Rep Pract Oncol Radiother, 2007; 12(1): 9-17

Original Paper

Received: 2006.07.24
Accepted: 2006.12.06
Published: 2007.03.09

Clinical data analysis using artificial neural networks (ANN) and principal component analysis (PCA) of patients with breast cancer after mastectomy

Adam Buciński[1ACDEF], **Tomasz Bączek**[2,3DEF], **Jerzy Krysiński**[4ADG], **Renata Szoszkiewicz**[5BD], **Jerzy Załuski**[5BD]

[1] Department of Biopharmacy, Faculty of Pharmacy, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
[2] Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gdańsk, Poland
[3] Tri-City Central Animal Labolatory – Research and Service Centre, Medical University of Gdańsk, Gdańsk, Poland
[4] Department of Marketing and Pharmaceutical Law, Faculty of Pharmacy, Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
[5] Great Poland Cancer Centre, Poznań, Poland

Authors' Contribution:
A Study Design
B Data Collection
C Statistical Analysis
D Data Interpretation
E Manuscript Preparation
F Literature Search
G Funds Collection

Source of support: The work was supported by Polish State Committee for Scientific Research Project No. 3 T11E 042 26 for years 2004–2007.

Summary

Background Exploitation of the several types of information on patient, disease and treatment variables ranging from sociological to genetic ones by means of chemometric analysis was considered and evaluated.

Aim Performance of modern data processing methods, namely principal component analysis (PCA) and artificial neural network (ANN) analysis, is demonstrated for predictions of the recurrence of breast cancer in patients treated previously with mastectomy.

Materials/Methods The data on 718 patients were retrospectively evaluated. 11 subject and treatment variables were determined for each patient. A matrix of 718×11 data points was subjected to PCA and ANN processing. The properly trained ANN was used to predict the patients with recurrence and without recurrence within a 10-year period after mastectomy.

Results It was found that the prognostic potency of the trained and validated ANN was reasonably high. Additionally, using the principal component analysis (PCA) method two principal components, PC1 and PC2, were extracted from the input data. They accounted cumulatively for 37.5% of the variance of the data analyzed. An apparent clustering of the variables and patients was observed – these have been interpreted in terms of their similarities and dissimilarities.

Conclusions It has been concluded that ANN analysis offers a promising implementation to established methods of statistical analysis of multivariable data on cancer patients. On the other hand, PCA has been recommended as an alternative to classical regression analysis of multivariable clinical data. By means of ANN and PCA practically useful systematic information may be extracted from large sets of data, which can be of value for prognosis and appropriate adjustment of the treatment of breast cancer.

Key words **breast cancer • mastectomy • recurrence model • artificial neural networks (ANN) • principal component analysis (PCA)**

Full-text PDF: http:/www.rpor.pl/pdf.php?MAN=10076

Word count: 2358
Tables: 6
Figures: 4
References: 47

Author's address: Adam Buciński, Department of Biopharmacy, Faculty of Pharmacy, Nicolaus Copernicus University, Collegium Medicum, Skłodowskiej-Curie 9 Str., 85-094 Bydgoszcz, Poland, e-mail: kizbiofarmacji@cm.umk.pl

BACKGROUND

Breast cancer has become the commonest malignant disease causing the death of women in the European Community [1–3]. Increasing incidence in all Western countries is observed. This includes Poland, where approximately 11 thousand cases of the disease are reported annually [4,5]. Breast cancer still confronts us with many unsolved problems and open questions about optimal treatment, prediction of recurrence and potential benefits of supplementary treatment.

Treatment strategy in breast cancer depends on the stage of development. Patients in an early stage of the disease are treated by primary surgery and/or radiotherapy. When the disease becomes more advanced or metastatic treatment starts with primary chemotherapy and/or radiotherapy and/or hormonotherapy. Further decisions concerning therapeutic strategy depend on the prediction of possible recurrence and response to therapy. However, reliable prediction is extremely difficult because of the lack of a single prognostic parameter or identified combinations of them [3,6].

