Clin Chem Lab Med 2008;46(10):1458–1463 © 2008 by Walter de Gruyter • Berlin • New York. DOI 10.1515/CCLM.2008.284

An expert system for the classification of serum protein electrophoresis patterns

Sara Altinier[1,*], Lorenzo Sarti[2], Mariacristina Varagnolo[1], Martina Zaninotto[1], Marco Maggini[2] and Mario Plebani[1]

[1] Department of Laboratory Medicine, University Hospital of Padova, Padova, Italy
[2] Department of Information Engineering, University of Siena, Siena, Italy

Abstract

Background: With the improvement of capillary electrophoresis, much progress has been made in terms of sensitivity and automation, but the interpretation of the patterns, actually, depends totally on expert personnel. The aim of this work was to evaluate Neurosoft-Sebia, an expert system developed to discriminate between regular and anomalous serum protein electrophoresis patterns performed on Capillarys™2.
Methods: Neurosoft-Sebia, based on six auto-associative neural networks, was trained to create the initial knowledge base. In the tuning phase, 3000 electrophoretic patterns were performed in three different laboratories, and the discordances between human experts and Neurosoft-Sebia classifications were added to the initial knowledge base. Finally, the performances of Neurosoft-Sebia were evaluated using a benchmark dataset.
Results: The initial knowledge base was created with 2685 fractions. In the tuning phase, 241 discordances were found: 56 as regular by Neurosoft-Sebia and anomalous by human experts, and 185 as anomalous by Neurosoft-Sebia and regular by human experts. Sensitivity values were evidenced as the ability of Neurosoft-Sebia in selecting anomalous fractions, with an increase from 66.67% using the initial knowledge base to 97.40% using the enriched knowledge base.
Conclusions: This work demonstrated how the ability of Neurosoft-Sebia in selecting anomalous pattern was comparable to that of human experts, saving time and providing rapid and standardized interpretations.
Clin Chem Lab Med 2008;46:1458–63.

Keywords: capillary electrophoresis; neural network; serum protein.

*Corresponding author: Sara Altinier, Department of Laboratory Medicine, University Hospital of Padova, Via Giustiniani 2, 35128 Padova, Italy
Phone: +39-049-8218708, Fax: +39-049-8218489,
E-mail: sara.altinier@sanita.padova.it
Received March 27, 2008; accepted June 22, 2008

Introduction

Electrophoresis of serum proteins has become a test of clinical relevance, allowing the detection of monoclonal gammopathies and providing information on several diseases (1). In recent years, sensitivity and automation in capillary electrophoresis have been greatly improved, and the approach has become less time-consuming (2). However, laboratory personnel are held responsible for the analysis and interpretation of electrophoresis patterns and other related crucial aspects, such as personnel training and experience; this has, in many cases, led to a lack of standardization, an unsatisfactory degree of attention, and time wasting.

In the last two decades, several neural networks (NNs), widely used machine learning techniques, have been employed for pattern recognition (3). In some applications, in which the patterns pertain to a small number of classes and the data have a high inter-class and a low intra-class variability, their performances have been surprising (4, 5).

The original approach in electrophoresis pattern classification, dating back to 1992, was based on feed-forward NNs (6). In 1993, Manner et al. (7) proposed another NN approach, which analyzed only the γ-fraction. Finally, in 2004, another method, able to detect anomalies in the β- and γ-fractions, was presented (8). These approaches were based on multilayer perceptrons (MLPs), the more common NN model, trained using the backpropagation algorithm (9). MLPs are particularly suited for classification tasks, when the data present a high inter-class and a low intra-class variability. On the other hand, in the machine learning field, electrophoresis pattern analysis can be considered as a verification task; in fact, regular examples present low intra-class variability, whereas anomalous patterns are particularly heterogeneous and not completely known a priori. Gori and Scarselli (10) claimed that MLPs are not adequate in solving verification tasks and tend to lead to the misclassification of negative patterns; on the contrary, other NN models, such as auto-associative neural networks (AANNs) and radial basis function NNs, are particularly suited to deal with verification problems. Moreover, such models allow exploiting unbalanced training sets that collect a very small number of examples belonging to the heterogeneous class, while the training of an MLP in this situation is really very difficult.

The present paper reports on the results of the training, tuning and validation of Neurosoft-Sebia (Sebia SA, Evry, Paris, France), an expert system based on six AANNs, each of which is dedicated to processing a distinct fraction, with a view to classi-

fying electrophoresis patterns produced by Capillarys™2 (Sebia SA) into two categories, regular or anomalous, and to identifying curve fractions presenting anomalous characteristics.

