[Turkish Journal of Medical Sciences](https://journals.tubitak.gov.tr/medical)

Volume 42 | Number 5 Article 20

1-1-2012

Prediction of mortality in stroke patients using multilayer perceptron neural networks

NECDET SÜT

YAHYA ÇELİK

Recommended Citation

SÜT, NECDET and ÇELİK, YAHYA (2012) "Prediction of mortality in stroke patients using multilayer perceptron neural networks," *Turkish Journal of Medical Sciences*: Vol. 42: No. 5, Article 20.
https://doi.org/10.3906/sag-1105-20
Available at: https://journals.tubitak.gov.tr/medical/vol42/iss5/20

Original Article

Turk J Med Sci
2012; 42 (5): 886-893
© TÜBİTAK
E-mail: medsci@tubitak.gov.tr
doi:10.3906/sag-1105-20

Prediction of mortality in stroke patients using multilayer perceptron neural networks

Necdet SÜT[1], Yahya ÇELİK[2]

Aim: We aimed to predict mortality in stroke patients by using multilayer perceptron (MLP) neural networks.

Materials and methods: A data set consisting of 584 stroke patients was analyzed using MLP neural networks. The effect of prognostic factors (age, hospitalization time, sex, hypertension, atrial fibrillation, embolism, stroke type, infection, diabetes mellitus, and ischemic heart disease) on mortality in stroke were trained with 6 different MLP algorithms [quick propagation (QP), Levenberg-Marquardt (LM), backpropagation (BP), quasi-Newton (QN), delta bar delta (DBD), and conjugate gradient descent (CGD)]. The performances of the MLP neural network algorithms were compared using the receiver operating characteristic (ROC) curve method.

Results: Among the 6 algorithms that were trained with the MLP, QP achieved the highest specificity (81.3%), sensitivity (78.4%), accuracy (80.7%), and area under the curve (AUC) (0.869) values, while CGD achieved the lowest specificity (61.5%), sensitivity (58.7%), accuracy (60.8%), and AUC (0.636) values. The AUC of the QP algorithm was statistically significantly higher than the AUCs of the QN, DBD, and CGD algorithms ($P < 0.05$ for all of the pairwise comparisons).

Conclusion: The MLP trained with the QP algorithm achieved the highest specificity, sensitivity, accuracy, and AUC values. This can be helpful in the prediction of mortality in stroke.

Key words: Multilayer perceptron neural networks, stroke, mortality, algorithm

Introduction

Stroke is a disease that is the world's third most common cause of death behind diseases of the heart and cancer. It occurs when a blood vessel is blocked by a clot or bursts; then part of the brain cannot get the blood it needs, and so it starts to die (1). Medical diagnosis and outcome prediction of diseases are complex processes. Neural networks can be used as classification or prediction tools in medical decision making in many diseases. One of the most popular neural network models is the multilayer perceptron (MLP) neural network, because of its clear architecture and comparably simple algorithm (2). The MLP can be trained with different classification algorithms, and these algorithms can produce different results. In many medical studies it has been trained with different algorithms as a classification or prediction tool, such as for diagnosing coronary artery disease (3), antenatal fetal risk assessment (4), identification of responsiveness to interferon therapy in multiple sclerosis patients (5), prediction of atrial fibrillation termination (6), prediction of influenza vaccination outcome (7), prediction of essential hypertension (8), and diagnosis of the obstructive sleep apnea syndrome from nocturnal oximetry (9). In a study by İçer et al., the MLP was trained with 3 algorithms in the diagnosis of cirrhosis disease (10). In a study by Güler and Übeyli, the MLP was trained with 4 algorithms in the diagnosis of partial epilepsy (11). In a study by Süt and Şenocak (3), the MLP was trained

Received: 10.05.2011 – Accepted: 01.08.2011

[1] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Trakya University, Edirne - TURKEY
[2] Department of Neurology, Faculty of Medicine, Trakya University, Edirne - TURKEY

Correspondence: *Necdet SÜT, Department of Biostatistics and Medical Informatics, Faculty of Medicine, Trakya University, Edirne - TURKEY*
E-mail: nsut@trakya.edu.tr

with 4 algorithms in the diagnosis of coronary artery disease and compared with logistic regression and quadratic discriminant analyses. The MLP has shown higher classification or prediction results than some statistical analyses (e.g. logistic regression analysis and discriminant analysis) (3,4,8). A search of the literature did not show the classification properties of neural networks in the prediction of mortality in stroke.

