

Artificial neural network, genetic algorithm, and logistic regression applications for predicting renal colic in emergency settings

Caner Eken · Ugur Bilge · Mutlu Kartal · Oktay Eray

Received: 3 March 2009 / Accepted: 13 April 2009 / Published online: 3 June 2009
© Springer-Verlag London Ltd 2009

Abstract

Background Logistic regression is the most common statistical model for processing multivariate data in the medical literature. Artificial intelligence models like an artificial neural network (ANN) and genetic algorithm (GA) may also be useful to interpret medical data.

Aims The purpose of this study was to perform artificial intelligence models on a medical data sheet and compare to logistic regression.

Methods ANN, GA, and logistic regression analysis were carried out on a data sheet of a previously published article regarding patients presenting to an emergency department with flank pain suspicious for renal colic.

Results The study population was composed of 227 patients: 176 patients had a diagnosis of urinary stone, while 51 ultimately had no calculus. The GA found two decision rules in predicting urinary stones. Rule 1 consisted of being male, pain not spreading to back, and no fever. In rule 2, pelviciceal dilatation on bedside ultrasonography replaced no fever. ANN, GA rule 1, GA

rule 2, and logistic regression had a sensitivity of 94.9, 67.6, 56.8, and 95.5%, a specificity of 78.4, 76.47, 86.3, and 47.1%, a positive likelihood ratio of 4.4, 2.9, 4.1, and 1.8, and a negative likelihood ratio of 0.06, 0.42, 0.5, and 0.09, respectively. The area under the curve was found to be 0.867, 0.720, 0.715, and 0.713 for all applications, respectively.

Conclusion Data mining techniques such as ANN and GA can be used for predicting renal colic in emergency settings and to constitute clinical decision rules. They may be an alternative to conventional multivariate analysis applications used in biostatistics.

Keywords Artificial neural network · Genetic algorithm · Logistic regression · Renal colic · Emergency department

Introduction

There are many articles in the medical literature measuring diagnostic tools for patients with acute flank pain. The main purpose of these studies is to find the safest and most cost-effective way to detect urinary stones in the emergency department (ED). Logistic regression is mostly used in these studies in processing multivariate data. Alternative methods in data mining, also known as artificial intelligence, should also be used to evaluate the medical data sheet.

An artificial neural network (ANN) is an information processing tool that is inspired by the structure and function of the human brain. The human central nervous system is composed of a series of interconnecting neurons separated by synapses, and scientists have demonstrated information transfer via a series of action potentials [1]. The brain learns by adjusting the number and strength of these connections.

The views expressed in this paper are those of the author(s) and not those of the editors, editorial board or publisher.

C. Eken (✉) · M. Kartal · O. Eray
Department of Emergency Medicine, Akdeniz University Medical Faculty,
Dumlupinar Bulvari, Kampus,
07059 Antalya, Turkey
e-mail: cenkereken@akdeniz.edu.tr

M. Kartal
e-mail: enmkartal@yahoo.com

U. Bilge
Department of Biostatistics, Akdeniz University Medical Faculty,
07059 Antalya, Turkey
e-mail: ubilge@akdeniz.edu.tr

McCulloch and Pitts [2] first described ANNs as a method of information processing using a network of binary decision elements or “neurons.” Later efforts were made in order to explain complex processes of the central nervous system [3]. A neural network is composed of a series of interconnecting parallel nonlinear processing elements (nodes) resembling biological neurons.

A genetic algorithm (GA) [4, 5] is a good method for finding quick solutions to hard combinatorial optimization problems where there are too many combinations to search and were first put forward by John Holland as a search and optimization technique. A GA typically generates a random population of possible rule hypotheses and tests them with a given fitness function. In a standard GA, solution hypotheses are represented as code strings; this is also called a genotype where genetic operators such as crossover and mutation are applied. The resulting phenotypes are then assessed by applying the fitness function to each one of them. At each iteration the GA applies the principles of natural evolution to the population of hypotheses, that is, selecting the fittest hypotheses using the fitness function as well as applying mutation and crossover to the population.

In our first article [6], we tested a current algorithm and investigated the clinical characteristics of renal colic patients in the ED. The objective of this study was to compare predictive values of multiple statistical models for diagnosing renal colic patients in the same data set.

