J Med Syst (2012) 36:2057–2061
DOI 10.1007/s10916-011-9668-3

ORIGINAL PAPER

Artificial Intelligence Models for Predicting Iron Deficiency Anemia and Iron Serum Level Based on Accessible Laboratory Data

Iman Azarkhish · Mohammad Reza Raoufy · Shahriar Gharibzadeh

Received: 6 January 2011 / Accepted: 16 February 2011 / Published online: 19 April 2011

Abstract Iron deficiency anemia (IDA) is the most common nutritional deficiency worldwide. Measuring serum iron is time consuming, expensive and not available in most hospitals. In this study, based on four accessible laboratory data (MCV, MCH, MCHC, Hb/RBC), we developed an artificial neural network (ANN) and an adaptive neuro-fuzzy inference system (ANFIS) to diagnose the IDA and to predict serum iron level. Our results represent that the neural network analysis is superior to ANFIS and logistic regression models in diagnosing IDA. Moreover, the results show that the ANN is likely to provide an accurate test for predicting serum iron levels with high accuracy and acceptable precision.

Keywords Artificial neural network · Iron level · Neuro-fuzzy · Diagnosis · Modeling

Introduction

Iron deficiency (ID) is a common disorder, affecting more than 1 billion people worldwide [1]. It may reduce work capacity in adults [2] and reduce mental development in children [3]. Clinically, the main problem is to diagnose true iron deficiency from other causes of iron-deficient erythropoiesis, such as the anemia of chronic disease [4]. Therefore, it is important to assess the blood iron level in patients who are suspected of anemia. The simplest methods available for assessing body iron concentrations are biochemical measurements of the serum iron concentration, transferrin saturation, and ferritin concentration [5], which are expensive for the patient, time consuming and not reachable in every place or hospital. Therefore, presenting an accurate, simple, and inexpensive method for diagnosing the ID and measuring the iron serum can be truly helpful.

In this research, we developed an artificial neural network (ANN), based entirely on routine laboratory data, in order to anticipate the serum iron level with an acceptable precision.

ANN is a mathematical method for analyzing nonlinear functions [6]. The capabilities and advantages of ANNs are due to their special features including adaptive and parallel processing [7]. Each neuron in ANN receives inputs either from other neurons or from an external stimulus. The weighted sum of these inputs passes through an optional function and the resulted argument is applied to an activation function that finally yields the output of the neurons.

The proposed neuro-fuzzy model in ANFIS (Adaptive Neuro-Fuzzy Inference System) is a multilayer neural network-based fuzzy system with five layers. In this connectionist structure, the input and output nodes represent the descriptors and the response, respectively. In the hidden layers, there are nodes functioning as membership functions (MFs) and rules.

The ANFIS and ANN models are generated in this research to determine which model is better for diagnosing Iron Deficiency Anemia (IDA) outcomes.

Materials and methods

Patient selection

The study was conducted on patients at Tehran Shariati hospital in 2010. The total number of patients enrolled in

I. Azarkhish · M. R. Raoufy · S. Gharibzadeh (✉)
Amirkabir University of Technology,
Tehran, Iran
e-mail: gharibzadeh@aut.ac.ir

this study was 203 (92 males; 111 females; mean age: 55.8± 17.78 years old). Based on examination and history, cases with hemorrhage and hemodialysis patients were excluded from this study. Patients' blood samples were collected and cell blood count and serum iron were measured in a reliable laboratory. The level that has been considered as a margin of anemia and normal state was 40 mcg/dl for all of the patients. The patients were randomly divided into train and test groups. The data of 149 patients was used to train the models. The remainder of the whole patients (54 cases), was used to test the models.

Effective variables

In a preliminary statistical study, among the information provided from patient's blood test, those patch of data who had the most effect on blood iron were denoted (Table 1). These laboratory data were used as inputs of the models. These inputs were: Mean Cell Volume (MCV) (I1), Mean Cell Hemoglobin (MCH) (I2), Mean Cell Hemoglobin Concentration (MCHC) (I3) and Hemoglobin (Hb) to Red Blood Cell (RBC) ratio (I4). The outputs of the models were supposed to be the presence (U=0) or absence (U=1) of the IDA. Also, the ability of models to predict the serum iron concentration was evaluated.

