International Journal of Innovation and Applied Studies
ISSN 2028-9324 Vol. 14 No. 2 Jan. 2016, pp. 297-303

http://www.ijias.issr-journals.org/

Contribution to a diagnosis system of the Whewellite- and Weddellite crystals in human urinary using data minig algorithms

A. Ait Ider[1-3], M. Echajia[1], C. Tcheka[2], A. Merbouha[3], and M. Mbarki[1]

[1]Department of chemistry and environment,
Sultan Moulay Slimane University, Faculty of Science and Technology,
Transdisciplinary Team of Analytical Science for Sustainable Development,
P.B 523, Beni Mellal, Morocco

[2]Department of chemistry,
Yaoundé 1 University Faculty of Science, Laboratory of Physical and Theoretical Chemistry,
P.B 812, Yaounde, Cameroun

[3]Department of data processing,
Sultan Moulay Slimane University, Faculty of Science and Technology,
Information Processing and Decision Support,
P.B 523, Beni Mellal, Morocco

ABSTRACT: The goal of this study is to develop a simple intelligent urolithiasis diagnosis system. The accuracy of the system was determined by comparing the recognition rates of the Artificial Neural Networks (ANNs)-, k-nearest Neighbor (kNN)-, and Support Vector Machines (SVM) algorithms. The results showed that the ANN model was superior to SVM and KNN models in prediction. We aimed through this work to classify the subjects in three classes according to the chemical concentrations of variables (Ca, Ox, pCaOx, Ca/ Ox) using and according to their clinical status. The ANN model, used to determine the first class that contains the subjects presenting their urine a calcium oxalate monohydrate (CaC_2O_4,H_2O : whewellite (Wh)) crystal type. This ANN model reached a correct prediction rate of 85.3%. Using SVM- and KNN model the correct prediction rate reached 82.6% and 65.55% respectively. The second class contains the subjects presenting a calcium oxalate dihydrate (CaC_2O_4,$2H_2O$ wedellite (Wd)) crystal type. The ANN-, SVM- and KNN model reached a 93.4%-, 94.2%- and 77.25% correct prediction rate, respectively. In third class that corresponds to the subjects who have negative crystalluria (NC), ANN-, SVM- and KNN model reached a 91.7%-, 87.8%- and 69.77% correct prediction rate, respectively. Compared to SVM- and KNN models, the developed system using ANN model has allowed us to discriminate the subjects. This system is important in clinical laboratories since it could be a helpful tool for provide information about the development, formation of urinary stones crystals and the determination of their crystal type.

KEYWORDS: Urinary Lithiasis, Whewellite, Wedellite, Artificial neural network, k-nearest neighbor, Support Vector Machines , Intelligent Diagnosis System.

1 INTRODUCTION

The disease of urolithiasis is a disease whose prevalence tendency is rising in the world [1], [2]. It is caused by several factors that include frequently calcium oxalate species in the urine. It crystallizes in the form of (Wd) and (Wh) [1], [3], [4], [5], [6], [7]. During diagnosis of the disease no precise decision can be easily made taking into account the individual concentrations of calcium and oxalates those individually. While it is necessary to introduce the two important variables calcium oxalate molar product (pCaOx) and the molar ratio calcium / oxalate (Ca/Ox) in order to ensure a better diagnosis [8].

It appears from the results of most studies in the literature that calcium oxalate is the most frequent and majority urinary crystals crystallize [9], [10], [11], [12], [13], [14]. However, rare are the studies dedicated to the investigation of the influence of the calciuria and the oxaluria on the formation of the crystals calcium oxalate in urines. As a matter of fact, the crystals are a promoting phase to the formation of urinary stones [15], [16]. Most of the studies have used databases of infrared spectroscopy coupled to chemometrics methods in order to determine the chemical composition of urinary stones and to classify them by means of artificial neural networks (ANN) [17]. In the present work, using the ANN-, KNN- and SVM classification algorithms, we have classified the subjects of the database into 3 classes, depending on the chemical concentrations of variables (Ca, Ox, pCaOx, and Ca / Ox) and on their clinical status. The first class contains subjects presenting a Wh crystal type. The second class contains subjects with a Wd crystal type and the third one includes subjects who are unaffected by this disease According to the obtained results of the classification, we developed a simple intelligent diagnosis system. This system is important for medical laboratories since it can provide useful information about the development, formation of the urinary stones. The determination of the crystal types of the urinary stones would be helpful to better manage, economically and socially, the treatment of this rising disease [18]