In this study, we aimed to examine the performance of an MLP trained with 6 different algorithms [quick propagation (QP), Levenberg-Marquardt (LM), backpropagation (BP), quasi-Newton (QN; BFGS method), delta bar delta (DBD), conjugate gradient descent (CGD) in the prediction of mortality in stroke.

Materials and methods

Multilayer perceptron neural networks

Artificial neural networks resemble the human brain in the following 2 ways: they obtain knowledge through learning, and the knowledge is stored within interneuron connection strengths known as synaptic weights (12). MLPs are among the most popular neural network architectures. They were originally described by Rumelhart and McClelland in 1986 (13) and were discussed by Bishop (14) at length in his neural network textbook (15). The MLP is known as a supervised network due to the fact that it requires a desired output in order to learn. MLPs consist of an input layer with neurons (input variables), an output layer with neurons (dependent variables), and 1 or more hidden layers containing neurons to help capture the nonlinearity in the data (11).

Gurgen et al. (4) described the MLP function as follows: "the basic idea of the technique is to efficiently compute partial derivatives of an approximating function F(**w**; **x**), realized by the network with respect to all the elements of the adjustable weight vector **w**, for a given value of input vector **x**, and output vector **y**. The difference between the network output and the supervisor output is minimized according to predefined error function (performance criterion) such as mean square error (MSE). This function helps to place the discriminator function to the right location and position." The formula for the MSE is:

$$MSE = \sum_i (y_i - F(w, x_i))^2$$

The MLP was trained with 6 different algorithms in this study. In the following section, their properties are explained briefly.

Quick propagation algorithm

The QP algorithm seems more inclined to instability and to getting stuck in local minima than BP, and these tendencies may determine whether QP is more appropriate for a particular problem (3,15). Weight changes were calculated using following formula in QP:

$$\triangle w(t) = \frac{s(t)}{s(t-1) - s(t)} \triangle w(t-1)$$

This formula is numerically unstable if *s(t)* is very close to, equal to, or greater than *s(t-1)*. In these cases, weight changes are calculated using the following formula:

$$\triangle w(t) = a \triangle w(t-1)$$

Here, *a* is the acceleration coefficient (15).

Levenberg-Marquardt algorithm

The LM is a least squares estimation algorithm based on the maximum neighborhood idea (16). It is an advanced nonlinear optimization algorithm that uses the following formula for weight updating. Consider the nonlinear model fitting $y = f(\theta, x)$ with the given data X_i and Y_i, $i = 1,\dots, m$, where X_i is of dimension k and θ is of dimension n. The LM method seeks θ, the solution of θ (locally) minimizing:

$$g(\theta) = \sum_{i=1}^{m} (Y_i - f(\theta, X_i))^2$$

The LM algorithm finds the solution by applying the following routine iteratively:

$$\theta_{j+1} = \theta_j - (J'J + \lambda D)^{-1} J' (Y - f(\theta, X_i))$$

where Y is the m × 1 vector containing $Y_1,\dots,Y_m$; X is the m × k matrix containing $X_1,\dots,X_m$; J is the m × n Jacobian matrix for f(θ,x) with respect to θ; and D is the n × n diagonal matrix to adjust scale factors (15).

Backpropagation algorithm

The BP algorithm was proposed by Rumelhart et al. in 1986 (17). It is one of the simplest and most general methods for the supervised training of MLPs (2,17,18). It uses the following formula for weight updating:

$$\triangle w_{ij}(t) = \eta\delta_j o_i + \alpha\triangle w_{ij}(t-1)$$

where η is the learning rate, δ is the local error gradient, α is the momentum coefficient, and o_i is the output of the *i*th unit (15).

Quasi-Newton algorithm

The QN algorithm is an advanced method of training MLPs. It calculates the error gradient as the sum of the error gradients on each training case. It maintains an approximation to the inverse Hessian matrix (15).

Delta bar delta algorithm

The DBD algorithm is an alternative to the BP algorithm. The average error gradient across all of the training cases is calculated on each epoch, and then the weights are updated once at the end of the epoch. It uses the following formula for weight updating:

$$\bar{\delta}(t) = (1-\theta)\delta(t) + \theta\bar{\delta}(t-1)$$

where *δ(t)* is the derivative of the error surface and θ is the smoothing constant.

