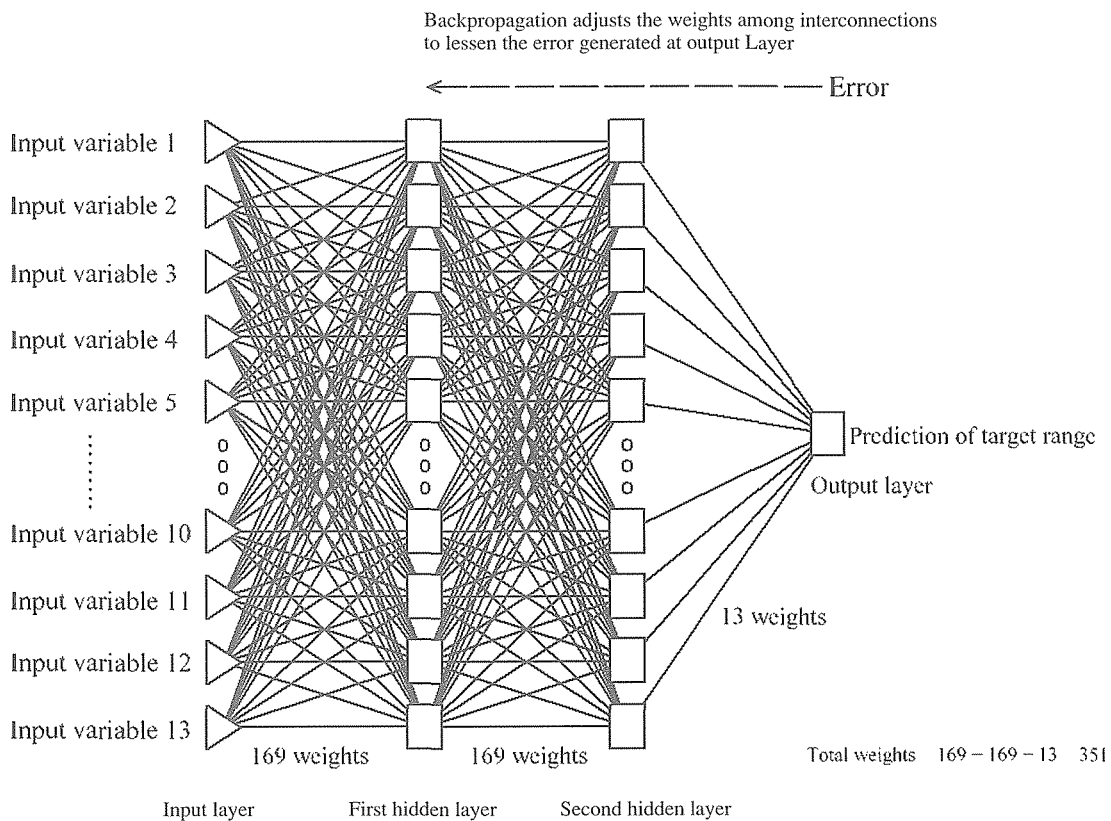


Figure 1. Graphical representation of our ANN model: a multilayer perceptron network with one input layer of 13 neurons (input variable), first hidden layer of 13 neurons, second hidden layer of 13 neurons, and one output layer with 1 neuron (outcome variable)

0.07) in the NBC model, respectively. Both ANN and NBC models had good-fit and similar calibration ($P > 0.05$). However, the ANN model had a lower H-statistic value with a higher P value than the NBC model that signified the good calibration with better fit for the ANN model.