Artificial neural network analysis

Several reviews on ANN have been published recently [8–16]. In order to design the ANN, we used the Neural Network Toolbox of the MATLAB 7 software. The ANN used in this study was a standard back-propagation neural network with three layers: an input layer, a hidden layer and an output layer. The input, hidden, and output layers contained four, 15 and one neurons, respectively. We determined the number of the network layers, hidden neurons and the stopping criteria via trial-and-errors process, because no commonly accepted theory exists for determining the optimum number of neurons in the hidden layer [17, 18]. Generally, the transfer functions are sigmoidal, hyperbolic tangent, or linear function, of which the sigmoidal function is the most widely used [19]. In this study the tansig and purelin functions were used for hidden layers and output layer respectively. We used the function newff to create the network object in training feed forward network and the Levenberg–Marquardt (trainlm) algorithm to train the back-propagation network. To simplify the problem for the network, we preprocessed the input and target values and mapped them in to the interval [1,−1]. The ANN was trained 1500 times (epochs). It takes less than 1 min. The mean standard error was 2.46203e-09.

ANFIS analysis

To predict the IDA, we developed an ANFIS model using MATLAB 7 software (fuzzy logic toolbox-ANFIS editor GUI). Training and testing datasets were the same as those used in ANN models. The number of input membership functions (MF) in this model was four, for each input. The MF type for input and output was psigmf and linear respectively. The model was trained 40 times (epochs) in Hybrid method. It takes about 7 min with a Pentium IV PC computer, having 2.6 Giga Hz CPU. The average testing error was 0.24269.

Logistic and linear regression analysis

Logistic regression and linear regression models were developed using SPSS for Windows 18.0 (SPSS Inc). For training and testing the models, we used datasets the same as those used in two other models.

Comparison of the models

The testing set was entered to all three models (ANN, ANFIS and Logistic Regression) as inputs and the diagnosis of IDA was assessed as an output. Accuracy (the number of correct predictions divided by total predictions), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood

Table 1 The distribution of the dataset between patients whit absence or presence of Iron Deficiency Anemia

	Presence of iron deficiency anemia	Absence of Iron Deficiency Anemia	*p* value
RBC	3.74(1.11)	3.69(0.78)	0.404
Hb	10.08(2.40)	10.51(2.08)	0.702
HCT	32.06(7.58)	30.91(6.06)	0.968
MCV	87.76(8.23)	89.87(7.22)	0.015
MCH	27.88(3.07)	28.52(2.55)	0.009
MCHC	31.74(1.47)	60.26(1.43)	0.024
Hb/RBC	2.79(0.31)	2.85(0.25)	0.009
Age	55.8(17.78)	53.6(19.11)	0.511

Data are mean (SD)

Table 2 The characteristics of patients: comparison of training and testing sets

	Training set(n=149)	Testing set(n=54)
Iron Deficiency Anemia	30(20%)	31(57%)
MCV	89.26(7.43)	89.17(8.04)
MCH	28.44(2.67)	28.03(2.88)
MCHC	31.84(1.40)	31.44(1.51)
Hb/RBC	2.85(0.27)	2.80(0.29)

Data are mean (SD) or numbers (percentage), *IDA* Iron Deficiency Anemia

ratio positive (LR+) and likelihood ratio negative (LR−) of these models were calculated. The ROC curves could measure the discriminating power of these diagnosis models [20]. The ROC curve shows the tradeoff between sensitivity and specificity (any increase in sensitivity will be accompanied by a decrease in specificity). Discriminatory power is measured by AUC, which is a particularly important metric for evaluating prediction tools because it is the average sensitivity over all possible specificities. AUC may range from 0 to 1, with area of 1.0 representing perfect discrimination and an area of 0.5 representing what is expected by chance alone [20].

To assess the blood iron magnitude as a result of ANN model, we calculated mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and R-squared (R^2). Also, we developed a linear regression model, as a classic test, and compared it with ANN results.

