Predicting 10-day Mortality in Patients with Strokes Using Neural Networks and Multivariate Statistical Methods

Güner Çelik, MD,* Ömer K. Baykan, PhD,† Yakup Kara, PhD,‡ and Hülya Tireli, MD§

Background: The aim of the present study was to evaluate the performance of 2 different multivariate statistical methods and artificial neural networks (ANNs) in predicting the mortality of hemorrhagic and ischemic patients within the first 10 days after stroke. *Methods:* The multilayer perceptron (MLP) ANN model and multivariate statistical methods (multivariate discriminant analysis [MDA] and logistic regression analysis [LRA]) have been used to predict acute stroke mortality. The data of total 570 patients (230 hemorrhagic and 340 ischemic stroke), who were admitted to the hospital within the first 24 hours after stroke onset, have been used to develop prediction models. The factors affecting the prognosis were used as inputs for prediction models. Survival or death status of the patients was taken as output of the models. *Results:* For the MLP method, the accuracies were 99.9% in a training data set and 80.9% in a testing data set for the hemorrhagic group, whereas 97.8% and 75.9% for the ischemic group, respectively. For the MDA method, the training and testing performances were 89.8%, 87.8% and 80.6%, 79.7% for hemorrhagic and ischemic groups, respectively. For the LRA method, the training and testing performances for the hemorrhagic group were 89.7% and 86.1%, and for the ischemic group were 81.7% and 80.9%, respectively. *Conclusions:* Training and test performances yielded different results for ischemic and hemorrhagic groups. MLP method was most successful for the training phase, whereas LRA and MDA methods were successful for the test phase. In the hemorrhagic group, higher prediction performances were achieved for both training and testing phases. **Key Words:** Predicting outcome—ischemic stroke—hemorrhagic stroke—models—statistical.

Introduction

Human life is extending and the population of elderly people is increasing in developed countries. It is expected that longevity of the human life will increase in the future.[1] The incidence of stroke increases with age, and the prognosis is more severe in the elderly people than the younger.[1,2] More recently, the addition of more promising and new treatment options has not yet pulled down high mortality and morbidity rates. It is expected that the incidence of stroke and the rate of dependent people will increase with the increase in longevity of human life. To decrease the rates of mortality and morbidity, it is of utmost importance to determine the appropriate treatment, care, and rehabilitation programs for each patient in early stage of stroke. For these purposes, a number of models are present in the literature for prediction of mortality and independent survival rates after stoke. It has been aimed to use these models for randomization of patients in clinical studies and provide case–controls in nonrandomized studies, compare patients in different groups, and analyze the results of randomized trials

From the *Department of Neurology, Faculty of Medicine, Baskent University, Konya; †Department of Computer Engineering, Selcuk University, Konya; ‡Department of Industrial Engineering, Selcuk University, Konya; and §Department of Neurology, Haydarpasa Numune Training and Research Hospital, Istanbul, Turkey.

Received November 12, 2013; revision received December 5, 2013; accepted December 16, 2013.

Address correspondence to Güner Çelik, MD, Department of Neurology, Faculty of Medicine, Baskent University, Konya, Turkey, Hocacihan Mh. Saray Cd. No. 1 Selçuklu, Konya, Turkey. E-mail: gunerkoyuncu@mynet.com.

1052-3057/$ - see front matter

http://dx.doi.org/10.1016/j.jstrokecerebrovasdis.2013.12.018

and meta-analysis. For this purpose, various models and methods have been developed up-to-date.[3-15] Most of these models are not widely accepted and routinely used in practice. In addition, the rates of the accuracy and reliability have generally not reached the satisfying level for these models. Clinical applicability of these models depends on its easy feasibility/applicability, reliability, and its success in predicting the clinical outcome.