Several factors have been recommended to help in the prognosis of both overall survival and recurrence-free interval in patients with breast tumours. However, considered separately those factors usually appear disputable. The best-known example is the patient's age. It has been demonstrated eventually (in statistical terms) that age under 35 is an independent prognostic factor for an unfavourable recurrence [7–9]. On the other hand, age over 50 years (post-menopausal women) was shown in several studies to provide a better prognosis [10]. It could not be recommended, however, to correlate linearly survival with age or to predict the recurrence and response to therapy of an individual patient based on age alone.

Among major risk factors of breast cancer, besides age, are often included: family history of breast cancer, menopause in old age, long supplementary hormone therapy, indolent proliterative breast disease, and carrying genes BRCA1 and BRCA2. Histolic grade, Nottingham Prognostic Index, lymph node stage, age, oestrogen receptor (ER) and progesterone receptor (PgR) status, and tumour size have previously been shown to be important prognostic indicators for distant recurrence of breast cancer [11,12].

In various types of cancer there have been proposed prognostic indices which are derived by multiple regression analysis of number of patients, diseases and treatment parameters [13,14]. For instance, when used simultaneously in a seven-variable regression equation, the levels of bicarbonate, alanine transaminase, alkaline phosphatase, sodium, potassium, urea and uric acid, together with erythrocyte sedimentation rate and patient age, appeared significant as prognostic factors for survival in small-cell lung cancer [13].

The fundamental problem with multiple regression analysis is that the parameters (independent variables) considered simultaneously cannot be mutually related, i.e. they should be orthogonal [15]. It is difficult to find a representative (and sufficiently large for statistical reasons) set of treatment parameters which would be orthogonal. Therefore, prognostic indices derived by means of multiple regression analysis, having otherwise advantages too (e.g. interpretability of the coefficients), can sometimes be of rather limited reliability if there are a lot of interactions among the data to be considered.

Artificial neural network (ANN) analysis is a new method of data analysis, which is to emulate the brain's way of working. Neural nets exhibit the way in which arrays of neurons possess function in biological learning and memory recognition. ANNs differ from classical computer programs because they "learn" from a set of examples rather than being programmed to get the right answer. The information is encoded in the strength of the network's "synaptic" connections [16,17]. In chemistry and related fields of research, interest in neural-network computing has been noted since 1986 [18,19]. ANNs have found application in compound classification, modelling of structure-activity relationships, identification of potential drug targets and the localization of structural and functional features of biopolymers [20–29]. ANNs are a sophisticated tool for exploration of complex data sets. Because of their ability to mimic a number of relationships, they are used to process clinical data, too. Confirmation of that fact can be found in the literature, including ANNs' usefulness in oncology and studies on breast cancer [29–33].

The general idea of principal component analysis (PCA) is to reduce the dimensionality of the original multivariable data set by a finding linear combination of those variables that explains most of the variability within the set of data considered. By means of PCA, systematic information initially dispersed over a large matrix of input variables (often intercorrelated) is extracted and condensed in a few abstract variables. Usually a few principal components (PCs) are used to determine the abstract variable projection on the plane or in three-dimensional space [34]. PCA allows the principal components to be found given either the original variables or a correlation or covariance matrix. Coefficients of each principal component are determined by computing the eigenvalues of the covariance matrix or the correlation matrix. Standard commercially available computer programs provide the eigenvalues, the component weight and the values of individual principal components. Especially convenient are plots of component weights for the first two PCs for each variable and the scatterplots for the first two PCs illustrating distribution of objects [18]. Projections of data points ascribed to individual objects (patients) and to individual variables reflect mutual similarities and dissimilarities among

Table 1. Variables considered in the analysis by artificial neural network (ANN).