2 MATERIALS AND METHODS

2.1 SAMPLING

The study was carried out on 106 urine samples from human volunteers from the region of Tadla Azilal.; all samples were analyzed and prepared in our laboratory [15]. Crystalluria has been characterized using optical polarized light microscopy. The apparent oxaluria and urinary calcium were determined by conventional volumetric assays on urines. The database contains 63 men and 43 women among whom 30 (28.3%) cases presented (Wh) crystal type their urine, 27 (25.5%) cases presented a (wd) crystal type and 49 (46.2%) cases would be not affected by this disease (Nc).

Table 1. The 106 samples according to age group

group	<20 years	20-40 years	>40 years	total
Man	5	34	24	63
Woman	5	21	17	43
Global	10	54	42	106

2.2 DATA ANALYSIS

The urine samples were divided into tow subsets, one subset for training and another subset for test. Three training subsets were employed 20 samples, 20 samples and 30 samples for respectively Wd, Wh and NC class. And three tests subsets were employed 10 samples, 7 samples and 19 samples for respectively Wd, Wh and NC. The chemometrics methods that have applied in this work were well known as SVM, KNN and ANN. These algorithms are applied on the same data subset.

2.2.1 K-NEAREST NEIGHBOR ALGORITHM (K-NN)

The k-Nearest Neighbor algorithm (k-NN) [19] is a simple algorithm that stores all the available cases and classifies new cases based on a similarity measure (e.g., distance functions). They have been used in various areas such as bioinformatics [20], image processing and data compression [21], document retrieval [22], computer vision [23], multimedia database [24], and marketing data analysis [25].

To classify a new observation x, the idea is to simulate a vote among the closest neighbors of this observation. The class of x is determined according to the majority class among the k (an integer to be determined) closesets neighbors of the observation x. It is a method for classifying objects based on the closest training examples in the feature space. Given a training set D and a test object x = (x', y') the algorithm computes the distance (or similarity) between z and all the training objects (x, y) belonging to D ((x, y) ∈ D) to determine its nearest-neighbor list, Dz. (x is the data of a training object, while y is its class. Likewise, x' is the data of the test object and y' is its class.) Once the nearest-neighbor list is obtained, the test object is classified based on the majority class of its nearest neighbors [26].

$$\text{majority voting}: y' = \underset{v}{\text{argmax}} \sum_{(xi,yi)\in Dz} I(v = yi)$$

Where v is a class label, yi is the class label for the i th nearest neighbor, and $I(.)$ is an indicator function that returns the value 1 if its argument is true and 0 otherwise.

The k-nearest neighbor classification algorithm:

Input: D, the set of k training objects and test object z (x', y') ∈ D. Select $Dz \subseteq D$ the set of k closest training objects to z.

Process: Compute d (x', x), the distance between z and every object, (x,y)

Output: $y' = \underset{v}{\text{argmax}} \sum_{(xi,yi)\in Dz} I(v = yi)$

2.2.2 SUPPORT VECTOR MACHINES

Support vector machines (SVM) are a set of techniques of supervised learning intended algorithms to solve problems of discrimination and regression. The SVM is a generalization of the linear classifiers [27]. The SVM set up a system model using a set of given training samples for the classification and prediction of the output based on the training samples and input samples. The aim of SVM is to create a distinction between two or more data classes. SVM uses the multi-class formulation described in [28-29], but optimizes it with an algorithm that is very fast in the linear case. One can find a training set (x_1, y_1) ... (x_n,y_n), presenting y_i in [1..k] labels [28][30].