The learning rate of each weight is updated using:

$$\triangle\eta(t) = \begin{cases} \mathrm{K} & \bar{\delta}(\mathrm{t}-1)\delta(\mathrm{t}) > 0 \\ -\phi\eta(\mathrm{t}) & \bar{\delta}(\mathrm{t}-1)\delta(\mathrm{t}) < 0 \\ 0 & \bar{\delta}(\mathrm{t}-1)\delta(\mathrm{t}) = 0 \end{cases}$$

where κ is the linear increment factor and φ is the exponential decay factor (15).

Conjugate gradient descent algorithm

The CGD algorithm calculates the error gradient as the sum of the error gradients on each training case. The initial search direction is given by:

$$d_o = -g_o$$

Subsequently, the search direction is updated using the Polak-Rebiere formula:

$$d_{j+1} = -g_{j+1}\beta_j d_j$$

$$\beta_j = \frac{\mathrm{g}_{j+1}^T(\mathrm{g}_{j+1} - \mathrm{g}_j)}{\mathrm{g}_j^T \mathrm{g}_j}$$

If the search direction is not downhill, the algorithm restarts using the line of steepest descent. It restarts regardless after W directions (where *W* is the number of weights), as at that point, the conjugacy has been exhausted. Line searches are conducted using Brent's iterative line search procedure, which utilizes a parabolic interpolation to locate the line minima extremely quickly (15).

Data and architecture of MLP

The hospital records of stroke patients were reviewed retrospectively using hospital automation software. We identified 584 stroke patients. They were classified as living or deceased. Among the 10 independent variables (age, hospitalization time, sex, hypertension, atrial fibrillation, embolism, stroke type, infection, diabetes mellitus, and ischemic heart disease), 8 variables were found to be prognostic factors on mortality in stroke using univariate statistical analysis (Student's t-test, Mann-Whitney U test, or chi-square test).

The MLP used in this study consisted of 3 layers including an input layer, a hidden layer, and an output layer. Eight input variables were used in the mortality prediction of stroke. The hidden layer consisted of 2 nodes, which were determined using trial and error. The most appropriate network configuration was 8 neurons for each hidden layer. The output layer consisted of 2 nodes, which corresponded to stroke outcome (living vs. deceased). We then architectured our MLP trained with the QP, LM, BP, QN, DBD, and CGD algorithms using Statistica 7.0 (StatSoft Inc., Tulsa, OK, USA) neural network toolbox. The values of the tuning parameters of the algorithms are shown in Table 1. Of the 584 patients, 408 (70%) were used for training and 176 (30%) were used for testing processes. Area under the curve (AUC) was computed using receiver operating characteristic (ROC) curves in order to compare performances of the algorithms and then AUCs were compared using *z* statistics. AUCs were compared using MedCalc statistical software version 11.1.1.0 (MedCalc Software, Mariakerke, Belgium). A flow chart of the research design is shown in Figure 1.

Results

Demographic and clinical characteristics of the patients are shown in Table 2. The mean age was significantly higher in the living patient group than in the deceased patient group. Hospitalization time, sex, hypertension, atrial fibrillation, embolism,

Table 1. The values of the tuning parameters of the algorithms.

Algorithms	Values of the tuning parameters
QP	Learning rate (α) = 0.01, Acceleration = 2, Add Gaussian noise = 0.1
LM	Decay factor = 0.01, Scale factor = 1.0
BP	Learning rate (α) = 0.01, Momentum (μ) = 0.3
QN	Decay factor = 0.01, Scale factor = 1.0
DBD	Learning rate (α): Initial = 0.01, Increment = 0.01, Decay = 0.8, Smoothing = 0.5, Add Gaussian noise = 0.1
CGD	Decay factor = 0.01, Scale factor = 1.0

stroke type, and infection were significantly different between the living and deceased stroke patients. However, diabetes mellitus and ischemic heart disease were not significantly different between the living and deceased stroke patients.