Variable No.	Variable Name
1	Age (in years) (1) <30 (2) 31–50 (3) 51–60 (4) >60
2	Period of hormonal activity (in years) (1) <10 (2) 11–20 (3) 21–30 (4) 31–40 (5) >40
3	Number of childbirths (1) 0 (2) 1 (3) 2 (4) 3 (5) >4
4	Size of tumour (1) no data available (2) <40 (3) >40
5	Involvement of auxiliary lymph nodes (1) no involvement (2) 1–3 (3) 4–8 (4) >8
6	Emboli from carcinoma cells in the vessels (1) yes (2) no
7	Degree of malignancy according to Bloom (1) I (2) II (3) III
8	The type of hormonal adjuvant therapy (1) no hormonal adjuvant therapy (2) tamoxifen
9	Familial incidence of cancer (1) yes (2) no
10	Professional activity (1) intellectual (2) physical (3) no profession
11	Residence place (1) Poznań (2) outside Poznań
Category	Category (1) recurrence (2) no recurrence

them. That way, the basic part of information on the objects and on the variables can be exploited by our mind, which naturally visualizes only relationships in up to three dimensional space.

The work was aimed at application of ANNs as convenient and reliable prognostics in breast cancer. Our former publications demonstrated preliminarily the use of PCA [35] and ANN [36] for 228 breast cancer patients after mastectomy. In this project the approach has been tested and revisited on the material available for 718 breast cancer patients after mastectomy that were treated and observed in the Chemotherapy Unit, Oncology Centre of Wielkopolska (Greatpoland Cancer Centre) in Poznań, Poland. Mastectomy was done in 1992/1994 in the Surgery Unit of the same institution. Patient observation was carried on until 2002/2004.

Aim

The purpose of the work is to demonstrate that ANN and PCA are convenient and reliable prognostic tools in prediction of recurrence incidences within a 10-year period after mastectomy in the case of breast cancer. By these methods one can exploit all the types of information on patient, disease and treatment, making use in a single analysis of variables ranging from sociological to genetic ones.

Materials and Methods

Patient characteristics

Data on 718 patients with breast cancer were retrospectively collected and analyzed. The variables considered in this study are presented in Table 1. A total number of 11 variables were subjected to ANN and PCA analyses. The final matrix of data subjected to ANN and PCA analyses was 718 patients times 11 variables. Table 2 presents the data for 6 selected patients with and without recurrence of breast cancer within a 10-year period after mastectomy.

ANN analysis

Artificial neural networks (ANN) were run on a personal computer using Statistica Neural Networks v. 6 software (StatSoft, Tulsa, OK, USA). In the first phase of the analysis, the number of hidden neurons was determined experimentally. For the choice of the appropriate network model, the coefficient value of the correct classification with respect to validation of the set of data was taken into account. No distinct differences in classification coefficient were observed for networks differing in the number of hidden neurons. The results are presented in Table 3. The size of individual sets of data was defined on the basis of the assumed division of the sample according to the scheme as follows: 358:180:180. That means that 50% of cases were assigned to the learning set, 25% of cases were in the validation set, and 25% of cases were in the testing set. Individual patients were classified randomly to each set. The evaluation of the influence of the decrease of learning set on the predictive properties was also evaluated (Table 4).

For further investigation, an artificial neural network based on a multilayer perceptron consisting of 11 artificial neurons in the input layer, 9 in the hidden layer and 1 neuron in the output layer was used. The architecture of the model used is depicted in Figure 1. The method of supervised learning with back-propagation strategy and conjugate gradient descent method was used. Variables for patients analyzed were divided into three sets: learning set with 358 patients, validating set with 180 patients and testing set with 180 patients. The learning process was completed when the artificial neural network was characterized by the least RMS error. In the case of the network applied, learning was completed in 50 epochs by back propagation (BP) method and 2 epochs by conjugate gradient descent (CGD) method. Data from the learning set were presented in a randomized manner during the learning process.

PCA analysis

Principal component analysis of the 718×11 data matrix was performed by means of the Statistica *v.* 6 computer program (StatSoft Inc, Tulsa, OK, USA) operated on a personal computer. It was found that the two first principal components, PC1 and PC2, accounted cumulatively for 37.5% of the total variance in the data described with 11 original variables.

The projection of points assigned to individual patients (principal component "scores") in the space determined by the first two principal components axes, PC1 and PC2, is depicted in Figure 2.

The variables positioned in the space determined by the first principal components produced a plot of principal component "loadings". The principal components most significant for separation of patients ("object scores") in Figure 2 are PC1 and

Table 2. Variables considered in analysis by artificial neural network (ANN) and their values for six example patients.