2.2.3 ALGORITHM OF ARTIFICIAL NEURAL NETWORKS

The Artificial Neural Networks (ANN) are used in various disciplines such as economics, ecology, environment, biology and medicine. These networks are applied in particular to solve problems of classification, prediction [31], [32], [33], categorization, optimization, recognition of the forms and the associative memory [34]. The ANN improves the diagnosis in medical sciences. In fact, the ANNs are used, for example, in cases of the prediction of myocardial [35, 36], lung pathologies [37], [38], diabetes [39], cancers [40], Alzheimer's disease [41] and electromyography and kinesiology [42]. In an ANN, the Neurons are connected via synapses (connection links), that modulate signals passing through them. Each synapse has an associated weight w. The net input N is the function of all transmitted signals X_i and their corresponding weights W_i in a neuron: $N = \sum W_i X_i$ (sum of weighted input signals). Each neuron applies an activation (transfer) function to its net input N in order to provide an output signal for each neuron. A neural network is characterized by its architecture or its pattern of connections between the neurons (see fig1). Neurons are arranged in several layers: an input layer that receives the inputs, on or several hidden layers which transform the input representation into a new 'hidden' representation and an output layer whose units send the predicted values out. Input data are signals xi of the input layer. Initial weights are random values. Processing the data from input layer to output layer provides the output, which is the result. A neural network is also characterized by its learning algorithm (*i.e.* its method of determining the weights on the connections). Before using a network for prediction, it must be trained with known data. This is necessary to ensure that the network provides useful results. The most used learning algorithm is based on the 'back-propagation of errors'. While learning, the network compares its output with observed (known) output values of learning data. The effectiveness is usually determined in terms of the root mean squared (RMS) error between the actual and the desired outputs averaged over the learning data. After comparison, the network changes weights backwards from the output layer to the input layer with respect to the output error. A neural network is also characterized by its activation (transfer) function such as hyperbolic tangent function and sigmoid transfer function, in order to determine its output. In this study a sigmoid transfer function was chosen as transfer function [43], [44]. Several research works were made to describe the functioning of ANNs [43], [44], [45], [46], [47], [48], [49], [50], [51], [52] this algorithms is used to improve the diagnosis in medical domain and chemistry [54], [55].

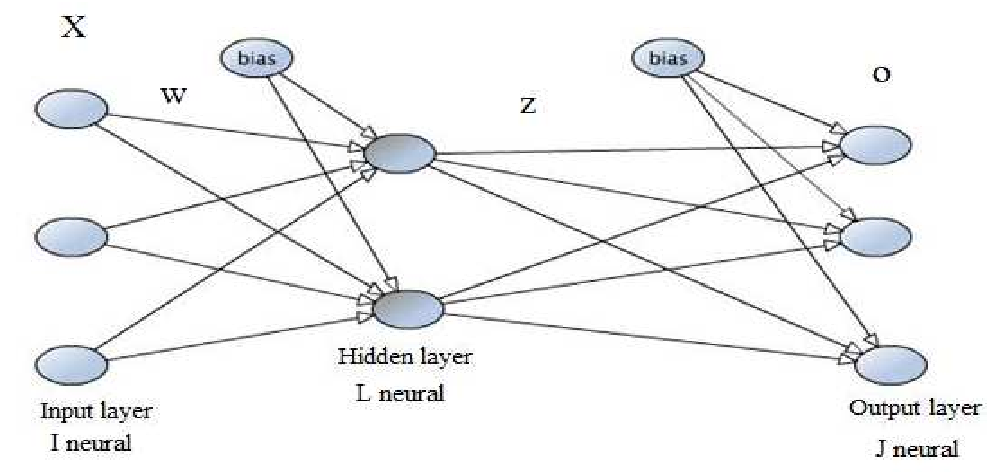

Fig. 1. Neural network architecture

3 RESULTS AND DISCUSSION

In the present study, several classification algorithms were applied to the same database in order to develop a simple intelligent diagnosis system. The aim of this paper was to compare the model performances of three supervised algorithms, namely SVM, KNN and ANN. The obtained results show that ANN and SVM are very powerful methods in prediction. The best results are obtained with the ANNs technique, with lower RMS values than those obtained in the case of the other classification methods. As a matter of fact, in some cases, the error is almost half of that obtained by using KNN