Discussion

Artificial intelligence provides the benefits for identifying complicated correlative interactions among the predictors. It can minimize the use of redundant information in predictors during the training procedure, improve with training, create highly nonlinear decision boundaries, allow maximization of the use of small training samples, and easily explore various forms of biomedical data that may be hidden in the biological nature. Hence, applications of artificial intelligence have been increasingly utilized as a helpful tool for clinical decision making in both diagnosis and treatment [38]. In this study, we tried to implement a simple application of artificial intelligence, the NBC model, to predict the real issue in clinical practice. But the resulting discrimination of the NBC model was not satisfied. The poor performance of the NBC model may be explained by the fact that it utilizes one dangerous hypothesis to make predictions. The NBC model permits the class conditional densities $p(x_j/C_j)$ to be calculated separately for each variable. Subsequently, the NBC model can avoid a multidimensional task and simplify the classification to a number of one-dimensional estimation. The NBC model is a simple probabilistic method that incorporates a strong independence assumption which often has no bearing in reality. Nevertheless, the assumption that the predictors are independent in the NBC analysis is not always true in the biology. The poor predictability of the NBC model approved that the settings of the physiology can not be enacted on the basis of an unreal assumption. Therefore, we applied another complicated application of artificial intelligence, the ANN model, to look for the better resolution.

As expected, the ANN model showed superior predictive performance for plasma concentration prediction of iPTH based on hemodialysis patients' clinical parameters compared to predictions made using a NBC model. These results indicate that the ANN model has some intellectual benefits. It has the ability to recognize complicated correla-

tive interactions between the input variables during the network's training process. Hence, the ANN model can handle clinical information that may be latent in the biological nature of the predictors. However, no clear direct cause-and-effect relation has been shown between input and outcome variables and this black box phenomenon remains a controversial issue. Some objectors claim that, to evaluate the relevance of a decision aid to a particular patient, the physicians need perception into the system's behavior. While it is still disputable whether human experts use hypothetico-deductive reasoning or a "hunch" more frequently in making a medical diagnosis, an accurate second opinion is often beneficial in medical decision-making with or without a detailed understanding of how "dues ex machine" operates [39]. Nonetheless, many clinicians will accept such a tool as an aid to clinical care if this tool can improve the effectiveness and efficiency for the patient care. Clinical predictive performance may be strengthening through the benefit of ANN process that is able to inspect nonlinear interactions among predictors (functional analysis). As a new development of medical technology with the utilization of positive emission tomography, this epochal evolution of functional analysis (ANN approach) will eventually take best advantage to the medical progression if applying these information technologies properly. On the other hand, the NBC model simplifies classification dramatically by allowing the class conditional densities to be calculated separately for each variable and a multidimensional task to be reduced to a number of one-dimensional ones. Although input variables must be independent in the NBC model, the assumption is not always accurate in the biomedical phenomenon and the posterior probabilities do not seem to be greatly affected, especially in regions near decision boundaries. By contrast, no assumption of variable distribution is necessary for the ANN model and the performance of the ANN model is improved over time when more data is collected and used to retrain the model. Therefore, every institute could implement their applications of artificial intelligence to increase the efficiency of their health-care resource usage. To the best of our knowledge after a comprehensive search of the PubMed literature, this is the first comparative analysis of ANN and NBC models in this theme.

Biochemical data monitored at monthly intervals in hemodialysis patients, as well as demographic and historical parameters, can give useful information for adaptations of proper treatment for renal osteodystrophy. However, a pertinent quantity of information concealed in these datasets, which could be very appropriate for the treatment of hemodialysis patients, is systemically lost and easily ignored because of complicated correlations and the nonlinear multi-dimensionality of the interconnections. By using pliable artificial intelligence with a clinical machine learning approach, the ANN model performed a better prediction of the target range of iPTH level in hemodialysis patients in terms of either discrimination or calibration. We also demonstrated how to utilize limited clinical variables to obtain good forecasting. For any predictive model to be useful in making clinical decisions it must use only data that are readily available to the physicians at the time of triage [40]. The challenge is to train a network to identify patterns without overfitting and avoid model complexity for physicians. It is not necessary to add more variables such as a patient's medications or dialytic prescriptions even though some are known to have a consequential impact on iPTH levels.

