ELSEVIER

Journal of Chromatography A, 828 (1998) 489–496

JOURNAL OF CHROMATOGRAPHY A

Artificial neural network classification based on capillary electrophoresis of urinary nucleosides for the clinical diagnosis of tumors

Ruihuan Zhao[a,*], Guowang Xu[a], Bingfang Yue[a], H.M. Liebich[b], Yukui Zhang[a]

[a]*National Chromatographic R. & A. Centre, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 161 Zhongshan Road, 116011 Dalian, China*

[b]*Medizinische Universitätsklinik, Otfried-Müller-Strasse, D-72076 Tübingen, Germany*

Abstract

Nucleosides in human urine have been studied frequently as a possible biomedical marker for cancers, acquired immune deficiency syndrome (AIDS) and the whole-body turnover of RNAs. A capillary electrophoretic method that can quantitatively analyze urinary normal and modified nucleosides in less than 40 min with a good resolution and sufficient sensitivity has been developed. Twelve kinds of normal and modified nucleosides were determined in urine samples from 25 healthy persons and 25 cancer patients of 14 kinds of cancers. Artificial neural networks have been used as a powerful pattern recognition tool to distinguish cancer patients from healthy persons. The recognition rate for the training set reached to 100% and above 85% of the members in the predicting set were correctly classified. In addition, the neural network technique was compared with methods of the principal component analysis and the canonical discriminant analysis. The results demonstrate that the predictive ability of the artificial neural network is stronger than the others in this study.

Keywords: Chemometrics; Neural networks, artificial; Principal component analysis; Canonical discriminant analysis; Nucleosides

1. Introduction

Nucleosides in human urine are of interest to scientists concerned with diagnosis and therapy of cancers [1–3]. It has been shown that nucleosides are excreted in abnormal amounts in the urine of cancer patients [4,5]. Elevated nucleoside concentrations have been suggested to be used as possible markers for leukemia, lymphoma, small cell lung cancer, esophagus cancer, breast cancer, nasopharyngeal cancer, brain cancer, bronchogenic carcinoma, colorectal carcinoma, cancer of the urinary organs or female genital tract, Hodgkin's diseases and the whole-body turnover of RNAs [6,7].

Reversed-phase high-performance liquid chromatography (RP-HPLC) [5,6,8] and immunoassays [7,9] have been used as the main analytical methods for urinary nucleosides. Recently, a new capillary electrophoretic method for normal and modified nucleosides in urine from healthy persons and cancer patients has been developed and optimized [10,11]. A data set of 12 normal and modified nucleosides in 25 urine samples from healthy persons and 25 from cancer patients were obtained by means of this method. A powerful pattern recognition method is

*Corresponding author.

0021-9673/98/$ – see front matter
PII: S0021-9673(98)00589-5

needed for correlating the nucleoside levels and cancers.

Artificial neural network (ANN) is a developing branch of chemometrics and particularly suitable for solving nonlinear multivariate problems such as modeling, optimizing and pattern recognizing, etc. [12–19]. It has been proved to be a most promising method for application in analytical chemistry [20–32].

The purpose of this work was to apply artificial neural networks to study the classification of urinary nucleosides as cancer markers. Binary values were used to represent two groups of persons, zero for normal cases and one for cancer patients. Neural networks are trained by back propagation algorithm to correlate normal case or cancer case with the nucleoside levels in the corresponding urine samples. The results of classification by the neural network were compared with those by principal component analysis and the canonical discriminant analysis. It was proved that the ANN method was the best one in this study.

2. Theory

A multilayer feedforward network model is composed of layers which consist of a large number of simple computational nodes (see Fig. 1), which are fundamental processing elements analogous to neurons in biological neural network [12]. Each node has a series of weighted inputs, which may be either external signals or outputs from other nodes. Nodes in the first layer or input layer perform no special processing but simply distribute the input vector to the next layer. For the nodes in the hidden and output layer, their inputs are weighted and the weighted sum net_{jk} is given by

$$net_{jk} = \sum_i w_{ij} o_{ik} + \theta_j \quad (1)$$

where o_{ik} represents the output of node i of previous layer in iteration k; w_{ij} is the weight between node i and node j; θ_j is a bias term equal to 1.0 multiplied by a weight responsible for accommodating non-zero offset in the data.