The comparison of the performances of all the models is presented in table 2 and figure 2. Using KNN, the error classification rate for the test samples is: 0.3445 for whewellite, 0.2275 for weddellite and 0.3015 for negative crystalluria. Using SVM, the error classification rate is 0.174 for whewellite, 0.058 for weddellite and 0.12 for negative crystalluria. The error prediction rate, using ANN, is 0.165 for whewellite, 0.066 for weddellite and 0.083 for negative crystalluria.

Therefore, it turns out that the average correct prediction rate for ANN is higher than the one obtained by using SVM and KNN. The results showed that the ANN is a helpful tool for the identification of these types of stones. Moreover we did not observe any unclassified samples in the dataset test.

Table 2. The correct prediction rate using the algorithms (in percentage)

algorithmes	Classe1	Class2	Classe3	Correct prediction rate
kNN	65.55	77.25	69.77	69.85
SVM	82.6	94.2	87.8	88.2
ANN	83.5	93.4	91.7	89.53

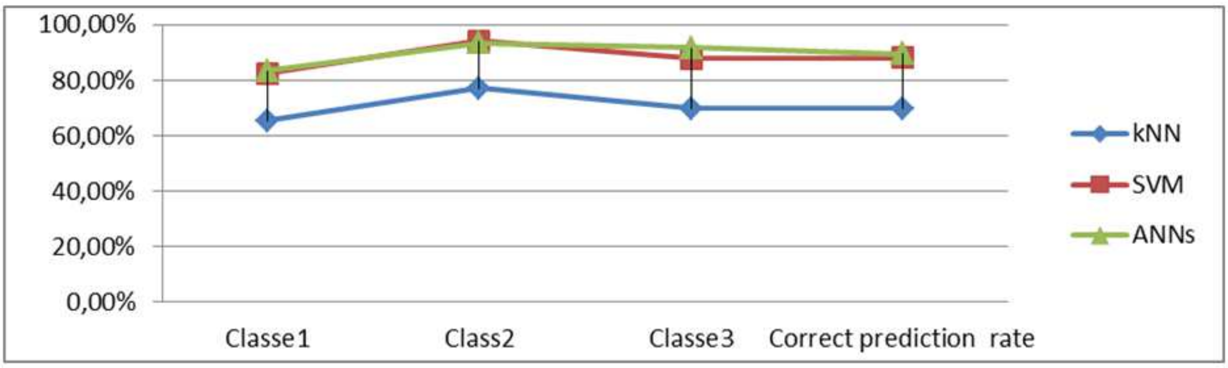

Fig. 2. The comparison of correct prediction rate between supervised algorithms

To explain the error classification rate of (10.47%) using the ANN model algorithm we can say that the data at hand generally include a number limited examples and a part of variable noise. This noise is mainly due to the presence of random disturbance and also to the complexity of the urinary environment. Given that fact that the neural networks are very flexible, there is a risk over fitting that is mistaking the noise for the signal and learning from the noise. Moreover, it is necessary to increase the number of samples in the learning database. This error classification rate can be also explained by the use of not very precise determination techniques of the concentrations in calcium and oxalates (chemical volumetric).