Parathyroid hormone has a major role in the pathogenesis of renal osteodystrophy, and hyperparathyroidism is also associated with an increase in the relative risk of death, cardiovascular, and fracture-related hospitalization in hemodialysis patients [41]. In most dialysis centers, the measurement interval of plasma iPTH concentration is usually more than three months. Even with popularity of hemodialysis centers and qualified dialytic care in Taiwan, the monitoring interval is usually every six months or longer. We noted that 90% of our patients were out of the target range for plasma iPTH concentration. Despite the fact that the cost of increasing the frequency of plasma iPTH level measurement might be great, we suggest monitoring plasma iPTH concentration frequently in hemodialysis patients for appropriate control of iPTH level. Meanwhile, the inhomogeneous distribution in our study could cause doubts about the generalizability of the ANN to different populations. This is intrinsic in the distribution of covariant patterns, i.e., there were more patients in the non-target group than target group. The distribution of cohorts is also important in the development of models.

Because there are many more covariant patterns for patients in the non-target group, the ANN model will train very well in detecting patients in the non-target group. The challenge is to develop a model to correctly classify patients who will be in the target group [42]. Actually, ANN trained by back propagation learning algorithm can optimize its predictive accuracy to memorize and recognize the biological patterns even though the regulation of PTH in dialysis patients is complicated [43]. As human brain, we can distinguish one red flower from plentiful green leaves without any difficulty; this means that the ratio is not always the problem to our human brain so do the ANN.

The investigations of artificial intelligence to predict iPTH level are limited in clinical medicine. One previous investigation was fostering an interested in time series analysis of PTH - the development and the attempted application of ANN to complex biological systems. Prank et al. used a feedforward ANN to predict future values of the time series of plasma PTH concentration in 12 healthy men, 3 men with idiopathic osteoporosis, and 3 postmenopausal women with osteoporosis [44]. Their data seemed well modeled and showed nonuniform dynamics in that the normal PTH time series alternated between periods of low and high predictability. This fact, that the dynamics of normal PTH secretion has bistable characteristics, implies that the system has a major nonlinear component. This in no way implies that the underlying dynamics are chaotic, but it suggests that the dynamics must be more interesting than the simple linear models in their study. Although we referenced and conducted such a similar research based on nonlinear phenomenon of their findings, we decided to use cross-section biochemical data as predictors since time-series data are rarely used in daily practice. The applicability of a complicated time-series method is to search the latent pattern in mining a dataset which has the characteristic of time sequence while a cross-section method could not provide proper analysis or explain underlying phenomenon in nature. In other respects, their goal and population were different with our study. The other pilot study by our previous results, we developed an ANN model to predict the plasma iPTH concentration in 23 uremic patients on continuous ambulatory peritoneal dialysis [30]. At that time, the results were acceptable with 4 predictors

(plasma calcium, phosphate, alkaline phosphatase concentrations, and calcium-phosphate product). The size of participants and the predictors we selected were the possible reasons. Therefore, we used more predictors and constructed a multilayer perceptron network of the ANN model which was similar to the previous architecture. Through these advances among sample size, participant characteristics, and model architectures, we can confirm that our ANN model not only had appropriate design with adequate predictors but also had good performance to successfully predict the target range of plasma iPTH concentration in hemodialysis patients. However, we look forward to establish a more integrated design to predict the plasma iPTH level not only for the patients with secondary hyperthyroidism but also for those with primary hyperthyroidism or hypothyroidism.

There are some limitations to our study that deserve comments. First, our sample size was relatively small, but by using the leave-one-out cross validation we could overcome this problem. Second, our study was carried out at a single institute, and the models were not evaluated prospectively in other institutions. To test our findings, further studies in different hemodialysis centers can be designed to lessen interinstitutional variation. Third, some physicians may consider that ANN is inconvenient to use. At the present time, with the help of physician-friendly computer hardware and software, the ANN model is in fact easier to use as well as more accurate, and might be applied to different populations of patients. Our future study will concentrate upon the evolution of a web-based platform using the ANN model, merging it into a single-kernel engine for physicians to do real-time forecasting.