The weighted sum net_{jk} is transformed by a transfer function f, then the node output is obtained from the equation $o_{jk} = f(net_{jk})$ and passed to the succeeding layer. The last layer is output layer consisting of one node for each variable to be investigated. In this paper, the following nonlinear input–output transformation function, sigmoid function [15] was used.

$$f(net_j) = \frac{1}{1 + e^{-(net_j / t_j)}} \quad (2)$$

where t_j is a gain, which can be adjusted and determines the shape of sigmoid function. A small value for the gain gives the sigmoid function a very steep transition from 0.0 to 1.0, whereas a large

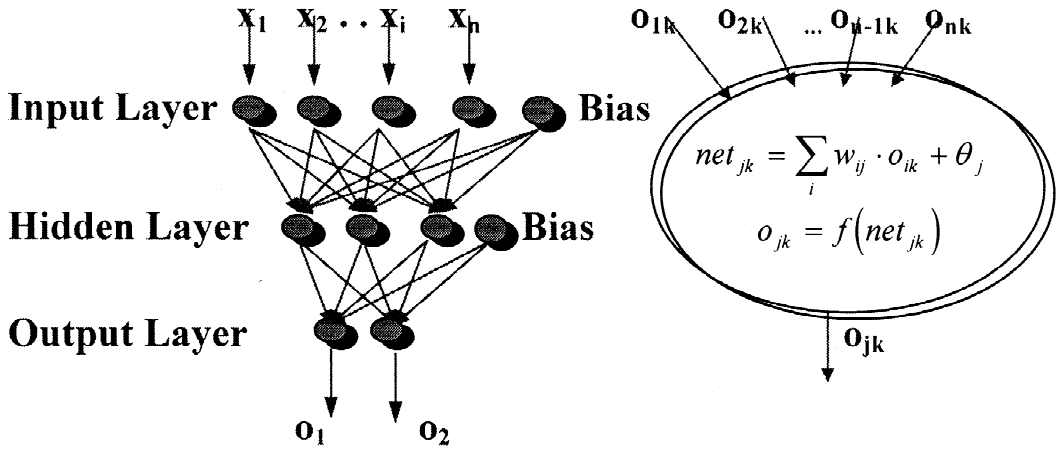

Fig. 1. A schematic view of the architecture of a three layer artificial neural network.

value gives a more gentle slope for the transition [15].

There are a lot of learning algorithms used to train neural networks. A frequently used one is called the error back propagation algorithm [12–15]. The solution of the weights in a feedforward neural networks is a multivariate optimization problem where the connective weights are the variables to be optimized in accordance with a cost function equal to the sum of squared residual errors. Generally the error function can be expressed as:

$$E = 1/2 \cdot \sum_{p=1}^{N} E_k = 1/2 \cdot \sum_{p=1}^{N} \sum_{q=1}^{M} (y_{pq} - \hat{y}_{pq})^2 \qquad (2')$$

where y_{pq} is the expected output value of node q in output layer for the training set and $\hat{y}_{pq}$ is the corresponding actual output, N is the pattern number in training set and M is the number of output nodes. The weights are adjusted by a gradient descent optimizing scheme and are improved by taking incremental changes that are proportional to $\partial E/\partial w_{ij}$, so in iteration $k+1$, weights can be modified practically by the following equation.

$$w_{ij}(k+1) = w_{ij}(k) + \alpha \cdot \left(\frac{\partial E}{\partial w_{ij}}\right)_{|w=w(jk)|} + \beta[(w_{ij}(k) - w_{ij}(k-1)] \qquad (3)$$

where α is the learning rate. A momentum term $0.0<\beta<1.0$ is frequently added to the weight correction in order to stabilize oscillation in the weights which occurs in the learning process. Conjugate gradient principle can be adapted to get a much faster convergence speed and better fitting model [18].

A more in-depth discussion of the theory concerning the artificial neural networks may be found in the book of Rumelhart et al. [12]. Details of the theory about principal component analysis, stepwise discriminant analysis, canonical discriminant analysis may be found in the literatures [33–35].

3. Experimental

3.1. Materials and methods

The collected urine samples were immediately frozen and stored at −20°C. For the analysis of ribonucleosides, the samples were defrosted at room temperature [10,11].

A phenyl boronate gel-affinity chromatographic method [9] was used to isolate the urinary nucleosides with a phenyl boronate column (Affi-gel, Bio–Rad). An extraction factor of 10 in comparing with the original urine has been achieved.