Materials and methods

Capillarys™ system

Capillarys™2, an automatic capillary electrophoresis system (11), is designed to perform serum and urine electrophoresis tests. The system exploits liquid flow electrophoresis using eight very narrow capillary tubes functioning concurrently and allowing a throughput of 90 serum protein results per hour. Direct protein detection, achieved by measurement at 200 nm, provides an electrophoretic pattern including albumin, α1-, α2-, β1-, β2- and γ-fractions.

Neurosoft-Sebia architecture

The electrophoresis patterns processed by Neurosoft-Sebia were originally represented by 300 pairs of x-y coordinates. Since the x-y representation is not particularly suited to locate the anomalies of patterns, a pre-processing phase is carried out to extract, for each fraction, a set of features which allow to describe the morphological characteristics of the patterns. The system core was made up by a committee consisting of six trained AANNs (3), each one specialized to process a distinct fraction.

The AANN neurons were organized in three layers: input, hidden and output. Each network input is a vector of features that are extracted from the corresponding fraction, and its output predicts the input class (anomalous or regular). The predictions produced by the six AANNs were merged to assess whether the entire electrophoresis pattern is regular or anomalous. If an AANN misclassifies one or more fractions, the user can provide feedback in order to adapt the behavior of Neurosoft-Sebia to his/her own experience and evaluation criteria (Figure 1).

To verify whether a given input vector belongs to the regular class or not, a trained AANN was used. The input vector U was fed to the input units of the network and the signals were propagated to compute the activation of the hidden and output neurons. The Euclidean distance between the output vector and the original input vector was computed once the output $O(U)$ was obtained. The classification was performed by comparing $d = \|U - O(U)\|$ with a threshold distance d_{thr}. If $d < d_{thr}$ then the input pattern was recognized as belonging to the regular class, otherwise it was considered as anomalous. The value d_{thr} is automatically computed by the expert system for each fraction. However, such values can be updated by the users to customize the behavior of Neurosoft-Sebia. In fact, decreasing d_{thr} enhances the sensitivity of the system; on the contrary, the increase of d_{thr} corresponds to improving its specificity.

All the features that were considered during the development of Neurosoft-Sebia represent the morphological characteristics of the patterns; quantitative values were not included in the feature set, since quantitative anomalies can be determined using a rule-based algorithm.

The chosen features can be divided into two categories: approximation and differential.

The key concept of approximation features was to approximate the electrophoresis pattern or a particular fraction, using a smooth function. The approximation error (obtained by computing the distance between the pattern and the approximation) could be used as a function describing the smoothness of the input pattern (Figure 1, inside the circle).

The differential analysis of a function provided information on the number of minimum and maximum points, their convexity and concavity and, in general, on their smoothness. Neurosoft-Sebia estimated the first, second and third derivatives of each fraction. Features used to describe the associated fraction, for each derivative, were, e.g., the number of times it changes sign, its area and its maximum and minimum values.

The system can represent each fraction by using a different set of features. Currently, albumin, α1, β1 and β2 are represented by 30 features, and α2 and γ, by 31 and 47 features, respectively, as they make up the majority of the anomalies in the γ-fraction from a statistical standpoint.