The chemometrics coupled to simple and less expensive chemical and physical analyses (volumetry and optical microscopy), on the one hand and methods of data analysis, on the other hand would be a quick and non expensive tool for a first diagnosis of the urolithiasis that can allow the decision-makers to avoid the high budgets, of coverage, assigned to the treatments of this disease. Thanks to the neural networks, we have been able to establish a decision-making model allowing the competent sanitary services to predict the lithogenous risk in the populations. The results of most works in the literature showed that calcium oxalates is the most frequent as chemical species in urinary and the majority of stones are crystalluria [9], [10], [11], [12], [13], [14]. However, rare are the studies dedicated to the investigation of the influence of urine calcium and apparent oxaluria on the formation of calcium oxalate crystals in the urine; which is known to initiate the formation of urinary stones. These crystals can in some conditions become stones that lead to the failure of urinary system.

4 Conclusion

This study has illustrated that the use of a simple and non expensive physicochemical analyses coupled to classification algorithms have been successfully employed. This coupling technique is fast and simple. It can be very useful for the first diagnostic which is of great advantage in this context. This work has shown that ANN and SVM are very powerful methods to develop a simple diagnostic system. Models based on KNN failed badly in most of the cases. By contrast, ANN and SVM achieved a good prediction performance and yield the highest average correct prediction rate. Using ANN yielded an improvement of the error classification rate for the by comparison to SVM and KNN. According to the obtained results, the ANN was recommended to develop a helpful decision support system. The system developed would be able to class new cases, and thus set up a diagnosis of the disease and the type of lithogenous crystals**.**

Acknowledgements

Our acknowledgements to the Tadla Azilal Regional Hospital and Dr Constant Cheka for providing samples of urinary stones.

References

[1] Renard-Penat R, Ayed A. Diagnostic et bilan des calculs urinaires.Radiologie et imagerie médicale : Génito-urinaire - Gynéco-obstétricale – Mammaire.334-173-C-10.
[2] Fekak H, Sahnoun A, Rabii R et coll. La lithiase urinaire : étude spectrophotométrique infrarouge (à propos de 80 cas). J Maroc Urol 2006;1:17-20.
[3] Daudon M. Epidémiologie actuelle de la lithiase rénale en France : Annales d'urologie 2005 ; 39 :209-31.
[4] Daudon M, Traxer O, Lechevallier E, Saussine C.La lithogenèse. Prog Urol 2008 ; 18 : 815 27.
[5] Daudon M, Hennequin, Jungers P, Lacour B, Drueke T. Les inhibiteurs de cristallisation. In: Funck-Brentano JL, Bach JF, Kreis H,Grunfeld JP (éds). Actualités Néphrologiqes Jean Hamburger - Hôpital Nec-ker, Paris, Médecine-sciences Flammarion, 1994 ; 165-22
[6] Marangella M. Daniele P G, Ronzani M, Sonego S, Linari F. Urine saturation with calcium salts in normal subjects and idiopathic calcium stone-formers estimated by an improved computer model system Urol Res 1985; 189-193.
[7] Daudon M. Apport de l'analyse morphologique et infrarouge des calculs urinaires : Service de Biochimie A Hôpital Necker-Enfants Malades, Paris.
[8] Jungers P, Daudon M, Le Duc A. Paris Flammarion Médecine-Sciences, 1989.
[9] Jungers P, Daudon M, Conort P. Lithiase rénale: diagnostic et traitement, ed Flammarion, paris, 1999.
[10] Dao NQ, Daudon M. Infrared, and Raman spectra of stones. Paris: Elsevier, 1997.
[11] Graf JD. Féraille E. Importance de l'analyse du calcul dans l'exploration de la lithiase rénale. Méd Hyg2000; 58: 431-433.
[12] Laziri F, Rhazi Filali F, Oussama A, Soulaymani A, Qarro A, Lezrek M. Facteurs impliqués dans l'épidémiologie des calculs urinaires marocains. J Maroc Urol 2010 ; 19 : 9 14.
[13] Belhaj R, Hayder N, Gargueh T, Zorguati M, Marrakchi O, Abdelhak S, Lakhoua R, Abdelmoula J. Diagnostic biochimique et moléculaire de l'hyperoxalurie primaire de type 1

[14] Amrani Hassani M, Hennequin C, Lacour B, Daudon M.Citraturie et cristallurie de weddellite. Prog Urol 2005 ; 15 : 650-5.