Variable No.	Variable name	Variable value for patient					
		No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
1	Age (in years)	2	3	2	2	3	2
2	Period of hormonal activity (in years)	3	4	3	4	4	4
3	Number of childbirths	3	1	1	3	2	1
4	Size of tumour	3	3	3	2	2	2
5	Involvement of auxiliary lymph nodes	4	3	3	2	1	1
6	Emboli from carcinoma cells in the vessels	2	2	1	2	2	2
7	Degree of malignancy according to Bloom	3	3	3	1	1	1
8	The type of hormonal adjuvant therapy	1	1	1	2	2	1
9	Familial incidences of cancer	2	1	2	2	2	1
10	Professional activity	2	2	1	2	1	1
11	Residence place	1	2	2	1	1	2
	Category	R	R	R	NR	NR	NR

Recurrence (R); No recurrence (NR).

Table 3. Classification results for six ANN models differing in the number of hidden neurons.

No.	Typ ANN	CR* learning set	CR validation set	CR testing set
1	MLP 11: 11-9-1: 1	0.757	0.767	0.767
2	MLP 11: 11-4-5-1: 1	0.737	0.761	0.761
3	MLP 11: 11-14-14-1: 1	0.737	0.761	0.794
4	MLP 11: 11-9-9-1: 1	0.735	0.761	0.772
5	MLP 11: 11-14-1: 1	0.743	0.756	0.778
6	MLP 11: 11-5-1: 1	0.743	0.750	0.767
7	MLP 11: 11-22-22-1: 1	0.740	0.733	0.778

* CR – Classification rate.

Table 4. Results for ANN with respect to different size of each set of data.

No.	Size of individual sets of data (learning: validation: testing)	ANN model type	CR learning set	CR validation set	CR testing set
1	478: 120: 120	MLP 11: 11-12-1: 1	0.799	0.717	0.625
2		MLP 11: 11-11-1: 1	0.791	0.725	0.658
3		MLP 11: 11-13-1: 1	0.795	0.708	0.633
4	538: 90: 90	MLP 11: 11-13-1: 1	0.818	0.767	0.678
5		MLP 11: 11-13-1: 1	0.801	0.667	0.689
6		MLP 11: 11-11-1: 1	0.810	0.722	0.689
7	538: 150: 30	MLP 11: 11-13-1: 1	0.803	0.700	0.500
8		MLP 11: 11-11-1: 1	0.812	0.740	0.600
9		MLP 11: 11-8-1: 1	0.783	0.720	0.533

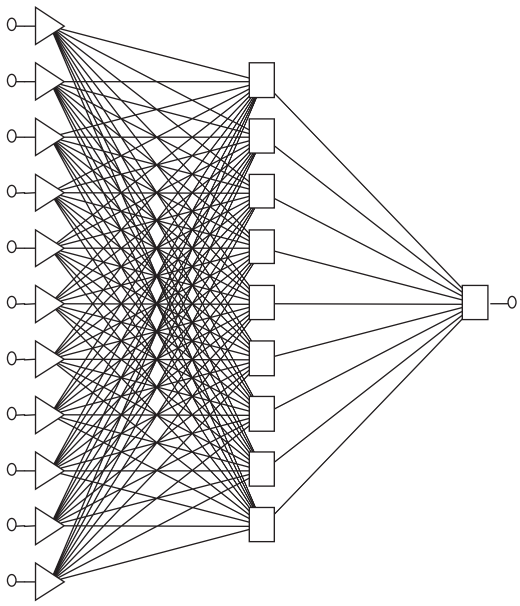

Figure 1. Architecture of artificial neural network used in predictions of recurrence of breast cancer within 10 years after mastectomy.