All separations were performed on a Dionex CES I capillary electrophoresis system in a 565 mm (500 mm to detection window)×50 μm I.D. and a 500 mm (437.5 mm to detection window)×50 μm I.D. uncoated capillary from Grom. Samples were introduced by gravity injection at 100 mm head height for 45 s with 10–15 μl of sample size.

The composition of the buffers and the voltage were varied to optimize the conditions for the separation of normal and modified nucleosides [11]. The optimal buffer for the separation of nucleosides extracted from urine was 300 mmol sodium dodecyl sulfate (S.D.S)-25 mmol borate–50 mmol phosphate (pH 6.7). The optimal voltage was 7.0 kV with the current of 47–49 μA. Under these conditions, all nucleosides of interest could be baseline-separated and the reproducibility of migration times and the peak areas was very good.

On-column UV detection was performed at 260 and 210 nm in sequence [10]. Data were collected using a dedicated computer system with a Dionex AC interface and Dionex AI-450 software. The peaks were identified by: (1) comparing migration times of the unknown peaks with those of the standard nucleosides eluted under the same conditions, and (2) by spiking the sample with stock standard solutions of nucleosides. The calibration curves plotting the peak areas vs. concentrations of the nucleosides were obtained by eluting the three standard solutions with different concentrations extracted under the same conditions as urine samples. This enabled the direct determination of nucleoside concentrations (μM) in urine, then transformed into nmol per μmol creatinine.

Urine samples were collected from 25 healthy persons (13 females, 12 males) and 25 cancer patients (nine females, 16 males) suffering 14 kinds of cancers including breast, bronchial, esophagus, rectum, hypopharynx, prostate, thigh, follicular, anaplastic, bladder, endocrine, floor of mouth cancer,

as well as glioblastoma and oligo-dendroglioma. These samples were analyzed under the optimum separation condition. Fig. 2 (260 nm only) shows the typical electropherograms of normal and modified nucleosides from urine of a cancer patient. The separation of normal and modified nucleosides was satisfactory.

3.2. Data processing

A data set of urinary nucleosides containing 50 patterns representing 25 healthy persons and 25 cancer patients has been obtained. Each pattern was described by 12 feature variables which were the concentrations of 12 kinds of nucleosides: Pseu, pseudouridine; U, uridine; C, cytidine; mU, 3-methyluridine + 5-methyluridine; I, inosine; m1I, 1-methylinosine; ac4C, *N*4-acetylcytidine; G, guanosine; m1G,1-methyl-guanosine; A, adenosine; X, xanthosine and m2G, 2-methylguanosine.

Preprocessing is numerically operating on the data (measurements) in order to increase the useful information for pattern recognition. According to the characteristics of sigmoid transformation function,