Results

In this research, 203 patients were studied. 149 patients allocated at random to the training set and the remaining 54 to the testing set (Table 2). The ANN was able to diagnose

Table 3 The ANN (Artificial Neural Network), ANFIS (Adaptive Neuro-Fuzzy Inference System) and logistic regression performance to diagnose IDA (Iron Deficiency Anemia)

		IDA		Total
		Presence (N=31)	Absence (N=23)	
ANN	Positive	30	1	31
	Negative	1	22	23
ANFIS	Positive	27	1	28
	Negative	4	22	26
Logistic Regression	Positive	29	18	47
	Negative	2	5	7

Table 4 Comparison of predictive performance of ANN (artificial neural network), ANFIS (Adaptive Neuro-Fuzzy Inference System), logistic regression to diagnose IDA (Iron Deficiency Anemia)

	ANN	Logistic Regression	ANFIS
Accuracy (%)	96.29	62.96	90.74
Sensitivity (%)	96.8	93.5	87.1
Specificity (%)	95.6	21.7	95.6
PPV (%)	96.7	61.7	96.4
NPV (%)	95.6	71.4	84.6
LR+	22	1.19	19.8
LR−	0.03	0.3	0.13

PPV positive predictive value, *NPV* negative predictive value, *LR+* likelihood ratio positive, *LR−* likelihood ratio negative

the IDA and also the absence of IDA with the accuracy of 97% for patients with IDA and 96% for patient without it. The accuracy of logistic regression model for this diagnosis is 93% for patients with IDA and 22% for patients without IDA. These results were 87% and 95% respectively for the ANFIS model. Table 3 shows the ANN, logistic regression and ANFIS performance in diagnosing the presence or absence of IDA in patients. Some predictive performance indices as sensitivity, specificity, PPV, NPV, LR+, LR−, and AUC are compared among ANN, logistic regression and ANFIS models in Table 4. Figure 1 represents the ROC curves for these three models. Results of the comparison of the ROC curves between three models are shown in Table 5.

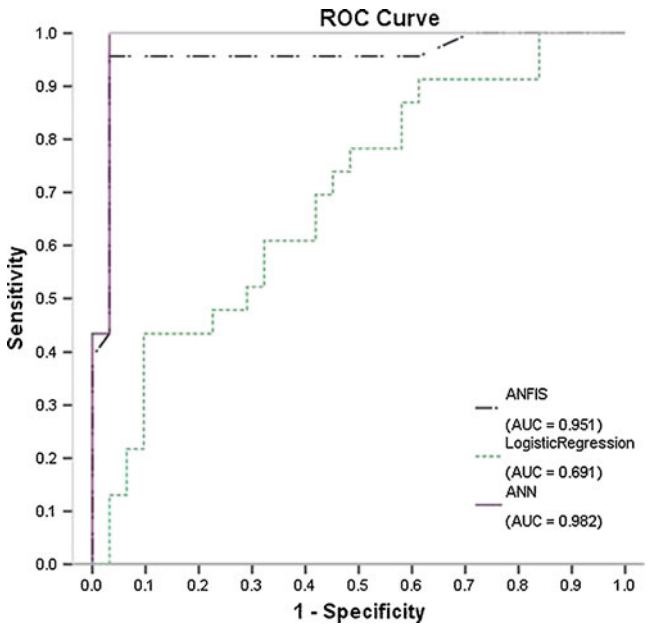

Fig. 1 Comparison of ROC (receiver operating characteristic) curves to diagnose IDA (Iron Deficiency Anemia). ANN, artificial neural network; ANFIS, Adaptive Neuro-Fuzzy Inference System

Table 5 Comparison of ROC (Receiver Operating Characteristic) curves to diagnose IDA (Iron Deficiency Anemia)

	AUC	Standard error	95% confidence interval	p value
ANN	0.982	0.019	0.000–1.000	<0.00001
ANFIS	0.954	0.033	0.000–1.000	<0.00001
Logistic Regression	0.691	0.073	0.549–0.834	0.017

ANN, artificial neural network; ANFIS, Adaptive Neuro-Fuzzy Inference System; AUC, area under the ROC curve

For predicting the serum iron level, the mean and standard deviation of ANN and real outputs were 50.8518±36.07 and 50.8412±36.06, respectively. Table 6 shows mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and R-squared (R^2) for the predicted values of the blood iron level for the ANN model in comparison with linear regression model.