Material and methods

Data collection

The data sheet of our previously published article [6] on renal colic was used to perform ANN, GA, and logistic regression analyses. We additionally included the body temperature of patients into the analyses although it had not been included in the previous data sheet. The data set consists of the following parameters:

Parameter	Type
Gender	Coded (2 codes, i.e., male, female)
Number of attacks	Coded (2 codes, i.e., first, more than one) Coded as discrete variable for GA
Triggered by exercise	Coded (2 codes; 1 for yes and 0)
Previous presentation	Coded (2 codes; 1 for yes and 0)
History of urolithiasis	Coded (2 codes; 1 for yes and 0)
Family history of urolithiasis	Coded (2 codes; 1 for yes and 0)
Comorbid disease	Coded (2 codes; 1 for yes and 0)
Nausea	Coded (2 codes; 1 for yes and 0)

Parameter	Type
Vomiting	Coded (2 codes; 1 for yes and 0)
Sweating	Coded (2 codes; 1 for yes and 0)
Fever	Coded (2 codes; 1 for yes and 0)
Dysuria	Coded (2 codes; 1 for yes and 0)
Hematuria	Coded (2 codes; 1 for yes and 0)
Unable to void	Coded (2 codes; 1 for yes and 0)
Tenderness in costovertebral region	Coded (2 codes; 1 for yes and 0)
Tenderness on ureteral tract	Coded (2 codes; 1 for yes and 0)
Suprapubic tenderness	Coded (2 codes; 1 for yes and 0)
Abdominal rigidity	Coded (2 codes; 1 for yes and 0)
Rebound	Coded (2 codes; 1 for yes and 0)
Positional discomfort	Coded (2 codes; 1 for yes and 0)
Positive urine analysis (≥ 10 red blood cells)	Coded (2 codes; 1 for yes and 0)
Pelvic/iceal dilatation on bedside ultrasonography	Coded (2 codes; 1 for yes and 0)
Colic pain	Coded (2 codes; 1 for yes and 0)
Radiation to the groin	Coded (2 codes; 1 for yes and 0) 0=no radiation 1=radiation to the belly 2=radiation to the groin 3=radiation to the back for GA codes

Statistical analyses

Artificial neural network

A feed-forward ANN with back-propagation was performed by JMP (release 6.0, a business unit of SAS). In large data sets, a data set is divided as a training set and a test set to avoid overfitting which is a problem when ANN learns the training set too accurately yet it cannot generalize when presented with a new test set. However, for the small data sets, the k-fold cross-validation model is suggested to avoid overfitting. The k-fold cross-validation method separates the data into k sets and assigns one of the k sets as the test set and the remaining as the training set. We performed 10-fold cross-validation with 2 hidden units, 500 iterations, and 20 tours. The overfit penalty was assigned as 0.001 and convergence criterion was chosen as 0.00001.

Genetic algorithm

The GA program SimMine is part of a data mining software package developed by SimWorld Limited, UK (www.simworld.co.uk). The system allows the selection of

flat or hierarchical chromosome types and allows preprocessing the data by scaling, taking first differences and dividing continuous data into categories. As GA normally operate on categorical data, when the input parameters are a mixture of numerical and coded values, the system converts numerical values into a number of categories. The system allows users to select the total number of chromosomes (the population size) and the number of top scoring individuals that survive to the next generation.

Fitness score

When the GA puts forward a rule hypothesis, the system counts the number of samples that are covered by this hypothesis. By this we mean *positive_count*; for example, if we are trying to find a rule that explains all the patients that have urinary stones, then *positive_count* would be the total number of patients with urinary stones. Then the system tests each hypothesis by counting the number of patients that are covered by the hypothesis, and this is the *true_positive*. Similar logic is applied to find *negative_count*, which is the total number of samples not covered by the hypothesis (i.e., total samples minus *positive_count*), and the *false_positive*, which is the total number of samples the hypothesis would falsely find positive, which is undesirable. So the fitness function would be:

$$\text{Fitness score} = (\text{true_positive}/\text{positive_count}) - (\text{false_positive}/\text{negative_count})$$

As can be seen above, the GA could achieve a maximum score of 1 when it identifies all positives correctly, and would have a *false_positive* of 0.

We used the software previously on data mining tasks at Akdeniz University Hospital with good results [3].

Logistic regression

The Statistical Package for the Social Sciences (SPSS 13.0 for Windows) was used for binary logistic regression analysis. Like in GA and ANN, 24 independent variables were assigned to predict the renal colic as the dependent variable. A stepwise forward logistic regression analysis was performed. We tested the fitness of the logistic regression model with the Hosmer-Lemeshow goodness-of-fit statistic.