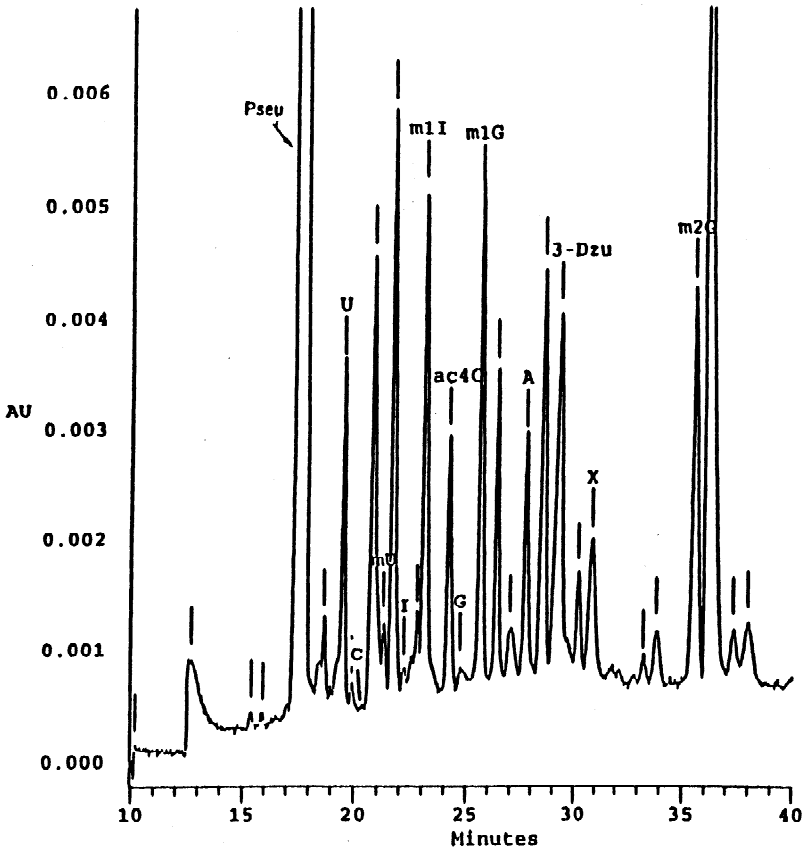

Fig. 2. Typical electropherograms of normal and modified nucleosides extracted from urine of a cancer patient. Sample: nucleosides extracted from a spontaneous urine from a patient with breast cancer; detection wavelength: 260nm; buffer: 300 m SDS–25 m*M* borate–50 m*M* phosphate pH 6.7 voltage: 7.0 kV with 49 μA of current; peak identification: Pseu, pseudouridine; Dhu, dihydrouridine; U, uridine; C, cytidine; mU, 3-methyluridine + 5-methyluridine; I, inosone; M1I, 1-methylinosine; ac4C, N4-aetylcytidine; G, guanosine; M1G, 1-methylguanosine; A, adenosine; 3-Dzu, 3-deazauridine; X, xanthosine; m2G, 2-methylguanosine; m6A, N6-methyladenosine.

the node output value lies in the range of 0 to 1. Therefore, 12 nucleoside concentrations were all normalized to interval of (0.1) using the following min–max procedure.

$$\text{factor} = \frac{(x_{i\ \max} - x_{i\ \min})}{(l_{\text{high}} - l_{\text{low}})}$$
$$x_{im} = \frac{(x_{i\ \text{old}} - x_{i\ \min})}{\text{factor} + l_{\text{low}}} \quad (4)$$

where $x_{i\ \max}$, $x_{i\ \min}$ are the maximum and minimum values for feature variable $x_{i\ \text{old}}$. l_{high} and l_{low} are two ends of the interval, in this case, 0 and 1.

3.3. Software

The back-propagation learning algorithm described in Theory section was programmed into a simulation software and used to develop neural network models. This software was implemented by authors using Borland C++ programming language and run on a IBM-486 compatible personal computer with Windows operating system and 4 Mb memory.

The principal component analysis, stepwise discriminant analysis, canonical discriminant analysis methods had been included in our statistical software package for special usage of pattern recognition.

4. Results and discussion

4.1. Feature selection

The problem of feature selection is to find the optimum combination of features. Features which are not relevant to the classification problem should be eliminated. In almost all chemical applications of pattern recognition the number of original raw features is too large and a reduction of the dimensionality is necessary.

The mean levels of 12 kinds of normal and modified nucleosides in urine of healthy persons and cancer patients have been established (Table 1). Table 1 shows that the levels of nucleosides in urine from cancer patients were generally elevated. According to their higher mean excretion levels, Pseu, m1I, m2G, m1G and ac4C may be more significant for pattern recognition of cancers.

Stepwise discriminant analysis was used to select feature variables. When SDA was performed on capillary electrophoretic data (original) containing 12 components of nucleosides, five nucleosides were selected as the variables which were more important for classification of healthy persons and cancer patients. They were Pseu, m1I, ac4C, m1G, m2G and consistent with the significance levels suggested by

Table 1
Individual nucleoside levels in urine of healthy persons and patients with cancers

Compound		Cancer patients			Healthy persons		
		Mean	S.D.	R.S.D. (%)	Mean	S.D.	R.S.D. (%)
1	Pseu	48.446	24.916	0.514	25.320	10.324	0.408
2	U	0.745	0.344	0.462	0.468	0.191	0.409
3	C	0.084	0.106	1.271	0.070	0.095	1.349
4	mU	0.735	2.125	2.891	0.089	0.123	1.383
5	I	0.486	0.504	1.037	0.141	0.099	0.704
6	m1I	2.517	1.525	0.606	1.270	0.475	0.374
7	ac4C	1.231	0.821	0.667	0.600	0.384	0.640
8	G	0.288	0.352	1.222	0.010	0.021	2.128
9	m1G	1.400	0.831	0.593	0.817	0.298	0.365
10	A	0.509	0.574	1.126	0.176	0.167	0.952
11	X	0.914	1.004	1.099	0.450	0.260	0.577
12	m2G	1.549	1.035	0.668	0.386	0.202	0.523

The results are the average concentrations of nucleosides (nmol μmol^{-1} creatinine) from 25 healthy persons and 25 patients with 14 kinds of different cancers.

mean levels of nucleosides in urine. So, the five nucleosides were selected as feature variables to construct a five dimensional pattern space for the classification in this study.