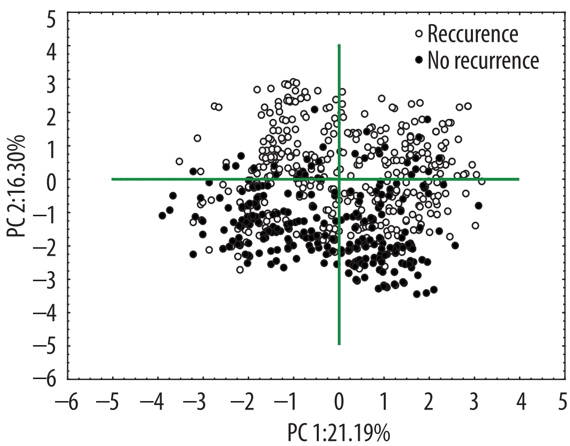

Figure 2. Projection of 718 points denoting patients (described by 11 variables listed in Table 1) in the space of the first two principal components, PC1 and PC2, from principal component analysis (PCA) of a 718×11 data matrix.

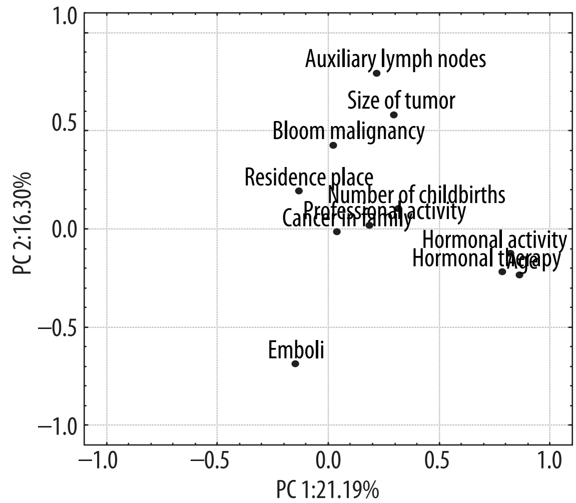

Figure 3. Projection of 11 variables from Table 1 on the plane of PC1 and PC2, from principal component analysis (PCA) of data for 781 patients.

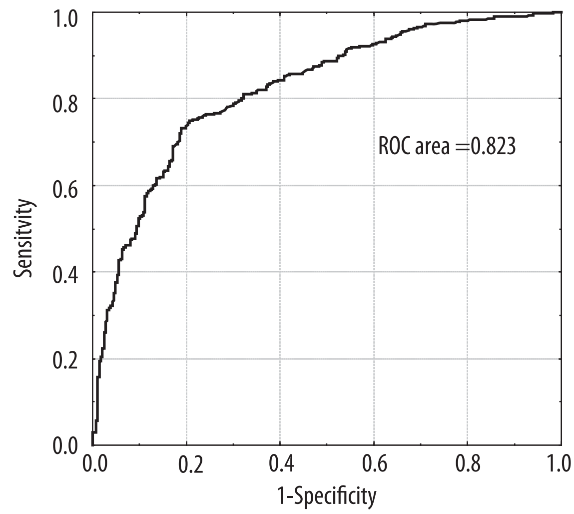

Figure 4. Receiver operating characteristic curves (ROC curves) for training set.

PC2. Therefore in Figure 3 the "loadings" of PC1 and PC2 by individual variables are presented.

Results

Figure 1 presents the architecture of the ANN model used for predictions of breast cancer recurrence within a 10-year period after mastectomy based on the input data from training, validating and testing data sets, respectively.

In Table 5 classification statistics are collected for training, validating and testing sets. Receiver operating characteristic curves (ROC curves) are shown in Figure 4.

Using the proposed method it was possible to differentiate patients in the testing group as with or without breast cancer recurrence with prognostic potency at the level of 75%. The prognostic potency of ANN regarding the set of test patients is reasonably

Table 5. Classification statistics for ANN used.

	Learning set		Validation set		Testing set	
	No recurrence	**Recurrence**	**No recurrence**	**Recurrence**	**No recurrence**	**Recurrence**
Total	153	205	67	113	66	114
Correct	108	157	58	82	52	90
Wrong	45	48	9	31	14	24

Table 6. Sensitivity analysis results for the variables considered in artificial neural network (ANN) analysis.