The confusion matrix of the testing processes is shown in Table 3. The testing results of the MLP neural networks (MLPNNs) trained with the QP, LM, BP, QN, DBD, and CGD algorithms to predict mortality in stroke are shown in Table 4. According to the testing results, mortality (deceased) status in stroke was predicted with accuracy rates varying from 60.8% to 80.7% by the MLPNNs trained with 6 different algorithms. Among the 6 algorithms, the QP algorithm achieved the highest accuracy rate (80.7%), while the CGD algorithm achieved the lowest accuracy rate (60.8%). When we investigated the sensitivity and specificity values, similarly, the QP algorithm achieved the highest rates (sensitivity = 78.4%, specificity = 81.3%), while the CGD algorithm achieved the lowest rates (sensitivity = 58.7%, specificity = 61.5%).

The ROC curves of the MLP trained with the QP, LM, BP, QN, DBD, and CGD algorithms are shown in Figure 2. The AUCs of the MLPNNs obtained from the ROC analyses are shown in Table 5. The AUCs for the algorithms were calculated as 0.869 for QP ($P < 0.001$), 0.853 for LM ($P < 0.001$), 0.817 for BP (P

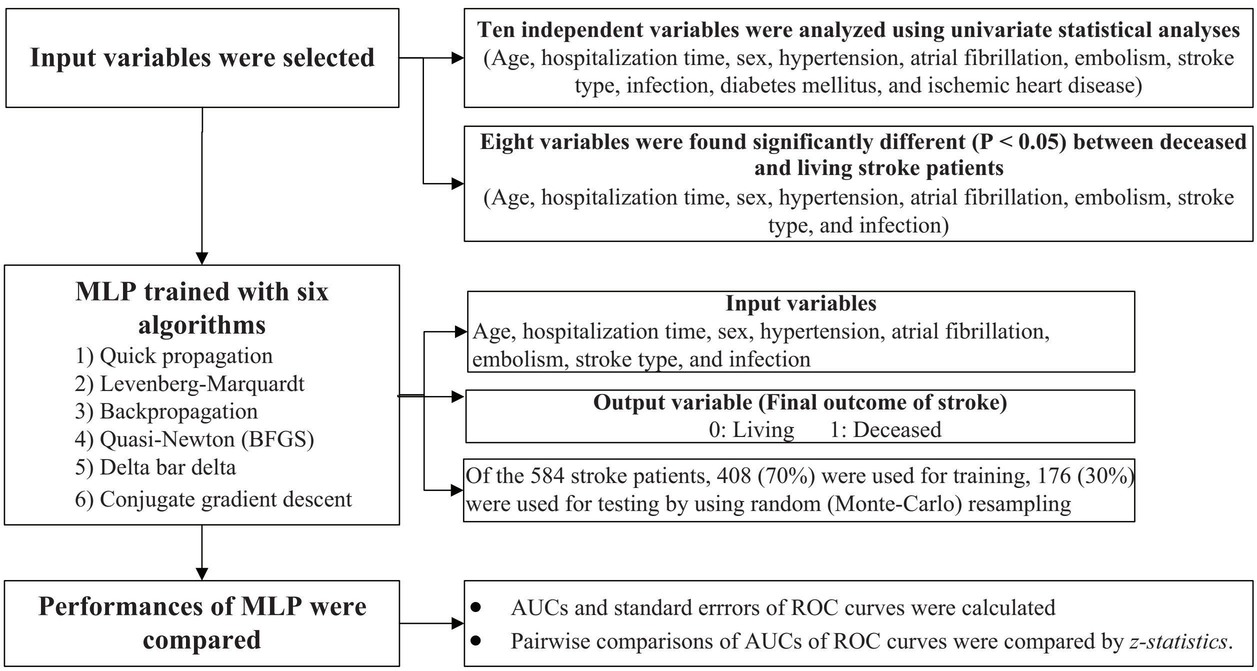

Figure 1. Flow chart of the research design.

Table 2. Demographic and clinical characteristics of the patients.

	Living (n = 457)	Deceased (n = 127)	P
Age	68.7 ± 10.9	71.3 ± 11.9	0.022[#]
Hospitalization time (days)	10.7 ± 6.2	9.5 ± 9.8	<0.001[‡]
Sex (male/female)	246/211	55/72	0.036[†]
Hypertension (–/+)	138/319	23/104	0.007[†]
Atrial fibrillation (–/+)	382/75	93/34	0.008[†]
Embolism (–/+)	371/86	92/35	0.032[†]
Stroke type (ischemic/hemorrhagic)	408/49	81/46	<0.001[†]
Infection (–/+)	330/127	44/83	<0.001[†]
Diabetes mellitus (–/+)	339/118	92/35	0.693[†]
Ischemic heart disease (–/+)	332/125	95/32	0.628[†]

[#]Student's t-test, [†]chi-square test, [‡]Mann-Whitney U test.