[15] Tcheka C, Mbarki M, Loura Benguellah B, Ketcha Mbadcam J, Amechrouq A, Amhoud A, Moubarik A. Apport de l'etude de la cristallurie et du dosage chimique a la recherche des parametres de risque de lacristallisation oxalocalcique urinair E, J Maroc Urol 2012 ; 27-28 : 25-32

[16] Daudon M, Brahimi M, Nguyenh V, Réveillaud RJ, Jungers P. Intérêt clinique de différencier les phases cristallines de l'oxalate de calcium dans les urines des sujets lithiasiques oxalo-calciques. Acta Pharm Biol Clin 1987 ; 4 : 461-4.

[17] Kuzmanovski I, Trpkovska M, Šoptrajanov B, Stefov V. Determination of the composition of human urinary stones composed of whewellite, weddellite and carbonate apatite using artificial neural networks Analytica Chimica Acta 491-2003; 211–218

[18] Monnier L, Piperno M, Colette C, J.-L. Schlienger: Applying nutritional and metabolic perspectives to the treatment of nephrolithiasis. Médecine des maladies Métaboliques - Septembre 2013 - Vol. 7 - N°4 doi:10.1016/S1957-2557(13)70593-5

[19] Altman, N S. An introduction to kernel and nearest-neighbor nonparametric regression. The American Statistician 46 (3): 1992; 175–185. doi:10.1080/00031305.1992.10475879.

[20] Nasibov E, Kandemir-Cavas C. Efficiency analysis of KNN and minimum distance-based classifiers in enzyme family prediction Computational Biology and Chemistry 33- 2009; 461–464

[21] Chen T, Zhang X, Jin S, Kimb O Efficient classification using parallel and scalable compressed model andits application on intrusion detection Expert Systems with Applications 41–2014; 5972–5983

[22] Songbo T.An effective refinement strategy for KNN text classifier Expert Systems with Applications 30 – 2006; 290–298

[23] Pedronette D C G, Penatti O A B, Torres R S. Unsupervised manifold learning using Reciprocal kNN Graphs in image re-ranking and rank aggregation tasks Image and Vision Computing 32– 2014; 120–130

[24] Yu C, Cui B, Wang S, Su J. Efficient index-based KNN join processing for high-dimensional data. Information and Software Technology 49– 2007: 332–344

[25] F. Crone S, Lessmann S, Stahlbock R. The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing.European Joul of Operational Research 173– 2006; 781–800

[26] Wu X, Kumar V, Quinlan J R, Ghosh J, Yang Q, Motoda H, J. McLachlan G, Ng A, Liu B, S. Yu P, Zhou Z, Steinbach M, J. Hand D, Steinberg D. Top 10 algorithms in data mining Knowl Inf Syst 2008; 14:1–37

[27] Vapnik V. Statistical learning theory. Wiley, New Tork 1998.

[28] Crammer K, Singer Y. On the Algorithmic Implementation of Multi-class SVMs, JMLR, 2001

[29] J. Zaki M, Meira Jr W. Fundamentals of Data Mining Algorithms. Cambridge University Press, 2010.

[30] Tsochantaridis I, Hofmann T, Joachims T, Altun Y. Support Vector Learning for Interdependent and Structured Output Spaces, ICML, 2004

[31] Cissé Y, Kinouchi Y, Nagashino H, Akutagawa M. Identification of homeostatic dynamics for a circadian signal source using BP neural networks ITBM-RBM 2000; 21: 24 32

[32] Chow TWS, Fei G, Cho SY. High order cumulants-based least squares for nonminimum-phase system identification. IEEE Trans Trans Signal Proc 1997; 45: 104.5-50.