Variable No.	Variable name	Rank	Error
1	Age (in years)	4	0.4313
2	Period of hormonal activity (in years)	3	0.4322
3	Number of childbirths	11	0.4267
4	Size of tumour	2	0.4467
5	Involvement of auxiliary lymph nodes	6	0.4299
6	Emboli from carcinoma cells in the vessels	7	0.4287
7	Degree of malignancy according to Bloom	1	0.4484
8	The type of hormonal adjuvant therapy	9	0.4274
9	Familial incidences of cancer	10	0.4268
10	Professional activity	5	0.4310
11	Residence place	8	0.4280

good and proves the appropriate choice and shape of the network proposed. 142 patients in the testing set of a total number of 180 have been correctly classified, which means that one is able to predict the breast cancer recurrence utilizing the variables used with high probability reaching 79%.

Parallel with ANN statistics, also ANN sensitivity analysis for input variables was performed (Table 6). Sensitivity analysis provides insight into the usefulness of individual variables. With this kind of analysis it is possible to judge what parameters are the most significant (with sensitivity value close to 1) and the least significant (with sensitivity value close to 10) during generation of the satisfactory ANN. According to sensitivity analysis the degree of malignancy according to Bloom is the most significant parameter to distinguish patients with and without breast cancer recurrence. Highly significant are also the size of tumour, period of hormonal activity and age. Less significant appear to be number of childbirths, familiar accidences of cancer and the type of hormonal adjuvant therapy.

Principal component analysis also extracts systematic information on the variables considered. The most interesting are variables directly connected with the negative prognosis of recurrence 10 years after the mastectomy. Almost all variables considered are directly associated with recurrence process in this kind of analysis, except emboli from carcinoma cells in the vessels and to a lesser extent the type of hormonal adjuvant therapy, period of hormonal activity and age. These variables are located in the lower side of Figure 3.

Lower values of PC2 for patients (closed circles in Figure 2) suggest no recurrence among these patients. On the basis of PCA analysis it could be assumed that the involvement of auxiliary lymph nodes, size of tumour and degree of malignancy according to Bloom generate in an especially significant manner the recurrence of breast cancer of the patients 10 years after mastectomy (Figure 3). This is confirmed also by the results obtained in sensitivity analysis of the ANN method. According to that analysis, size of tumour and degree of malignancy according to Bloom considering the possibility of breast cancer recurrence of patients 10 years after mastectomy should be taken into special account. However, the involvement of auxiliary lymph nodes was treated in this case as less significant.

DISCUSSION

The prognostic factors identified by both types of analysis (ANN and PCA) as Bloom grade and tumour size are well known in breast cancer prediction [38–44]. Another analyzed histological grade – involvement of lymph nodes – is usually identified as a significant prognostic factor in different types of multivariate and univariate analysis and it is always highly correlated with the other two histological components [39]. The undertaken ANN analysis did not recognize its strong prognostic power (rank 6 of sensitivity). One reason could be the character of the database, where almost half of the group was negative lymph nodes. Regarding the representativeness of the analyzed matrix, the advantageous utilization of just two of them (Bloom grade and tumour size) can be concluded.

Among the most significant predictive variables, menopausal status (ANN sensitivity rank 3) and age (ANN sensitivity rank 4) were identified, which are often recognized as sensitive factors [40]. The PCA analysis revealed that these two factors represent a different direction of variability (PCA axis 1) than Bloom grade and tumour size (PCA axis 2). Combining variables representing both revealed directions of variability would benefit in sensitiveness of the proposed combination of risk factors.

CONCLUSIONS

The large potential of the neural network has been proved based on clinical studies in breast cancer and several comparisons with other analytical techniques have been undertaken [32,45–47]. The presented study shows the valuable application of ANN and PCA to breast cancer recurrence prediction due to the valuable original data set analysis based on uniform long-term records. The undertaken analysis is mostly explanatory in character, which may help to identify a combination of factors providing an effective treatment and a good prognosis, the more so that ANN and PCA analyses allow for testing a practically unlimited number of either mutually related or apparently unrelated factors and cases.