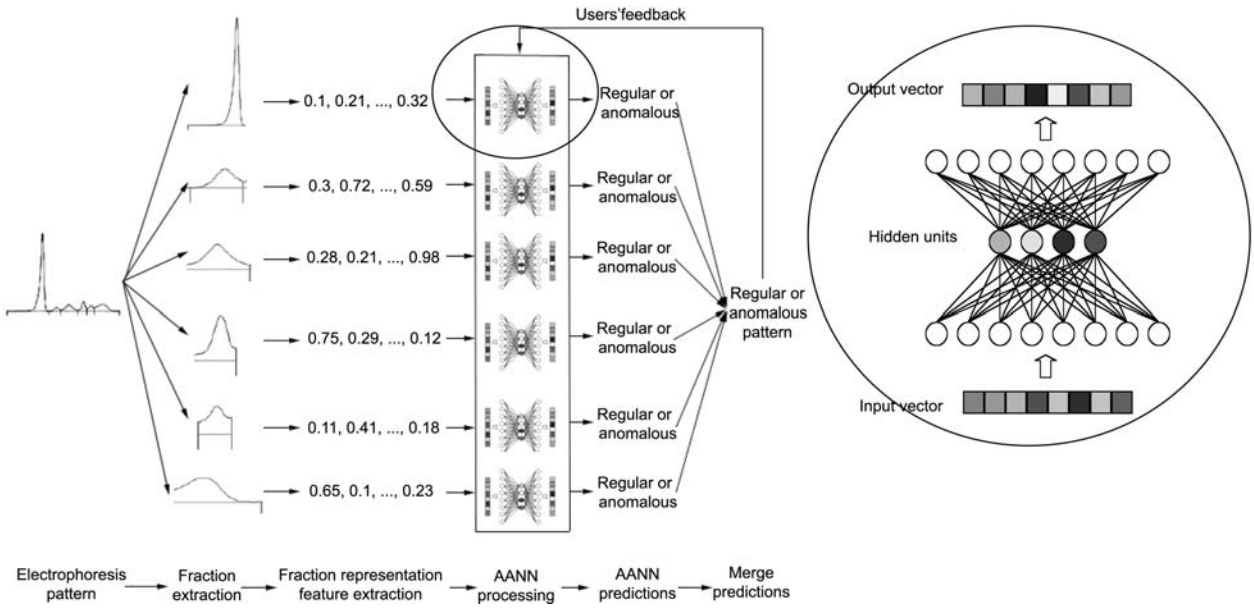

Figure 1 Design of the Neurosoft-Sebia system.

Neurosoft-Sebia training, tuning and validation phases

To evaluate and optimize the performances of the expert system, an evaluation procedure, divided into three distinct phases, was carried out.

During the first phase, approximately 5000 fractions were obtained using the Sebia Capillarys™2 and analyzed by four human experts working in Sebia (the manufacturer of the electrophoresis system to which Neurosoft-Sebia is applied). To reduce the presence of ambiguous situations, only the fractions obtaining a univocal analysis result (regular or anomalous) were considered as training examples, which made up the initial knowledge base.

At the end of the training phase, to evaluate the performances of the trained system, three distinct sets, each one composed of 1000 different electrophoresis patterns (total 3000 patterns), were analyzed in three different Italian laboratories (Forlì, Bergamo and Padova). The classifications (regular or anomalous) made by the human experts working in these laboratories were compared to those carried out by Neurosoft-Sebia using the initial knowledge base. The laboratories were chosen on the basis of the degree of experience of the personnel; each laboratory is located in a university hospital, and carries out a relevant number of electrophoresis each year (47,000, 100,000 and 125,000 analyses/year, respectively). Furthermore, the operative procedure employed for routine analysis by the different laboratories involved are very similar, all of them using capillary electrophoresis; the mean time that each operator spends in evaluating electrophoresis (from 25 to 30 s/sample) is also comparable. At the end of the tuning phase, the discordances between Neurosoft-Sebia and human expert classifications were added to the initial knowledge base, simulating the enrichment of the training set by means of users' feedback, and the system was retrained considering the new learning data thus creating the enriched knowledge base.

Finally, during the validation phase, the performances of Neurosoft-Sebia were evaluated using a public benchmark dataset (http://www.dii.unisi.it/~neurosoft/nsbd.html) composed of 1000 patterns, comparing the predictions obtained using both the initial knowledge base and the enriched knowledge base. The patterns belonging to the benchmark dataset were classified by Neurosoft-Sebia, by a human expert in the Padova laboratory and by three human experts in Sebia. Then, the provided classifications were compared among them to determine both the performances of the system and of the human experts. For each comparison, the predictions of the evaluated subject (Neurosoft-Sebia or a human expert) were compared against the evaluations of the remaining subjects. The results are reported regarding sensitivity and specificity, where sensitivity is defined as the number of anomalous patterns identified, divided by the number of anomalous patterns identified plus the number of ''false regular'' (classified as regular by the subject and anomalous by the remaining ones), while specificity is defined as the number of regular patterns identified, divided by the number of regular patterns identified plus the number of ''false anomalous'' (classified as anomalous by the subject and regular by the remaining ones).

Results

In the training phase, 2685 fractions were univocally classified (2598 as regular and 87 as anomalous), thus defining the initial training set (Table 1).

Table 1 Initial training set composition.

Fraction	Regular	Anomalous	Total
Albumin	541	0	541
α1	538	4	542
α2	542	3	545
β1	410	0	410
β2	256	15	271
γ	311	65	376

Subsequently, on the basis of this initial knowledge base, the ability of Neurosoft-Sebia to classify the curves was tested, comparing its results with those of the experts working in the three chosen laboratories. Table 2 reports the results showing the classification discordances between Neurosoft-Sebia and human experts in relation to each electrophoresis pattern fraction.