< 0.001), 0.750 for QN (P < 0.001), 0.720 for DBD (P < 0.001), and 0.636 for CGD (P = 0.008). The QP algorithm produced the lowest standard error level (0.0315), while the CGD algorithm produced the highest (0.0513).

The pairwise comparisons of the AUCs of the ROC curves obtained from the MLPNNs trained with the QP, LM, BP, QN, DBD, and CGD algorithms by the testing process are shown in Table 6. The AUC of the QP algorithm was significantly higher than the AUCs of the QN (P = 0.021), DBD (P = 0.005), and CGD (P < 0.001) algorithms. The AUC of the LM algorithm was significantly higher than the AUCs of the DBD (P = 0.014) and CGD (P < 0.001) algorithms. The AUC of the BP algorithm was significantly higher than that of the CGD (P = 0.005) algorithm. However, there were no statistically significant differences in the other AUC pairings (P > 0.05) for any of the comparisons.

Table 3. Confusion matrix for the testing process.

		Final status of stroke	
		Living	Deceased
QP prediction	Living	113	8
	Deceased	26	29
LM prediction	Living	106	13
	Deceased	26	31
BP prediction	Living	100	14
	Deceased	29	33
QN prediction	Living	100	18
	Deceased	28	30
DBD prediction	Living	89	17
	Deceased	39	31
CGD prediction	Living	80	19
	Deceased	50	27

Discussion

We examined the performance of the MLP trained with the QP, LM, BP, QN, DBD, and CGD algorithms to predict mortality in stroke. We found that the QP algorithm achieved the highest accuracy rates, while the CGD algorithm achieved the lowest rates.

When we investigated our predictive results, we observed that mortality in stroke was predicted with accuracy rates varying from 60.8% to 80.7% by the MLP trained with the 6 different algorithms. Among the 6 algorithms, the QP algorithm achieved the highest accuracy rate (80.7%), while the CGD

Table 4. Predictive results of the MLP algorithms.

	Sensitivity (%)	Specificity (%)	Accuracy (%)
QP	78.4	81.3	80.7
LM	70.5	80.3	77.8
BP	70.2	77.5	75.6
QN	62.5	78.1	73.9
DBD	64.6	69.5	68.2
CGD	58.7	61.5	60.8