To compare the overall performances of all applications, sensitivity, specificity, positive predictive value, and negative predictive value of all applications were determined. And also a receiver-operating characteristic curve (ROC) analysis was performed; $p \geq 0.05$ was accepted as significant.

Results

The study population was composed of 227 patients; 147 (64.8%) patients were men and their mean age was 38.4 ± 14 . Diagnostic studies and follow-up revealed 176 patients with urinary stone, while 51 had no calculus. Demographics of 24 independent variables are shown in Table 1.

Artificial neural network

ANN revealed 174 of 176 patients with renal colic and 40 of 51 patients without renal colic. It assigned 11 (21.6%) patients falsely as having renal colic and 9 (5.1%) renal colic patients as having no calculus. So it had a sensitivity of 94.9% and a specificity of 78.4%. The positive predictive value of ANN was 93.8% and negative predictive value was 81.6%. The positive likelihood ratio (PLR) of ANN was 4.4 and negative likelihood ratio (NLR) was 0.06. The area under the curve (AUC) for the ANN model was 0.928 ($p=0.000$). Figure 1 shows an ANN diagram with three layers: input layer, hidden layer with two nodes, and output layer.

Genetic algorithm results

We used the GA population size of 128, and the top scoring 32 individuals were selected as fit to survive in the next generation of solutions; the search was stopped after no improvement was seen for the last 25 iterations of each run. The default value for the crossover rate was 60%, and the mutation rate was 20% per chromosome.

When we apply the GA, the system quickly converges into simple rules. Modifying the crossover rate or the mutation rate does not improve the results.

The best rule to explain **urinary stone=1** is presented below:

Rule 1:
Kidney stone (1): gender (1) AND spread (0 OR 1 OR 2) AND no fever (0)

Total samples=227

Positive_count=176 (total of 176 cases of urinary stone is 1)

True_positive=119 (the rule above explains 119 of the 176)

Negative_count=51 (total of 51 cases where urinary stone is 0)

False_positive=12 (the rule is wrong in 12 of 51 cases)

Fitness score=0.4408

Sensitivity=67.61%

Specificity=76.47%

Table 1 Demographics and univariate data analysis of patients with renal colic

Variable	Patients with urinary stone	%	<i>p</i>
Age	38.4±14		
Gender: male/female	128/48	87/60	0.000
More than one attack	98	81	0.182
Colic pain	24	75	0.711
Radiation to the groin	120	81.6	0.045
Positional discomfort	23	79.3	0.806
Previous presentation	69	89.6	0.002
History of urolithiasis	104	86	0.001
Family history of urolithiasis	85	81.7	0.163
Comorbid disease	33	73.3	0.451
Nausea	126	83.4	0.003
Vomiting	73	85.9	0.020
Sweating	74	84	0.060
Fever	9	60	0.092
Dysuria	80	80	0.429
Hematuria	57	86.4	0.041
Unable to void	26	84	0.363
Tenderness in costovertebral region	150	80.2	0.036
Tenderness on ureteral tract	101	79	0.573
Suprapubic tenderness	40	70.2	0.124
Abdominal rigidity	9	90	0.334
Rebound	3	100	0.348
Positional discomfort	36	80	0.658
Positive urine analysis	121	83	0.010
Pelvic/iceal dilatation on bedside US	142	81.6	0.008

US ultrasonography

Positive predictive value=90.84%
 Negative predictive value=40.6%
 Positive likelihood ratio: 2.9
 Negative likelihood ratio: 0.42
 Area under the curve=72.04%

This rule can be explained in a more articulate form as “the patient has urinary stone if: the gender is male, and the pain radiates to groin and belly or no pain radiation (codes: 0, 1, or 2), and has no fever”. As there are 4 types of pain (they are 0, 1, 2, and 3 as can be seen in the “Data collection” section), we can reword the rule as “the patient has urinary stone if: the gender is male, and the pain does not radiate to the back, and the patient has no fever.”