Discussion

In this study, based entirely on routine and inexpensive laboratory data, we developed ANN, ANFIS and Logistic Regression models in order to diagnose IDA in patients. Also we tested ANN and linear regression for predicting the serum iron level in patients.

In designing the ANN, we used accessible variables that were not directly diagnostic of IDA. For validating the ANN, we used testing data which were not used for training.

The AUC is a measure of a model's discriminatory power. According to the observation by Swets et al. [21], an AUC of ≥0.7 is diagnostically useful. The ANN model significantly outperform the logistic regression model (AUC=0.982 vs. 0.691, $p<0.00001$) in the testing set. The ANN model also has better simultaneous sensitivity and specificity than both ANFIS and logistic regression models. As it shown in Table 6, linear regression has remarkable error. The ANFIS model can predict the IDA with high accuracy, but it is not a preferable model to predict the exact values of iron magnitude. The basic structure of the ANFIS is a model that maps input characteristics to input membership functions, input membership function to rules, rules to a set of output characteristics, output characteristics to output membership functions, and the output membership function to a single-valued output or a decision associated with the output. The if-then rule statements of this model are used to formulate the conditional statements that comprise fuzzy logic. Although these rules are powerful in distinguishing categorized variables with a good accuracy, they cannot predict accurate amounts of continuous variables. A neural network approach is also preferable in that ANNs are model independent and flexible in being able to use mixes of categorical and continuous variables. ANNs also have the advantage that they can learn to anticipate arbitrarily complex nonlinear relationships between independent and dependent variables by including more processing elements in the hidden layer or more hidden layers in the ANN. These advantages make the ANN a more robust paradigm for application to a real-world setting [17, 22].

The real-time use of the ANN is not difficult. The number of hospitals that have an electronic medical record is growing rapidly [19]. On the other hand, the ferritin and iron serum are not available in all hospitals. The data used by our ANN is the standard information routinely collected in a simple blood test. Once trained, the ANN could reside in the background of the clinical information systems. Once entered into the electronic record, these data could then be used by the ANN to generate the probability of the predicted outcome. All conventional statistical models such as regression-based models presume that the data is linear; however, most biological data are nonlinear [6]. Neural networks-based models are more suited for pattern recognition in multidimensional nonlinear interactions [23]. ANN accuracy could also be continuously improved over time because it can constantly be retrained as more patients are accumulated. ANN models are dynamic and continue to learn each time a new patient's data are entered [6]; in contrast, a logistic regression model is not dynamic in that the derivation cohort for most models would not change over time. Portability will be a critical factor to the future use of the ANN in this setting [24].

Overall, using the routine laboratory data, the designed model was able to diagnose the IDA in patients. This model is a novel high accuracy, non-invasive, inexpensive, and rapid method, which can be used clinically.

There are a number of limitations to this study that need to be addressed. First, the ANN was not tested clinically. It is not clear how hematologists will respond to the ANN results. Second, this study was conducted at a single institution. These findings must be corroborated on patients from multiple locations.

Table 6 ANN (Artificial Neural Network) accuracy to predict the serum iron level in comparison with linear regression

	MAE	MAPE	RMSE	R^2
ANN model	0.022	230.115	0.136	0.93
Linear Regression Model	7.042	12.401	9.175	0.092

ANN artificial neural networks, *MAE* mean absolute error, *MAPE* mean absolute percentage error, *RMSE* root mean square error, R^2 determination coefficient