REFERENCES

1. McGuire WL: Adjuvant therapy of node-negative breast cancer. N Engl J Med, 1989; 320: 525–7
2. O'Higgins N: Aspects of breast cancers. Chirurgie. 1992; 118: 324–7
3. von Kleist S: Prognostic factors in breast cancer: theoretical and clinical aspects. Anticancer Res, 1996; 16: 3907–12
4. Zatoński WA, Didkowska J: Epidemiology of malignant cancers. in: Clinical oncology (Ed. Krzakowski M), Borgis, Warsaw; 2001
5. Landis SH, Murray T, Bolden S: Cancer statistics. Ca J Clin. 1999; 49: 8–31
6. Dhingra K, Hortobagyi GN: Critical evaluation of prognostic factors. Semin Oncol, 1996; 23: 436–45
7. Mariani L, Coradini D, Biganzoli E, Boracchi P: Prognostic factors for metachronus contralateral breast cancer: A comparison of the linear Cox regression model its artificial neural network extension. Breast Ca Res Treat, 1997; 44: 167–78
8. Bonnier P, Romain S, Charpin C et al: Age as a prognostic factor in breast cancer: relationship to pathologic and biological features. Int J Ca, 1995; 62: 138–44
9. Chung M, Chang HR, Bland KI, Wanebo HJ: Younger women with breast carcinoma has a poorer prognosis than older women. Cancer, 1996; 77: 97–103
10. Aaltoma S, Lipponen PK, Kosma VM: Prognostic factors after 5 years follow-up in female breast cancer. Oncology, 1992; 49: 93–8
11. Elston CW, Ellis IO: Pathological prognostic factors in breast cancer: 1.The value of histological grade in breast cancer-Experience from a large study with long term follow-up. Histopathology, 1991; 19: 403–10
12. Galea MH, Blamey RW, Elston CW: The Nottingham Prognostic Index in primary breast cancer. Breast Ca Res Treat, 1992; 22: 207–19
13. Beyer J, Kramar A, Mandanas R et al: High-dose chemotherapy as salvage treatment in germ cell tumours: a multivariate analysis of prognostic variables. J Clin Oncol, 1996; 14: 2638–45
14. Maestu I, Pastor M, Gomez-Codina J et al: Pretreatment prognostic factors for survival in small-cell lung cancer: a new prognostic index and validation of three known prognostic indices on 34 patients. Ann Oncol, 1997; 8: 547–53
15. Jollife IT: Principal Component Analysis, Springer, New York, 1986
16. Zupan J, Gasteiger J: Neural Networks for Chemists. An Introduction, VCH, Weinheim, 1993
17. Zupan J, Gasteiger J: Neural networks: A new method for solving chemical problems or just a passing phase? Anal Chim Acta, 1991; 248: 1–30
18. Kaliszan R: Structure and Retention in Chromatography, A Chemometric Approach, Harwood Academic, Amsterdam, 1997

19. Schneider G, Wrede P: Artificial networks for computer-based molecular design. Prog Biophys Mol Biol, 1998; 70: 175–222

20. Isu Y, Nagashima U, Hosoya H, Aoyma T: Development of neural network simulator for structure-activity correlation of molecules. J Chem Software, 1994; 2: 76–95

21. Andrea TA, Kalayeh H: Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J Med Chem, 1991; 34: 2824–36

22. So SS, Richards WG: Application of neural networks. Quantitative structure-activity relationships of the derivatives of 2,4-diamino-5-(substituted-benzyl) pyrimidines as DHFR inhibitors. J Med Chem, 1992; 35: 3201–7

23. Ajay A: Unified framework for using neural networks to build QSARs. J Med Chem, 1993; 36: 3565–71

24. Brickley MR, Shepherd JP, Armstrong RA: Neural networks: a new technique for development of decision support systems in dentistry. J Dent, 1998; 26: 305–9

25. Snow PB, Rodvold DM, Brandt JM: Artificial neural networks in clinical urology. Urology, 1999; 54: 787–90

26. Wei JT, Tewari A: Artificial neural networks in urology: pro. Urology, 1999; 54: 945–8

27. Krongrad A, Lai S: Artificial neural networks in urology: con. Urology, 1999; 54: 949–51

28. Lee CW, Park JA: Assessment of HIV/AIDS-related health performance using an artificial neural network. Inform Manage, 2001; 38: 231–8