Figure 2 shows some examples of conflicting results in electrophoretic patterns. In each electrophoretic pattern, Neurosoft-Sebia provides an interpretative aid identifying the degree of compliance of each fraction with its classification as regular or anomalous (green = regular, orange = anomalous). The majority of discrepancies in terms of false negatives were found for albumin (Figure 2A) and γ-fraction (Figure 2B), while false anomalous fractions were found in β-fractions in particular (Figure 2C).

To retrain Neurosoft-Sebia, thus reducing the disagreements and improving its performances, all the discordant patterns were added to the training data, with human expert results being used as target values.

Subsequently, 1000 additional electrophoresis patterns belonging to a public benchmark dataset were used to evaluate Neurosoft-Sebia. The patterns belonging to such a dataset were evaluated by four human experts, obtaining some ambiguous situations due to contradictory evaluations provided by different experts. To determine the performances of Neurosoft-Sebia and of the human experts, we chose to select, for each comparison, only the subset of the benchmark dataset constituted by the patterns that were unanimously evaluated by all the subjects, except for the evaluated one (Table 3). It is worth noting that the higher the number of regular examples in the selected

Table 2 Discordances between laboratory human experts and Neurosoft-Sebia in the classification of electrophoretic patterns.

Fraction	Regular according to Neurosoft-Sebia and anomalous according to human experts (false negative)	Anomalous according to Neurosoft-Sebia and regular according to human experts (false positive)
Albumin	10	1
α1	0	0
α2	0	0
β1	12	142
β2	8	42
γ	26	0