A second rule that was also found, which has slightly less fitness value but worth reporting, is shown below:

Rule 2:

Urinary stone (1): gender (1) AND pain radiation (0 OR 1 OR 2) AND pelvic/iceal dilatation on bedside ultrasonography (1)

Total samples=227

Positive_count=176 (total of 176 cases of urinary stone is 1)

True_positive=100 (the rule above explains 100 of the 176)
 Negative_count=51 (total of 51 cases where urinary stone is 0)
 False_positive=7 (the rule is wrong in 7 of 51 cases)
 Fitness score=0.4309
 Sensitivity=56.82%
 Specificity=86.27%

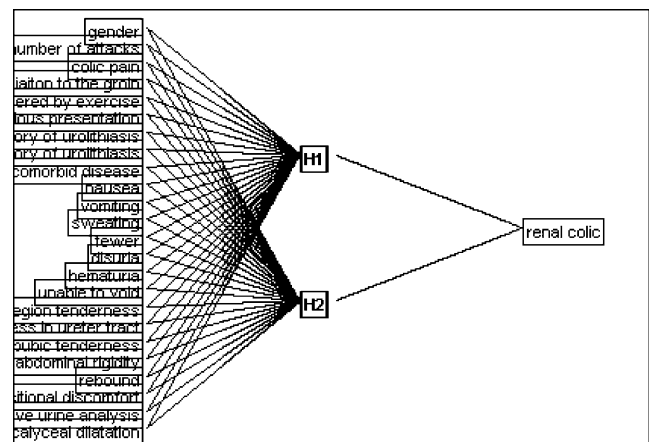


Fig. 1 ANN diagram exhibiting the input layer with 24 nodes, hidden layer with 2 nodes, and output layer

Positive predictive value=93.46%
 Negative predictive value=36.7%
 Positive likelihood ratio: 4.1
 Negative likelihood ratio: 0.5
 Area under the curve=71.55%

This rule can be also explained in a more articulate form as “the patient has urinary stone if: the gender is male, and the pain does not radiate to the back, and has pelvicaliceal dilation on bedside ultrasonography (US).”

Logistic regression

With the forward stepwise conditional method, the logistic regression model with 24 independent variables mentioned above revealed 168 of 176 patients with renal colic and 24 of 51 patients without renal colic. So it had a sensitivity of 95.5% and specificity of 47.1%. The positive predictive value was 86.2% and negative predictive value was 75.3%. The positive predictive value of logistic regression was 1.8 and negative predictive value was 0.09. The AUC for the logistic regression model was 0.841 ($p=0.000$).

Gender (to be male) [odds ratio (OR): 5.567, 95% confidence interval (CI): 2.34–13.24, $p=0.000$], history of urolithiasis (OR: 2.72, 95% CI: 1–7.3, $p=0.047$), and nausea (OR: 3, 95% CI: 1.12–8.5, $p=0.029$) were found to be independent factors in predicting urinary stones in logistic regression analysis.

Table 2 shows the overall performances of each model. Figure 2 depicts the comparison of ANN and logistic regression applications after the ROC analysis according to the prediction rates. The AUC for GA rules may be realized as smaller than ANN and logistic regression. However, these findings represent the binary outcomes of GA rules. Also, a ROC analysis was performed to see the overall performances of each model after the outcomes were dichotomized to have renal colic or not. Figure 3 shows a comparison of the three models with ROC analysis.

Discussion

We are in the process of applying the technology to other data mining tasks in the hospital databases. We are aiming

for an automated approach to explore and reveal hidden, unknown relationships in data. We wanted to enrich the ANN and GA data mining techniques with a logistic regression processing module and also to compare the fitness functions using statistical techniques.

In medical research, efforts are expended to detect a disease most accurately or predict the prognosis of a specific group of patients. Logistic regression analysis is the most frequently used statistical analysis for evaluating multiple independent variables through a dependent outcome in medical research. Although ANN and GA have been used commonly in other industries, they do not have widespread use in medicine so far. Here we show that both ANN and GA can be used for constituting prediction models and decision rules as well as logistic regression analysis.

ANN emerged as the best technique in predicting renal colic according to the results of this study. For ANN both the sensitivity (94.9%) and PLR (4.4) were higher than for the other models. The AUC was also found to be higher than with logistic regression and GA. However, overfitting is a major problem in ANN applications. In order to prevent overfitting, the data set must be separated randomly as the training set and test set. After the learning process of ANN by using the training set, the model should be tested in unseen data, a test set and sometimes a validation set. However, this is not always possible for small data sets. The k-fold cross-validation that we performed in this study separates the data sets into k groups and uses the groups as the data test set and the remaining as the training set. Although k-fold cross-validation was used, to test ANN in an unseen data test would be better. Even so the results are promising for ANN. ANN previously has been used for predicting genetic factors in stone disease [7], stone composition [8], spontaneous passage of ureteral calculi [9], and outcome after extracorporeal shock wave lithotripsy (ESWL) [10]. So far there is only one study detecting the performance of ANN in predicting the presence of renal calculi [11]. Tanthanuch and Tanthanuch studied 168 patients with renal calculi, 100 patients for the training set and 68 patients for the test set. After assuming a probability of 65% or more as having renal calculi, they found that ANN was 100% accurate in predicting renal calculi for the testing data. However, the study population they investigated