29. Mian S, Ball G, Hornbuckle J et al: A prototype methodology combining surface-enhanced laser desorption/ionization protein chip technology and artificial neural network algorithms to predict the chemoresponsiveness of breast cancer cell lines exposed to Paclitaxel and Doxorubicin under *in vitro* conditions. Proteomics, 2003; 3: 1725–37

30. Biganzoli E, Boracchi P, Coradini D et al: Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors. Ann Oncol, 2003; 14: 1484–93

31. Lisboa PJ, Wong H, Harris P, Swindell R: A Bayesian neural network approach for modelling censored data with an application to prognosis after surgery for breast cancer. Artif Intell Med, 2003; 28: 1–25

32. Jerez-Aragones JM, Gomez-Ruiz JA, Ramos-Jimenez G et al: A combined neural network and decision trees model for prognosis of breast cancer relapse. Artif Intell Med, 2003; 27: 45–63

33. Abbass HA: An evolutionary artificial neural networks approach for breast cancer diagnosis. Artif Intell Med, 2002; 25: 265–81

34. Kaliszan R: Quantitative Structure-Chromatographic Retention Relationships, Wiley, New York, 1987

35. Buciński A, Załuski J, Krysiński J, Kaliszan R: A principal component analysis of patients, disease and treatment variables: a new prognostic tool in breast cancer after mastectomy. Rep Pract Oncol Radiother, 2000; 5: 83–89

36. Buciński A, Bączek T, Kaliszan R et al: Artificial neural network analysis of patient and treatment variables as a new prognostic tool in breast cancer after mastectomy. Adv Clin Exp Med, 2005; 14: 973–9

37. *www.statsoft.com*

38. Pinder SE, Murray S, Ellis IO et al: The importance of the histologic grade of invasive breast carcinoma and response to chemotherapy. Cancer, 1998; 83: 1529–39

39. Genestie C, Zafrani B, Asselain B et al: Comparison of the prognostic value of Scarff-Bloom-Richardson and Nottingham histological grades in a series of 825 cases of breast cancer: Major importance of the mitotic count as a component of both grading systems. Anticancer Res, 1998; 18: 571–6

40. Le Goff JM, Lavayssiere L, Rouesse J, Spyratos F: Nonlinear discriminant analysis and prognostic factor classification in node-negative primary breast cancer using probabilistic neural networks. Anticancer Res, 2000; 20: 2213–8

41. Hery M, Delozier T, Ramaioli A: Natural history of node-negative breast cancer: are conventional prognostic factors predictors of time to relapse? Breast, 2002; 11: 442–8

42. Robles-Frias A, Gonzalez-Campora R, Martinez-Parra D et al: A Robinson cytologic grading of invasive ductal breast carcinoma – correlation with histologic grading and regional lymph node metastasis. Acta Cytolog, 2005; 49: 149–53

43. Dalton LW, Pinder SE, Elston CE et al: Histologic grading of breast cancer: Linkage of patient recurrence with level of pathologist agreement. Modern Pathol, 2000; 13: 730–5

44. Balslev I, Axelsson CK, Zedeler K et al: The Nottingham-Prognostic-Index applied to 9,149 patients from the studies of the Danish-Breast-Cancer-Cooperative-Group (DBCG). Breast Ca Res Treat, 1994; 32: 281–90

45. Floyd CE, Lo JY, Yun AJ et al: Prediction of breast-cancer malignancy using an artificial neural-network. Cancer, 1994; 74: 2944–48

46. Retsky M, Demicheli R, Hrushesky W et al: Recent translational research: computational studies of breast cancer. Breast Ca Res, 2005; 7: 37–40

47. Marian L, Coradini D, Biganzoli E et al: Prognostic factors for metachronous contralateral breast cancer: A comparison of the linear Cox regression model and its artificial neural network extension. Breast Ca Res Treat, 1997; 44: 167–78