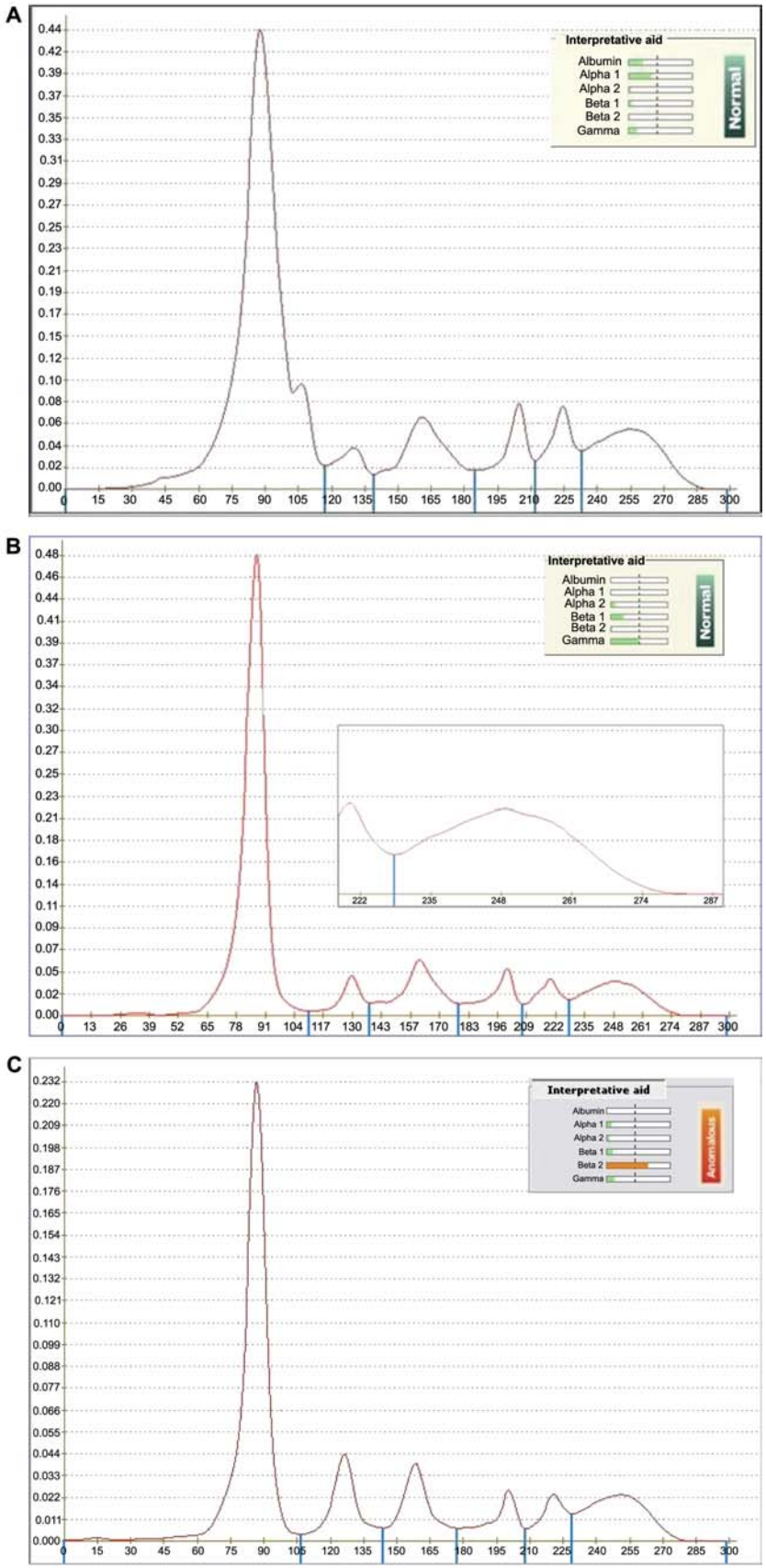

Figure 2 Examples from pattern discordances between Neurosoft-Sebia and human experts' classification of the electrophoresis patterns.
(A) Bisalbumin classified as regular by Neurosoft-Sebia (d=0.450, d_{thr}=0.730); (B) γ-fraction classified as regular by Neurosoft-Sebia (d=0.740, d_{thr}=0.746); (C) β2-fraction classified as anomalous by Neurosoft-Sebia (d=0.950, d_{thr}=0.730). d is the value produced by the system for each analyzed fraction; d_{thr} is the classification threshold. Green, classification as regular; orange, clarification as anomalous.

Table 3 Benchmark subset composition.

Evaluated subject	Regular	Anomalous
Neurosoft-Sebia	559	192
Expert 1	578	209
Expert 2	590	195
Expert 3	569	223
Expert 4	621	195

The subset associated with each expert was created taking into consideration the evaluation of the other experts on the whole benchmark dataset and selecting only the curves which obtained unanimous classifications.

subsets the higher the sensitivity of the evaluated subjects.

Finally, Table 4 reports the sensitivity and specificity obtained by Neurosoft-Sebia using both the initial and the enriched knowledge base, together with the results obtained by the human experts.

Discussion

Neurosoft-Sebia, an expert system based on NNs and designed to classify electrophoresis curves, has the ability to improve its knowledge on the basis of new information obtained from novel electrophoresis patterns. The present paper describes the training, tuning and validation of this expert system to classify serum electrophoretic patterns, obtained using a typical capillary technique used in many laboratories to carry out routine serum protein electrophoresis.

The training of an automatic classifier calls for a large body of data to provide examples representing the real-world data distribution (12, 13). On the contrary, relatively small datasets were used in the development of the expert system described in the present paper. This choice mainly depended on the difficulty involved in assigning a unique target label to each training example and, therefore, in providing the system with unequivocal examples.

On considering the classification of electrophoresis patterns, the learning procedure becomes extremely difficult if a large set of patterns is randomly selected from a laboratory database, since, in this applicative context, an uncontrolled training set is likely to contain many contradictory examples. On the other hand, the selection of a set of carefully validated patterns makes it easier to learn the distinctive characteristics of the regular and the anomalous classes.

In our study, the initial system training was carried out by employing electrophoresis patterns, thus obtaining a univocal analysis result (regular or anomalous) which made up the initial knowledge base of the system. A comparison then made between the performances of Neurosoft-Sebia and the performances of expert physicians, and the discordances found were added to the training set (feedback). Finally, the retrained system was re-tested to verify the improvement of its performances.

The algorithm, which includes training, tuning and validation of the expert system appears to be of value in view of the difficulty inherent to the definition of a unique evaluation criterion between different operators.

The results obtained in terms of sensitivity and specificity also show how human experts disagree in some predictions, and how Neurosoft-Sebia behaves like an ''average human expert'', with an ability to select anomalous fractions (sensitivity) that is equal to, or even greater, than that of each physician.

The specificity of Neurosoft-Sebia, which remains lower than those of human experts, even after the retraining phase, is more than acceptable, the identification of anomalous patterns being the target of the expert system. In other words, a review by human experts of all the patterns designed as ''anomalous'' by the system seems to be mandatory, while its capacity to automatically identify and release regular patterns is highly requested in routine practice.