Table 2 Overall performances of each application

Application	Sensitivity	Specificity	PLR	NLR	AUC
ANN	94.9	78.4	4.4	0.06	0.867
GA rule 1	67.6	76.47	2.9	0.42	0.720
GA rule 2	56.8	86.3	4.1	0.5	0.715
Logistic regression	95.5	47.1	1.8	0.09	0.713

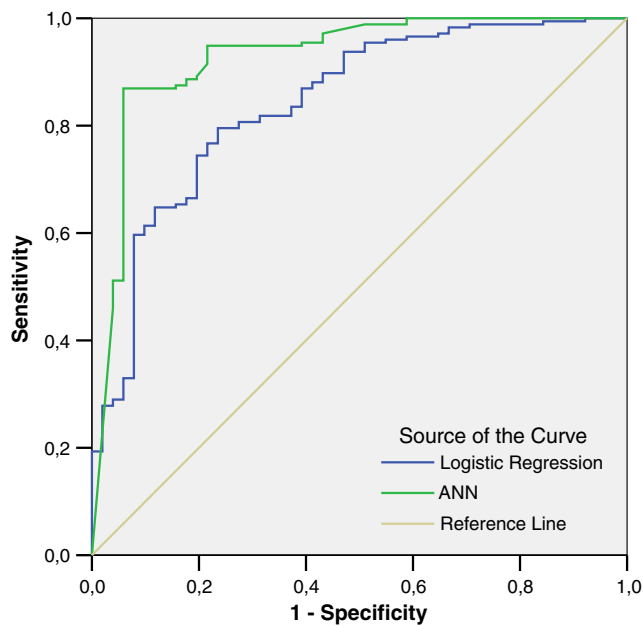


Fig. 2 ROC analysis of logistic regression and ANN in predicting the probability of urinary stones

consisted of only patients with renal calculi. Thus, ANN was trained and tested only in patients with renal calculi. The performance of ANN was not tested in patients without a renal calculus. This is a limitation of their study.

GAs are relatively new search techniques which have been used in optimization, scheduling, and planning tasks, and there is a huge amount of background literature on them. As in other industries, they are increasingly being used in medical data mining tasks. Our application of GA found useful relationships for predicting renal colic. Especially the PLR of the second rule was worth mentioning. The advantage of GA is to discover simple predictive rules. In the first rule of GA, to be male, having pain not radiating to the back, and having no fever composed the independent variables. In the second rule, pelviciceal dilatation on bedside US replaced having no fever. The PLR (4.1) of the second rule seems better than the first rule. And the specificity of the second rule (86.3%) was better than the first rule (76.5%). However, the sensitivity of the first rule (67.6%) was better than the second rule (56.8%).

Logistic regression emerges as the best model in this study in excluding renal colic. The NLR of logistic regression was 0.09, and also the NLR of ANN was 0.06, which is also reasonable to mention. But the overfitting problem has to be taken into consideration for ANN as mentioned above. Furthermore, gender (to be male), nausea, and history of urolithiasis were found to be the independent variables in predicting urinary stones.

One can also constitute simple prediction rules both by ANN and logistic regression applications. After analyzing

the significant variables in the univariate analysis or by using the likelihood ratios, it is possible to compose a prediction model by assigning significant variables into the ANN and logistic regression analysis. Furthermore, the independent variables that are statistically significant in a multivariate regression analysis can also be used to compose a prediction model. We preferred to use all of the variables (24 variables) for all the models in order to prevent bias. Different models for renal colic need to be tested in further studies.

Conclusion and future work

Data mining techniques, particularly ANN rather than GA, have been used in order to solve problems in the ED like diagnosing pathological conditions such as cardiac ischemia, craniocervical junction injury in trauma patients, and pneumonia, predicting the prognosis of patients in the ED, and also possible ED visits of patients because of respiratory symptoms [12–16]. This study also showed that data mining techniques such as ANN and GA could be used for predicting or excluding renal colic in emergency settings and to constitute clinical decision rules. Data mining techniques should be useful tools to solve the sophisticated issues in the ED in the future. They may also be an alternative to conventional multivariate analysis applications used in biostatistics. Further studies are needed to validate these data mining techniques in predicting either renal colic or complicated points in the ED.