Conclusion


The ANN model is better than the NBC model in forecasting the target range of plasma iPTH concentration in hemodialysis patients, and the ANN model might serve as a promising tool to follow or predict the response to all therapeutic interventions for renal osteodystrophy. Our ANN model is not meant to substitute for experienced physicians or specialists, but fill the role of an adjuvant tool for decision-making or for using this predictive information to decide whether further monitoring is necessary. Whether this

application of artificial intelligence can save the cost of frequent iPTH measurements merits further investigations.

References

1. National Kidney Foundation. K/DOQI clinical practice guidelines for bone metabolism and disease in chronic kidney disease. *Am J Kidney Dis* 2003;42:S1-201.
2. Chiu JS, Chong CF, Lin YF, Wu CC, Wang YF, Li YC. Applying an artificial neural network to predict total body water in hemodialysis patients. *Am J Nephrol* 2005;25:507-513.
3. Zini G. Artificial intelligence in hematology. *Hematology* 2005;10:393-400.
4. Grossi E. How artificial intelligence tools can be used to assess individual patient risk in cardiovascular disease: problems with the current methods. *BMC Cardiovasc Disord* 2006;6:20.
5. Forsstrom JJ, Dalton KJ. Artificial neural networks for decision support in clinical medicine. *Ann Med* 1995;27:509-517.
6. Allison JS, Heo J, Iskandrian AE. Artificial neural network modeling of stress single-photon emission computed tomographic imaging for detecting extensive coronary artery disease. *Am J Cardiol* 2005;95:178-181.
7. Ohlsson M. WeAidU-a decision support system for myocardial perfusion images using artificial neural networks. *Artif Intell Med* 2004;30:49-60.
8. Haraldsson H, Ohlsson M, Edenbrandt L. Value of exercise data for the interpretation of myocardial perfusion SPECT. *J Nucl Cardiol* 2002;9:169-173.
9. Toft J, Lindahl D, Ohlsson M, et al. The optimal reference population for cardiac normality in myocardial SPET in the detection of coronary artery stenoses: patients with normal coronary angiography or subjects with low likelihood of coronary artery disease? *Eur J Nucl Med* 2001;28:831-835.
10. Stefaniak B, Cholewinski W, Tarkowska A. Prediction of left ventricular ejection fraction in patients with coronary artery disease based on an analysis of perfusion patterns at rest. Assessment by an artificial neural network. *Nucl Med Rev Cent East Eur* 2004;7:7-12.
11. Scott JA. The lung scan and the abnormal chest X-ray:

- difficult diagnoses. *Nucl Med Commun* 2004;25:1137-1141.
12. Evander E, Holst H, Jarund A, Ohlsson M, Wollmer P, Astrom K, Edenbrandt L. Role of ventilation scintigraphy in diagnosis of acute pulmonary embolism: an evaluation using artificial neural networks. *Eur J Nucl Med Mol Imaging* 2003;30:961-965.
 13. Eng J. Predicting the presence of acute pulmonary embolism: a comparative analysis of the artificial neural network, logistic regression, and threshold models. *AJR Am J Roentgenol* 2002;179:869-874.
 14. Holst H, Mare K, Jarund A, et al. An independent evaluation of a new method for automated interpretation of lung scintigrams using artificial neural networks. *Eur J Nucl Med* 2001;28:33-38.
 15. Stefaniak B, Cholewinski W, Tarkowska A. Application of artificial neural network algorithm to detection of parathyroid adenoma. *Nucl Med Rev Cent East Eur* 2003;6:111-117.
 16. Vesselle H, Turcotte E, Wiens L, Haynor D. Application of a neural network to improve nodal staging accuracy with ¹⁸F-FDG PET in non-small cell lung cancer. *J Nucl Med* 2003;44:1918-1926.
 17. Houston AS, White DR, Sampson WF, Macleod MA, Pilkington JB. An assessment of two methods for generating automatic regions of interest. *Nucl Med Commun* 1998;19:1005-1016.
 18. Hamilton D, O'Mahony D, Coffey J, et al. Classification of mild Alzheimer's disease by artificial neural network analysis of SPET data. *Nucl Med Commun* 1997;18:805-810.
 19. Page MP, Howard RJ, O'Brien JT, Buxton-Thomas MS, Pickering AD. Use of neural networks in brain SPECT to diagnose Alzheimer's disease. *J Nucl Med* 1996;37:195-200.
 20. deFigueiredo RJ, Shankle WR, Maccato A, et al. Neural-network-based classification of cognitively normal, demented, Alzheimer disease and vascular dementia from single photon emission with computed tomography image data from brain. *Proc Natl Acad Sci U S A* 1995;92:5530-5534.
 21. Lee JS, Lee DS, Kim SK, et al. Localization of epileptogenic zones in F-18 FDG brain PET of patients with temporal lobe epilepsy using artificial neural network. *IEEE Trans Med Imaging* 2000;19:347-355.
 22. Shiomi S, Kuroki T, Kuriyama M, et al. Diagnosis of chronic liver disease from liver scintiscans by artificial neural networks. *Ann Nucl Med* 1997;11:75-80.
 23. Hamilton D, Miola UJ, Mousa D. Interpretation of captopril transplant renography using a feed forward neural network. *J Nucl Med* 1996;37:1649-1652.
 24. Yousef M, Nebozhyn M, Shatkay H, Kanterakis S, Showe LC, Showe MK. Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. *Bioinformatics* 2006;22:1325-1334.
 25. Pakhomov SV, Buntrock J, Chute CG. Prospective recruitment of patients with congestive heart failure using an ad-hoc binary classifier. *J Biomed Inform* 2005;38:145-153.
 26. Zhu Y, Hollmen J, Raty R, et al. Investigatory and analytical approaches to differential gene expression profiling in mantle cell lymphoma. *Br J Haematol* 2002;119:905-915.
 27. Zaffalon M, Wesnes K, Petrini O. Reliable diagnoses of dementia by the naive credal classifier inferred from incomplete cognitive data. *Artif Intell Med* 2003;29:61-79.
 28. Demsar J, Zupan B, Kattan MW, Beck JR, Bratko I. Naive Bayesian-based nomogram for prediction of prostate cancer recurrence. *Stud Health Technol Inform* 1999;68:436-441.
 29. de Francisco AL. Secondary hyperparathyroidism: review of the disease and its treatment. *Clin Ther* 2004;26:1976-1993.
 30. Chiu JS, Lin WT, Li YC, Wang YF. Neural network modeling to predict intact parathyroid hormone in uremic patients on continuous ambulatory peritoneal dialysis. *Ann Nuc Med Sci* 2005;18:135-141.
 31. Verweij PJ, Van Houwelingen HC. Cross-validation in survival analysis. *Stat Med* 1993;12:2305-2314.
 32. Molinaro AM, Simon R, Pfeiffer RM. Prediction error estimation: a comparison of resampling methods. *Bioinformatics* 2005;21:3301-3307.
 33. Banerjee R, Das A, Ghoshal UC, Sinha M. Predicting