Furthermore, the system evaluated in the present paper allows the users to update its behavior providing the system with feedback, indicating misclassified patterns and also enriching the training data. Periodically, Neurosoft-Sebia can be retrained by exploiting the enriched learning set in order to personalize its behavior and create a ''confidence relationship'' between the system and human experts.

The findings obtained in the present study demonstrate that Neurosoft-Sebia is reliable in classifying electrophoresis patterns: it allows continuous system training, thus constantly improving performances, enables rapid and standardized interpretations in the evaluating process for serum electrophoresis, saves time, minimizes the need for personnel, and maximizes the clinical availability and the usefulness of protein analyses.

In our laboratory, e.g., located in a university hospital, where approximately 500 serum protein electrophoresis patterns are analyzed daily, being 30% of the

Table 4 Sensitivity and specificity values obtained by Neurosoft-Sebia and by four human experts.

	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Neurosoft-Sebia initial knowledge base	66.67 (61.33–72.01)	58.29 (53.74–62.84)
Human expert 1	92.82 (91.02–94.63)	95.67 (94.25–97.10)
Human expert 2	98.46 (97.60–99.32)	94.41 (92.00–96.01)
Human expert 3	87.89 (85.62–90.16)	98.95 (98.23–99.66)
Human expert 4	99.49 (99.00–99.98)	88.89 (86.73–91.05)
Neurosoft-Sebia enriched knowledge base	97.40 (96.26–98.53)	79.07 (76.16–81.98)

CI, confidence interval.

proportion of regular curves, Neurosoft-Sebia might allow the selection of anomalous curves that need visual inspection, saving approximately 2 h/day (on the basis of a mean value of 10 s for the inspection of regular curves). Therefore, in other laboratories, where there are a greater number of patients showing a regular electrophoresis pattern, the use of this expert system could lead to an even more interesting optimization and rationalization of resources.

Acknowledgements

We acknowledge Dr. Arialdo Vernocchi and Dr. Cosimo Ottomano and all the physicians working with them at the Clinical Laboratories of Ospedale Pierantoni in Forlì and Ospedali Riuniti in Bergamo (Italy) for their cooperation in evaluating serum protein electrophoresis patterns.

References

1. O'Connell TX, Horita TJ, Kasravi B. Understanding and interpreting serum protein electrophoresis. Am Fam Physician 2005;71:105–12.
2. Bossuyt X. Advances in serum protein electrophoresis. Adv Clin Chem 2006;42:43–80.
3. Kramer MA. Autoassociative neural networks. Comput Chem Eng 1992;16:313–28.
4. Waibel A, Hanazawa T, Hinton G, Shikano K, Lang K. Phoneme recognition using time-delay neural networks. IEEE Trans Acoust Speech Signal Process 1989;37: 328–39.
5. Le Cun Y, Boser J, Denker D, Henderson D, Howard R, Hubbard W, et al. Backpropagation applied to handwritten zip code recognition. Neural Comput 1989;1:541–51.
6. Kratzer MA, Ivandic B, Fateh-Moghdam A. Neuronal network analysis of serum electrophoresis. J Clin Pathol 1992;45:612–5.
7. Manner GA, Schweiger CR, Soregi G, Pohl AL. Detection of monoclonal gammopathies in serum electrophoresis by neural networks. Clin Chem 1993;39:1984–5.
8. Ognibene A, Motta R, Caldini A, Terreni A, Dalla Dea E, Fabris M, et al. Artificial neural network-based algorithm for the evaluation of serum protein capillary electrophoresis. Clin Chem Lab Med 2004;42:1451–2.
9. Werbos PJ. The roots of backpropagation: from ordered derivatives to neural networks and political forecasting. New York: J. Wiley, 1994.
10. Gori M, Scarselli F. Are multilayer perceptrons adequate for pattern recognition and verification? IEEE Trans Pattern Anal Mach Intell 1998;11:1121–32.
11. Yang Z, Harrison K, Park YA, Chaffin CH, Thigpen B, Easley PL, et al. Performance of the Sebia CAPILLARYS 2 for detection and immunotyping of serum monoclonal paraproteins. Am J Clin Pathol 2007;128:293–9.
12. Haykin S. Neural networks: a comprehensive foundation. Upper Saddle River, NJ: Prentice Hall, 1994.
13. Mitchell TM. Machine learning. New York: McGraw-Hill, 1997.