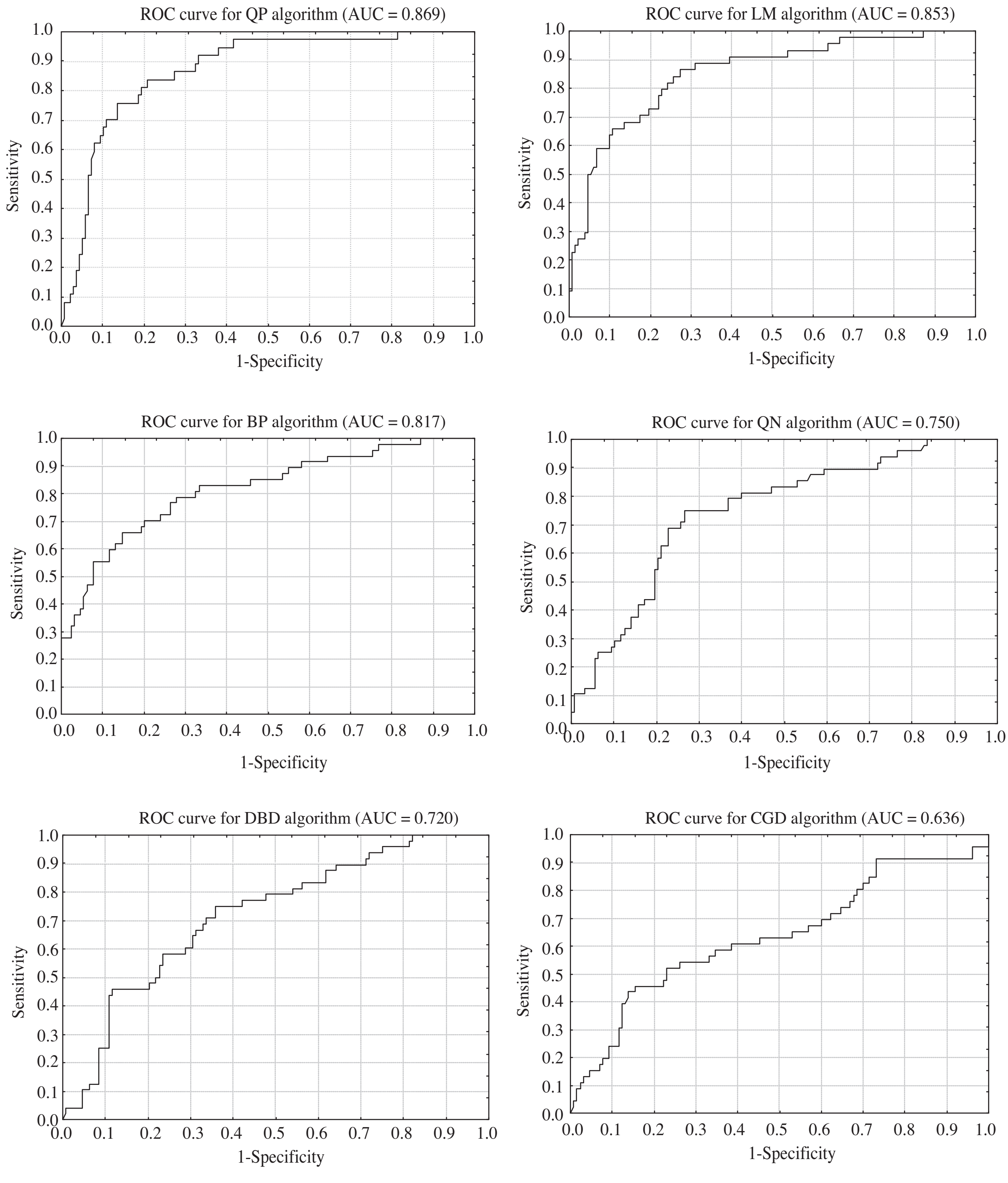

Figure 2. ROC curves of the MLP trained with the QP, LM, BP, QN, DBD, and CGD algorithms.

algorithm achieved the lowest accuracy rate (60.8%). When we investigated the specificity and sensitivity rates, similarly, the QP algorithm achieved the highest rates (specificity = 81.3%, sensitivity = 78.4%), while the CGD algorithm achieved the lowest rates (specificity = 61.5%, sensitivity = 58.7%).

Table 5. Results of the ROC curves of the MLP algorithms.

	Area under the ROC curve (AUC)	Standard error	95% Confidence interval for AUC	z-statistic	P
QP	0.869	0.0315	0.811-0.916	11.735	<0.001
LM	0.853	0.0340	0.792-0.902	10.386	<0.001
BP	0.817	0.0380	0.752-0.871	8.331	<0.001
QN	0.750	0.0406	0.679-0.812	6.136	<0.001
DBD	0.720	0.0420	0.547-0.785	5.229	<0.001
CGD	0.636	0.0513	0.560-0.707	2.655	0.008

In a study by İçer et al. (10), the MLP was trained with 3 algorithms (resilient propagation, LM, and scaled conjugate gradient algorithms) in the diagnosis of cirrhosis disease. They found that LM was the most efficient algorithm. In a study by Güler and Übeyli (11), the MLP was trained with 4 algorithms [QP, BP, DBD, and extended DBD (EDBD)] in the diagnosis of partial epilepsy; they found that QP was the most efficient algorithm. Similarly, in a study by Süt and Şenocak (3), the MLP was trained with 4 algorithms (QP, BP, DBD, and EDBD) and the statistical methods were compared in the diagnosis of coronary artery disease. They found that QP was the most efficient among the 4 algorithms. Consistent with these studies, our classification results and statistical parameters showed that the QP algorithm was the most efficient among the 6 algorithms for the mortality prediction of stroke.

When we investigated the AUCs of the ROC curves, we observed that the QP algorithm achieved the highest AUC (0.869), while the CGD algorithm achieved the lowest AUC (0.636). The AUC of the QP algorithm was statistically significantly higher than the AUCs of the QN, DBD, and CGD algorithms. The AUC of the LM algorithm was significantly higher than the AUCs of the DBD and CGD algorithms. The AUC of the BP algorithm was significantly higher than that of the CGD algorithm. These findings showed that the QP was the most efficient and powerful algorithm in mortality prediction for patients with stroke.