The aim of the present study was to determine the performance of 2 different multivariate statistical methods (logistic regression analysis [LRA] and multivariate discriminant analysis [MDA]) and artificial neural networks (ANNs) particularly in early (first 10 days) prognosis of ischemic and hemorrhagic strokes, and to compare them with each other. By this way, it was aimed to provide contribution to the practice with the results of a rigorous experiment conducted based on a wide data set.

Materials and Methods

Research Data

The data of 1725 patients with stroke, admitted to the emergency service in the first 24 hours of stroke and hospitalized in the Neurology Department in Haydarpasa Numune Training and Research Hospital, have been used in the study. The patients diagnosed of transient ischemic attack, sinus thrombosis, subarachnoid hemorrhage, and brain tumor were excluded from the data set after computed tomography or magnetic resonance imaging. The remaining 968 patients were included in the study. Of these patients 293 (30.3%) died and 675 (69.7%) survived. On the basis of the results of our preliminary experiments, it was decided to include equal number of dead and survived patients in a group to improve the training performance of proposed models. Finally, 2 groups, hemorrhagic with 230 patients and ischemic with 340 patients, were obtained and used in the experiments. To increase the reliability of the results, a 5-fold cross-validation technique was used. The number of patients in the entire data set and modeling data set are given in Table 1.

The factors affecting the prognosis of hemorrhagic and ischemic strokes have been determined by analyzing similar studies in the literature and were used as inputs in prediction models. Although age, gender, hypertension, diabetes mellitus, smoking, mean blood pressure on admission, and Scandinavian Stroke Scale score[16] were used as common inputs for both groups, inputs for ischemic stroke included blood glucose levels, lacunar infarct, nonlacunar infarct, ischemic stroke with undetermined etiology,[12] congestive heart failure, coronary heart disease, myocardial infarction, atrial fibrillation, history of cerebrovascular disease, and transient ischemic attack; inputs for the hemorrhagic group included pulse pressure, localization of hemorrhage (putaminal, thalamic, cerebellar, pontine, lobar, and others [caudate, thalamocapsular, intraventricular]), volume of hemorrhage, ventricular drainage, and the presence or absence of a midline shift. In the determination of the early-stage outcomes of patients with acute stroke, the findings of average 10-day hospitalization were used. Return and normalization of the cerebral function, however, is variable according to the age, risk factor, and site and size of the lesion 10 days is the average duration of hospitalization in our clinic. Output for both groups was survival or death within the first 10 days after the stroke.

Table 1. *Number of cases in data sets*

Stroke type	Data set	Dead	Survived	All subjects
Hemorrhagic	Entire	119	138	257
	Modeling	115	115	230
Ischemic	Entire	174	537	711
	Modeling	170	170	340
Total	Entire	293	675	968
	Modeling	285	285	570

Tables 2 and 3 show summary statistics for hemorrhagic and ischemic strokes, respectively.

ANN Model

A multilayer perceptron (MLP) neural network is a feed-forward neural network model proposed by Rumelhart et al[17] and consists of 1 input layer, 1 or more hidden layers, and 1 output layer.[18] The number of neurons in input and output layers is determined by the number of input and output vectors used in the data set. The number of neurons in the hidden layer can be determined experimentally or based on experience. A neuron on a layer is connected to all neurons of the adjacent layer with its weights. Usually, the initial values of weights are randomly selected. Outputs of hidden and output layer neurons are produced depending on selected transfer function and weighted neuron inputs. The MLP model uses a supervised neural network and is trained by a gradient descent method to minimize an error function.[19] Generalized delta learning rule based on least square can be used to train an MLP network, so these weights are adjusted for a given set of input–output pairs.[20] A typical MLP architecture adopted in this study is shown in Figure 1.

Multivariate Statistical Models

In this group of prediction techniques, 2 multivariate statistical methods MDA and LRA were used to predict the stroke mortality.

MDA is concerned with the classification of distinct sets of observations, and it tries to find the combination of variables that predicts the group to which an observation belongs. The combination of predictor variables is