References

1. Grosbois, B., Decaux, O., Cador, B., Cazalets, C., and Jego, P., Human iron deficiency. *Bull. Acad. Natl Méd.* 189:1649–1663, 2005.
2. Haas, J. D., and Brownlie, T., IV, Iron deficiency and reduced work capacity: a critical review of the research to determine a causal relationship. *J. Nutr.* 131(2 suppl):676S–88S, 2001. discussion 688S–90S.
3. Halterman, J. S., Kaczorowski, J. M., Aligne, C. A., Auinger, P., and Szilagyi, P. G., Iron deficiency and cognitive achievement among school-aged children and adolescents in the United States. *Pediatrics* 107:1381–6, 2001.
4. Cook, J. D., and Skikne, B. S., Iron deficiency: definition and diagnosis. *J. Intern. Med.* 226(5):349–55, 1989.
5. Worwood, M., The laboratory assessment of iron status: an update. *Clin. Chim. Acta* 259:3–23, 1997.
6. Cross, S. S., Harrison, R. F., and Kennedy, R. L., Introduction to neural networks. *Lancet* 346:1075–1079, 1995.
7. Dariani, S., Keshavarz, M., Parviz, M., Raoufy, M. R., and Gharibzadeh, S., Modeling force-velocity relation in skeletal muscle isotonic contraction using an artificial neural network. *Biosystems* 90(2):529–34, 2007.
8. Forsstrom, J. J., and Dalton, K. J., Artificial neural networks for decision support in clinical medicine. *Ann. Med.* 27:509–17, 1995.
9. Rodvold, D. M., McLeod, D. G., Brandt, J. M., Snow, P. B., and Murphy, G. P., Introduction to artificial neural networks for physicians: taking the lid off the black box. *Prostate* 46:39–44, 2001.
10. Boutsinas, B., and Vrahatis, M., Artificial nonmonotonic neural networks. *Artif. Intell.* 132:1–38, 2001.
11. Lisboa, P., A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw.* 15:11–39, 2002.
12. Ramesh, A. N., Kambhampati, C., Monson, J. R., and Drew, P. J., Artificial intelligence in medicine. *Ann. R. Coll. Surg. Engl.* 86:334–8, 2004.
13. Sharpe, P. K., and Caleb, P., Artificial neural networks within medical decision support systems. *Scand. J. Clin. Lab. Invest. Suppl.* 219:3–11, 1994.
14. Tafeit, E., and Reibnegger, G., Artificial neural networks in laboratory medicine and medical outcome prediction. *Clin. Chem. Lab. Med.* 37:845–53, 1999.
15. Winkler, D. A., Neural networks as robust tools in drug lead discovery and development. *Mol. Biotechnol.* 27:139–68, 2004.
16. Dytch, H. E., and Wied, G. L., Artificial neural networks and their use in quantitative pathology. *Anal. Quant. Cytol. Histol.* 12:379–93, 1990.
17. Tu, J. V., Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* 49:1225–1231, 1996.
18. Eftekhar, B., Mohammad, K., Ardebili, H. E., Ghodsi, M., and Ketabchi, E., Comparison of artificial neural network and logistic regression models for prediction of mortality in head trauma based on initial clinical data. *BMC Med. Inform. Decis. Mak.* 5:3, 2005.
19. Raoufy, M. R., Vahdani, P., Alavian, S. M., Fekri, S., Eftekhari, P., and Gharibzadeh, S., A novel method for diagnosing cirrhosis in patients with chronic hepatitis B: artificial neural network approach. *J. Med. Syst.* 35(1):121–6, 2011. Epub 2009 Jul 21.
20. Hanley, J. A., and McNeil, B. J., The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36, 1982.
21. Swets, J. A., Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293, 1988.
22. Lin, C. S., Li, Y. C., Mok, M. S., Wu, C. C., Chiu, H. W., and Lin, Y. H., Neural network modeling to predict the hypnotic effect of propofol bolus induction. *Proc. AMIA Symp*. 450–454, 2002.
23. Ghoshal, U. C., and Das, A., Models for prediction of mortality from cirrhosis with special reference to artificial neural network: a critical review. *Hepatol. Int.* 2(1):31–8, 2008. Epub 2007.
24. Chong, C. F., Li, Y. C., Wang, T. L., Chang, H., Stratification of adverse outcomes by preoperative risk factors in coronary artery bypass graft patients: an artificial neural network prediction model. *AMIA Annu. Symp. Proc*. 160–164, 2003.