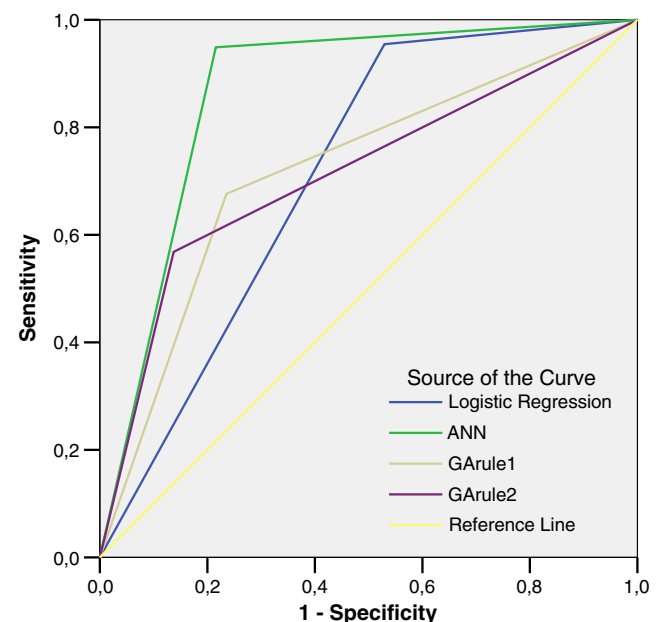


Fig. 3 ROC analyses comparing all the applications in predicting urinary stones as binary outcome

Acknowledgement This study was supported by Akdeniz University Investigation Foundation. We are also grateful to the residents and physicians in the emergency department.

References

1. Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol* 117:500–544
2. McCulloch WS, Pitts WA (1943) A logical calculus of ideas immanent in nervous activity. *Bull Math Biophys* 5:115–133
3. Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 160:106–154
4. Holland J (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor
5. Goldberg DE (1989) *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley, Reading
6. Kartal M, Eray O, Erdogru T, Yilmaz S (2006) Prospective validation of a current algorithm including bedside US performed by emergency physicians for patients with acute flank pain suspected for renal colic. *Emerg Med J* 23:341–344
7. Chiang D, Chiang HC, Chen WC, Tsai FJ (2003) Prediction of stone disease by discriminant analysis and artificial neural networks in genetic polymorphisms: a new method. *BJU Int* 91:661–666
8. Kuzmanovski I, Zografski Z, Trpkovska M et al (2001) Simultaneous determination of composition of human urinary calculi by use of artificial neural networks. *Fresenius J Anal Chem* 370:919–923
9. Cummings JM, Boullier JA, Izenberg SD et al (2000) Prediction of spontaneous ureteral calculus passage by an artificial neural network. *J Urol* 164:326–328
10. Poulakis V, Dahm P, Witzsch U et al (2003) Prediction of lower pole stone clearance after shock wave lithotripsy using an artificial neural network. *J Urol* 169:1250–1256
11. Tanthanuch M, Tanthanuch S (2004) Prediction of upper urinary tract calculi using an artificial neural network. *J Med Assoc Thai* 87:515–518
12. Harrison RF, Kennedy RL (2005) Artificial neural network models for prediction of acute coronary syndromes using clinical data from the time of presentation. *Ann Emerg Med* 46(5):431–439
13. Jaimes F, Farbiarz J, Alvarez D, Martínez C (2005) Comparison between logistic regression and neural networks to predict death in patients with suspected sepsis in the emergency room. *Crit Care* 9(2):R150–R156 Epub 2005 Feb 17
14. Heckerling PS, Gerber BS, Tape TG, Wigton RS (2003) Prediction of community-acquired pneumonia using artificial neural networks. *Med Decis Making* 23(2):112–121
15. Bektaş F, Eken C, Soyuncu S, Kilicaslan I, Cete Y (2008) Artificial neural network in predicting craniocervical junction injury: an alternative approach to trauma patients. *Eur J Emerg Med* 15(6):318–323
16. Bibi H, Nutman A, Shoseyov D, Shalom M, Peled R, Kivity S, Nutman J (2002) Prediction of emergency department visits for respiratory symptoms using an artificial neural network. *Chest* 122(5):1627–1632

Çenker Eken is 34 years old and has been working in the Emergency Department of Akdeniz University Medical Faculty, in Turkey, as an assistant professor. He started his residency program in 2000 and finished it in 2004. He is particularly interested in research, biostatistics, and evidence-based medicine.