- mortality in patients with cirrhosis of liver with application of neural network technology. *J Gastroenterol Hepatol* 2003;18:1054-1060.
34. Chiu JS, Lin CS, Yu FC, Li YC. What is the better model in burn patients? *Burns* 2005;31:941.
 35. McNeil BJ, Hanley JA. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. *Med Decis Making* 1984;4:137-150.
 36. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-1293.
 37. Lemeshow S, Hosmer DW, Jr. A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol* 1982;115:92-106.
 38. Ramesh AN, Kambhampati C, Monson JR, Drew PJ. Artificial intelligence in medicine. *Ann R Coll Surg Engl* 2004;86:334-338.
 39. Penny W, Frost D. Neural networks in clinical medicine. *Med Decis Making* 1996;16:386-398.
 40. Das A, Ben-Menachem T, Cooper GS, et al. Prediction of outcome in acute lower-gastrointestinal haemorrhage based on an artificial neural network: internal and external validation of a predictive model. *Lancet* 2003;362:1261-1266.
 41. Block GA, Klassen PS, Lazarus JM, Ofsthun N, Lowrie EG, Chertow GM. Mineral metabolism, mortality, and morbidity in maintenance hemodialysis. *J Am Soc Nephrol* 2004;15:2208-2218.
 42. Ottenbacher KJ, Linn RT, Smith PM, Illig SB, Mancuso M, Granger CV. Comparison of logistic regression and neural network analysis applied to predicting living setting after hip fracture. *Ann Epidemiol* 2004;14:551-559.
 43. Rodvold DM, McLeod DG, Brandt JM, Snow PB, Murphy GP. Introduction to artificial neural networks for physicians: taking the lid off the black box. *Prostate* 2001;46:39-44.
 44. Prank K, Nowlan SJ, Harms HM, et al. Time series prediction of plasma hormone concentration. Evidence for differences in predictability of parathyroid hormone secretion between osteoporotic patients and normal controls. *J Clin Invest* 1995;95:2910-2919.
- 

預測血液透析病患放射免疫量測完整副甲狀腺素之 目標範圍：人工智慧臨床應用

邱建勳¹ 黃世學² 胡宗明³ 陳彥宇⁴ 李友專⁵ 王昱豐¹

¹佛教大林慈濟綜合醫院 核子醫學科 暨慈濟大學醫學系

²佛教大林慈濟綜合醫院 資訊系統管理科

³行政院衛生署玉里醫院 成人精神科

⁴彰化基督教醫院 神經內科

⁵國立陽明大學 生物醫學資訊研究所

背景：在臨床醫學研究中，人工智慧逐漸用來探勘關於生物本質之預測因子與結果的關係。本研究應用兩種人工智慧模型—人工類神經網路與簡易貝氏分類器—預測血液透析病患血漿完整副甲狀腺素濃度之目標範圍。

方法：本研究包括130位穩定之血液透析病患。預測因子包括人口統計資料、相關疾病、各種血液生化檢驗、鈣磷乘積與運鐵蛋白飽和度。根據腎臟疾病結果與品質臨床指引，建議使用放射免疫分析來量測透析病患之血漿完整副甲狀腺素濃度。如果血液透析病患之血漿完整副甲狀腺素濃度介於150 ng/mL與300 ng/mL，表示達到標準範圍，此為目標組；血漿完整副甲狀腺素濃度小於150 ng/mL或大於300 ng/mL，表示未達標準範圍，則為非目標組。達到目標組與否，則為人工智慧模型之二元相依變數。另外，使用鐮刀式交叉驗證之反覆取樣技術，可以避免模型不易普遍化的問題。為了比較人工類神經網路與簡易貝氏分類器的效能，必須同時使用接收器運作指標曲線下面積與適合度檢定兩種統計量作為鑑別與校正的比較。

結果：人工類神經網路與簡易貝氏分類器的接收器運作指標曲線下面積分別是 0.90 ± 0.06 與 0.62 ± 0.08 ，因此人工類神經網路的鑑別能力較為優異 ($P < 0.01$)。而兩者的適合度統計量分別是6.88 ($P = 0.08$) 與6.97 ($P = 0.07$)，因為人工類神經網路有比較低的適合度與比較高的P值，所以人工類神經網路擁有比較好的適合度。

結論：人工類神經網路可以用來預測血液透析病患血漿完整副甲狀腺素濃度的目標範圍，是一種具有發展潛力的預測工具。

關鍵詞：人工智慧，血液透析，簡易貝氏分類器，類神經網路，副甲狀腺素

核子醫誌2006;19:149-159