4.2. Neural network optimization

The performance of a neural network is effected by the following parameters: network architecture, initial weight value, learning rate, momentum term and gain of sigmoid function. Optimization of a neural network is difficult and time-consuming. The first step is to find the best neural network architecture including the number of hidden layers and nodes in each hidden layer. It has been claimed that an arbitrary nonlinear mapping of an input domain to an output one can be achieved by using three layers in a neural network [12]. So neural networks with three layers were used in this work. Since there is no theoretical way to choose the number of hidden nodes. It is necessary to test networks with different numbers of hidden nodes and choose the best configuration in practice.

Networks with five input nodes, one output node and hidden nodes ranging from 1 to 6 were tested. All weights were initialized to random values between -0.1 and 0.1. The learning rate and the momentum term were varied for each of the networks, with each configuration being trained several times from different starting points (i.e. initial random weights).

It was found that the network architecture consisting of five input nodes, one output node and three hidden nodes, usually gave better results for any given condition. The gain parameter t_j of 1.0 was found to be preferable to give the best classification performance. The optimal learning rate and momentum were 0.5. The suitable training iterations were about 20 000 while multipattern training was adopted and learning rate as well as momentum were adjusted dynamically.

4.3. Results of classification

The data set of 50 patterns was randomly divided into two parts, a training set including 30 patterns (15 healthy persons; 15 cancer patients) used to train a network model and a predicting set including 20 patterns (ten healthy persons; ten cancer patients) used to evaluate the network model.

As the criterion for classification, an output value >0.5 was considered as cancer case and a value <0.5 as normal case. In fact, the great majority of the output values was close to 0.9 or 0.1.

The classification results are shown in Fig. 3 and Table 2. The pattern points close to (0.1, 0.1) and (0.9, 0.9) in Fig. 3 are classified correctly, otherwise it is misclassified. From Table 2, it can be seen that the recognition rate of 100% for the training set and 85% for the predicting set were obtained.

4.4. Comparison with principal component analysis (PCA) and canonical discriminant analysis (CDA)

In order to assess ANN method, the 5-dimensional feature vectors were subjected to PCA and CDA.

Fig. 4 (Eigenvector plot) shows the two-dimensional display for all 50 patterns by PCA. The normal patterns are clustered while the cancer pat-

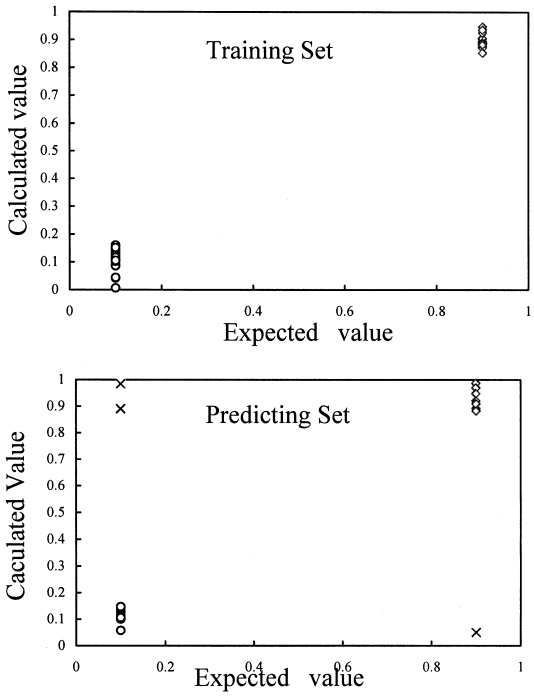

Fig. 3. Plot of output value versus expected value of ANN for the classification of healthy persons and cancer patients. ×, three sample points classified as the fault category.

Table 2
Classification results of nucleoside data by ANN and CDA

Class	Training Set			Predicting Set		
	No.[a]	Healthy	Cancer	No.	Healthy	Cancer
ANN method						
Healthy	15	15	0	10	8	2
Cancer	15	0	15	10	1	9
Recognition rate (%)		100.0		Prediction rate (%)		85.0
CDA method						
Healthy	15	14	1	10	5	5
Cancer	15	4	11	10	3	7
Recognition rate (%)		83.3		Prediction rate (%)		60.0

[a] no = the number of patterns.

terns are rather scattered. A decision line could be found, which separates two areas of healthy persons and cancer patients. It indicates that the five selected features are significant and two classes of patterns are distinguished mostly. The total consistency rate of 50 patterns were 86% for PCA and less than 94% for ANN (Fig. 3).