In conclusion, the MLP trained with the QP algorithm achieved the highest specificity (81.3%), sensitivity (78.4%), and accuracy (80.7%) rates, and so it can be a helpful tool in the prediction of mortality in stroke.

Table 6. Pairwise comparisons of the AUCs of the ROC curves obtained from the MLP algorithms.

	QP	LM	BP	QN	DBD
LM	z = 0.345 P = 0.730	–	–	–	–
BP	z = 1.054 P = 0.292	z = 0.706 P = 0.480	–	–	–
QN	z = 2.316 **P = 0.021**	z = 1.945 P = 0.052	z = 1.205 P = 0.228	–	–
DBD	z = 2.838 **P = 0.005**	z = 2.461 **P = 0.014**	z = 1.713 P = 0.087	z = 0.514 P = 0.608	–
CGD	z = 3.870 **P < 0.001**	z = 3.526 **P < 0.001**	z = 2.835 **P = 0.005**	z = 1.743 P = 0.081	z = 1.267 P = 0.205

References

1. American Stroke Association. About stroke. Dallas (TX): ASA. Available from URL: http://www.strokeassociation.org/STROKEORG/AboutStroke/About-Stroke_UCM_308529_SubHomePage.jsp. Accessed 17 January 2011.
2. Yan H, Jiang Y, Zheng J, Peng C, Li Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Syst Appl 2006; 30: 272-81.
3. Süt N, Şenocak M. Assessment of the performances of multilayer perceptron neural networks in comparison with recurrent neural networks and two statistical methods for diagnosing coronary artery disease. Expert Syst 2007; 24: 131-42.
4. Gurgen F, Guler N, Varol F. Antenatal fetal risk assessment by blood-flow velocity waveforms. IEEE Eng Med Biol Mag 2000; 19: 88-93.
5. Calcagno G, Staiano A, Fortunato G, Brescia-Morra V, Salvatore E, Liguori R et al. A multilayer perceptron neural network-based approach for the identification of responsiveness to interferon therapy in multiple sclerosis patients. Inform Sciences 2010; 180: 4153-63.
6. Sun R, Wang Y. Predicting termination of atrial fibrillation based on the structure and quantification of the recurrence plot. Med Eng Phys 2008; 30: 1105-11.
7. Trtica-Majnaric L, Zekic-Susac M, Sarlija N, Vitale B. Prediction of influenza vaccination outcome by neural networks and logistic regression. J Biomed Inform 2010; 43: 774-81.
8. Türe M, Kurt İ, Kurum AT, Özdamar K. Comparing classification techniques for predicting essential hypertension. Expert Syst Appl 2005; 29: 583-8.
9. Marcos JV, Hornero R, Álvarez D, Campo F, Zamarrón C, López M. Utility of multilayer perceptron neural network classifiers in the diagnosis of obstructive sleep apnoea syndrome from nocturnal oximetry. Comput Meth Prog Bio 2008; 92: 79-89.
10. İçer S, Kara S, Güven A. Comparison of multilayer perceptron training algorithms for portal venous doppler signals in the cirrhosis disease. Expert Syst Appl 2006; 31: 406-13.
11. Güler I, Übeyli ED. An expert system for detection of electrocardiographic changes in patients with partial epilepsy using wavelet-based neural networks. Expert Syst 2005; 22: 62-71.
12. Santhosh Baboo S, Kadar Shereef I. An efficient weather forecasting system using artificial neural network. IJESD 2010; 1: 321-6.
13. Rumelhart DE, McClelland J, PDP Research Group. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge (MA): MIT Press; 1986.
14. Bishop CM. Neural networks for pattern recognition. Oxford: Oxford University Press; 1995.
15. StatSoft. StatSoft electronic textbook. Available from URL: http://www.statsoft.com/textbook/stathome.html. Accessed 15 January 2011.
16. Güler NF, Übeyli ED, Güler İ. Recurrent neural networks employing Lyapunov exponents for EGG signals classification. Expert Syst Appl 2005; 29: 506-14.
17. Rumelhart D, Hinton G, Williams R. Learning internal representations by back-propagating errors. Cambridge (MA): MIT Press; 1986.
18. Duda RO, Hart PE, Stork DG. Pattern classification. New York: Wiley, 2001.