[33] Faraway J, Chatfield C. Time series forecasting with neural network, a comparative study using the airline data. Appl Stat J R Stat Sot Ser 1998; 47: 231-50

[34] Liu Q, Huang T. A neural network with a single recurrent unit for associative memories based on linear optimization. Neurocomputing 118– 2013; 263–267

[35] Baxt W G. Use of an artificial neural network for data analysis in clinical decision making: the diagnosis of acute coronary occlusion, neural computation 1990; 2:480-489.

[36] Baxt W G. Analysis of the clinical variables driving decision in an artificial neural network trained to identify the presence of myocardial infarction. Volume 21, Issue 12, December 1992; Pages 1439–1444

[37] Patil S, Henry J W, Rubenfvie M, Stein P D. Neural Network in the clinical diagnosis of acute pulmonary embolism, Chest1993; 104: 1685-1689.

[38] Wu Y, Wu Y, Wang J,Yan Z, Qu L, Xiang B,Zhang Y. An optimal tumor marker group coupled artificial neural network for diagnosis of lung cancer. Expert Systems with Applications 38– 2011; 11329–11334.

[39] Armoni A, Use of neural networks in medical diagnosis, Medical Diagnosis Computing 1998; 15: 100-104

[40] Han M, Snow P B, Brandt J M, Partin A W.Evaluation of artificial neural networks for the prediction of pathologic stage in prostate carcinoma, Cancer 91(S8): 2001; 1661-1666.

[41] HAMILTON D, O’MAHONY D, COFFEY J, MURPHY J, O’HARE N, FREYNE P, WALSH B, COAKLEY D. Classification of mild A lzheimer’s disease by artificial neural network analysis of SPET data Nuclear Medicine Communications 1997; 1 8 : 805-810.

[42] Pradhan C,Wuehr M, Akrami F, Neuhaeusser M, Huth S, Brandt T, Jahn K, Schniepp R. Automated classification of neurological disorders of gait using spatio-temporal gait parameters. Journal of Electromyography and Kinesiology 25–2015; 413–422
[43] Kalman B L, Kwasny S C. Proceeding at the International Joint Conference on Neural Networks, vol. 1, Publisher: IEEE Neural Networks Council, Baltimore,MD, June 1992, pp. 800–805.
[44] Pierna J A F, Lecler B, Conzen J P, Niemoeller A, Baeten V, Dardenne P. Comparison of various chemometric approaches for large near infrared spectroscopic data of feed and feed products: Analytica Chimica Acta 705– 2011; 30– 34.
[45] Baxt W G. Application of artificial neural networks to clinical medicine, The Lancet 346: 1995; 1135-1138
[46] Davalo E, Naïm P, Des réseaux de neurones, Eyrolles, Paris 1990.
[47] Bourret P, Reggia J, Samuelides M. Réseaux de neurones artificiels : une approche connexionniste de l'Intelligence Artificielle, Teknea. 1991
[48] Touzet C. Les réseaux de neurones artificiels. Introduction au connexionisme, 1992 ; EC2 Ed, Nanterre.
[49] ABDI H. Les réseaux de neurones, Presses Universitaires de Grenoble, 1994 ; Grenoble.
[50] Jodouin J.F. Les réseaux de neurones, principes et définitions, Hermès, Paris. 1994
[51] Gross G W. Boone J M, Bishop D M. Pediatric skeletal age: determination with neural networks, Radiology 1995; 689-695.
[52] Penny W, Frost D. Neural networks in clinical medicine, Medical Decision Making: 1996; 16: 386-398.
[53] Schmitt A, Le Blanc B, Corsini M.-M, Lafond C, Bruzek J. Les réseaux de neurones artificiels. Bulletins et mémoires de la Société d'Anthropologie de Paris, 2001 ; 13 (1 2)
[54] Zupan J, Gasteiger J. Neural Networks in Chemistry and Drug Design, Wiley, New York, 1999.
[55] Bos A, Bos M, Van der Linden W E. Anal. Chem. Acta 1992- 256- 133.