While performing canonical discriminant method on the same data sets, a decision line was used and all sample points was arbitrarily divided into the two classes. The results are listed in Table 2. It is obvious that both recognition rate and prediction rate by CDA are much lower than those by ANN.

Although there are differences in the classification ability of these three methods, most patterns classified incorrectly (observable in Fig. 4) are the same or similar to some degree. The reason for the existence of a small group of bad patterns partially may be that 25 cancer patients were suffering from 14 kinds of cancers which have different characteristics while urine samples were collected, most of the patients were being treated by chemotherapy and/or radiation, some patients were operated several months or years ago, but the carcinoma recurred and existed at collecting time. As much more samples are collected and more efficient methods applied, the predictive performance should be promoted.

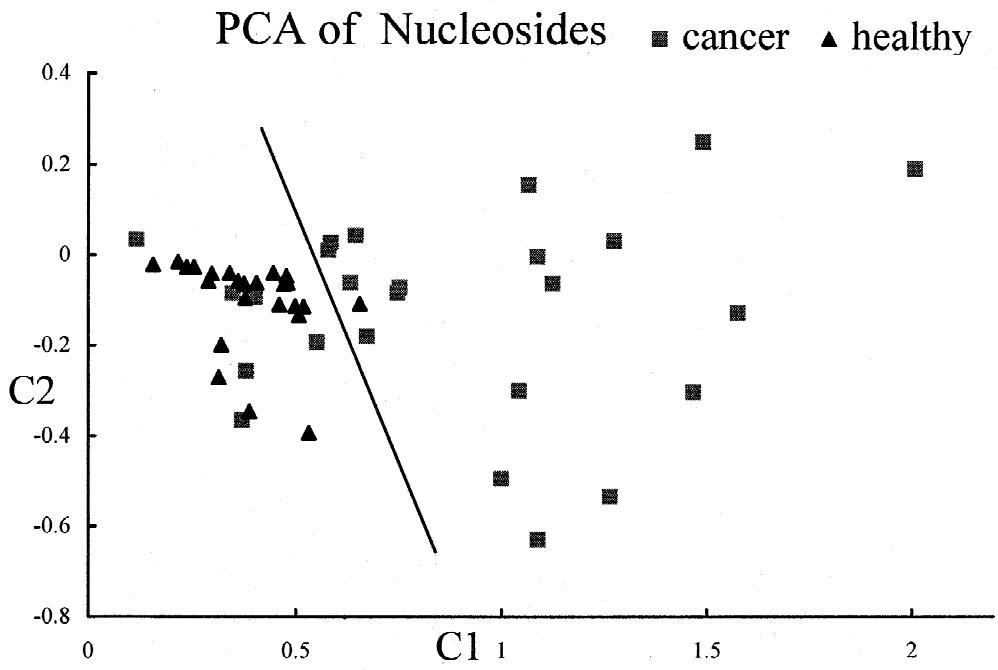

Fig. 4. PCA using five kinds of nucleosides selected as feature variables.

5. Conclusion

In this study, it was suggested that the urinary nucleosides analyzed by capillary electrophoresis were possible to be considered as a candidate of the markers for cancers. Being compared with PCA and CDA methods, the classification by artificial neural network was more satisfactory. More study has to be done to shed light on the usefulness of the modified nucleosides as the tumor markers in clinical diagnosis. Further work is carrying on through analyzing more samples and improving neural network method.

Acknowledgements

Financial support from the National Natural Science Foundation of China is gratefully acknowledged. G.X. gratefully acknowledges the fellowship from the Max-Planck-Institut of Germany.

References

[1] A. Apffel, J.A. Chakel, S. Fisher, K. Lichtenwalter, W.S. Hancock, Anal. Chem. 69 (1997) 1320.
[2] A.S. Cohen, S. Terabe, J.A. Smith, B.L. Karger, Anal. Chem. 59 (1987) 1021.
[3] H.J. Gaus, S.R. Owens, M. Winniman, S. Cooper, L.L. Cummins, Anal. Chem. 69 (1997) 313.
[4] T.P. Waalkes, M.D. Abeloff, D.S. Ettinger, K.B. Woo, C.W. Gehrke, K.C. Kuo, E. Borek, Cancer 50 (1982) 2457.
[5] R.W. Trewyn, R. Glaser, D.R. Kelly, D.G. Jackson, W.P. Graham, C.E. Speicher, Cancer 49 (1982) 2513.
[6] G. Schoech, G. Sander, H. Topp, G. Heller-Schoch, in: C.W. Gehre, K.C. Kuo (Eds.), Chromatography and Modification of Nucleosides, Part C, Elesevier, Amsterdam, 1990, p. C389.
[7] M. Masuda, T. Nishihira, K. Itoh, M. Mizugaki, N. Ishida, S. Mori, Cancer 72 (1993) 3571.
[8] H.M. Liebich, C. Di Stefano, A. Wixforth, H.R. Schmid, J. Chromatogr. A 763 (1997) 193.
[9] C. Reynaud, C. Bruno, P. Poullanger, J. Grange, S. Barbesti, A. Niveleau, Cancer Lett. 61(3) (1992) 255.
[10] H.M. Liebich, G. Xu, C. Di Stefano, R. Lehmann, H.U. Häring, P. Lu, Y. Zhang, Chromatographia 45 (1997) 396.
[11] H.M. Liebich, G. Xu, C. Di Stefano, R. Lehmann, J. Chromatogr. A 793 (1998) 341.
[12] D.E. Rumelhart, J.L. McClelland, PDP Research Group Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Part I, MIT Press, Cambridge, MA, 1986.
[13] P.B. de Harrington, Anal. Chem. 65 (1993) 2167.
[14] Z. Li, Z. Cheng, L. Xu, T. Li, Anal. Chem. 65 (1993) 393.
[15] P.B. de Harrington, Anal. Chem. 66 (1994) 802.
[16] I.V. Tetko, D.J. Livingstone, A.I. Luik, J. Chem. Inf. Comput. Sci. 35 (1995) 826.
[17] I.V. Tetko, D.J. Livingstone, A.I. Luik, J. Chem. Inf. Comput. Sci. 36 (1996) 794.
[18] H.J. Miao, M.H. Yu, S. Hu, J. Chromatogr. A 749 (1996) 5.
[19] I.C. Jou, S.-S. You, L.-W. Chang, Pattern Recognition 27 (1994) 864.
[20] G.M.J. West, J. Chem. Inf. Comput. Sci. 35 (1995) 21.
[21] M.E. Munk, M.S. Madison, E.W. Robb, J. Chem. Inf. Comput. Sci. 36 (1996) 231.
[22] S. Ventrura, M. Silva, D. Perez-Bendito, C. Hervas, J. Chem. Inf. Comput. Sci. 37 (1997) 287.
[23] J. Polanski, J. Chem. Inf. Comput. Sci. 36 (1996) 694.
[24] A. Eghbaldar, T.P. Forrest, D. Cabrol-Bass, A. Cambon, J.-M. Guigonis, J. Chem. Inf. Comput. Sci. 36 (1996) 637.
[25] X.-H. Song, Z. Chen, R.-Q. Yu, Computer Chem. 17 (1993) 303.
[26] J.R. Long, H.T. Mayfield, M.V. Henley, Anal. Chem. 63 (1991) 1256.
[27] J.M. Sutter, P.C. Jurs, Anal Chem. 69 (1997) 856.
[28] T.B. Blank, S.D. Brown, Anal. Chem. 65 (1993) 3081.
[29] G.M. Maggiera, D.W. Elrod, R.G. Trenary, J. Chem. Info. Comput. Sci. 32 (1992) 732.
[30] M. Bos, A. Bos, W.E. van Der Linden, Anal. Chim. Acta 233 (1990) 31.
[31] S.N. Kavuri, V. Venkatasubramanian, Comput. Chem. Eng. 17 (1993) 765.
[32] T.D. Booth, K. Azzaouri, I.W. Wainer, Anal Chem. 69 (1997) 3879.
[33] K.R. Beerwinkle, R.C. Beier, B.P. Mundy, Comput. Chem. 10 (1986) 3.
[34] H. Abe, S. Kanaya, Y. Takahashi, S.-I. Sasaki, Anal. Chim. Acta 215 (1988) 155–168.
[35] K. Varmuza, Pattern Recognition in Chemistry, Springer–Verlag, Berlin, Heidelberg, 1980.