

ADAM BUCIŃSKI^{1, 2}, TOMASZ BĄCZEK³, ROMAN KALISZAN³, ANTONI NASAL³,
JERZY KRYSIŃSKI⁴, JERZY ZAŁUSKI⁵

Artificial Neural Network Analysis of Patient and Treatment Variables as a Prognostic Tool in Breast Cancer after Mastectomy

Analiza za pomocą sztucznych sieci neuronowych danych dotyczących pacjentek oraz podjętej terapii jako metoda prognostyczna w raku sutka po mastektomii

¹ Nicolaus Copernicus University, Collegium Medicum in Bydgoszcz, Department of Biopharmacy, Poland

² Institute of Animal Reproduction and Food Research of Polish Academy of Sciences, Division of Food Science, Olsztyn, Poland

³ Medical University of Gdansk, Department of Biopharmaceutics and Pharmacodynamics, Poland

⁴ Nicolaus Copernicus University, Collegium Medicum in Bydgoszcz, Department of Pharmacoeconomics and Pharmaceutical Marketing, Poland

⁵ Greatpoland Cancer Center, Poznan, Poland

Abstract

Material and Methods. Satisfactory performance of a modern data processing method, namely the artificial neural network (ANN) analysis, in the prediction of the surgery treatment results in the case of breast cancer has been demonstrated. The data on 228 patients treated and observed in the same oncology unit were retrospectively evaluated. Sixteen subject and treatment variables were determined for each patient. A matrix of 228×16 data points was subjected to the ANN processing. The patients were randomly divided into three groups. A group of 168 patients formed the training set and 30 patients formed the validation set. The remaining 30 patients served as a testing set. The properly trained neural network was used to predict patients with recurrence and without recurrence within a 5-year period after mastectomy. For the patients' classification with and without recurrence artificial neural network of the type of multilayer perceptron was used.

Results. Classification coefficients for properly distributed patients for the training, validating and testing set of data were 85.7, 90 and 86.7%, respectively.

Conclusions. It was found that the prognostic potency of ANN regarding the testing set of patients was very high. It has been concluded that the ANN analysis offers a promising alternative to established methods of statistical analysis of multivariable data on cancer patients (*Adv Clin Exp Med* 2005, 14, 5, 973–979).

Key words: artificial neural networks, breast cancer, recurrence.

Streszczenie

Materiał i metody. Przedmiotem badań były retrospektywne dane opisujące pacjentki leczone operacyjnie z powodu raka sutka. Wszystkie pacjentki były leczone i obserwowane w tym samym ośrodku. Dla każdej z 228 pacjentek dysponowano zestawem 16 charakterystycznych parametrów. Pacjentki te zostały losowo podzielone na trzy zbiory. Zbiór uczący liczył 168 pacjentek oraz zbiory walidacyjny i testujący po 30 pacjentek. Do zaklasyfikowania pacjentek na te, u których wystąpiła znowa choroba nowotworowej i te, u których znowy nie zaobserwowano, wykorzystano sieć neuronową typu perceptronu wielowarstwowego.

Wyniki. Współczynniki poprawnych klasyfikacji dla pacjentek wchodzących w skład zbiorów: uczącego, walidacyjnego i testującego wynosiły odpowiednio 85,7, 90; 86,7%.

Wnioski. Badania wykazały, że wytrenowana sztuczna sieć neuronowa charakteryzowała się dużą zdolnością prognostyczną. Na tej podstawie można wyciągnąć wniosek, że analiza metodą sztucznych sieci neuronowych jest obiecującą alternatywą obecnie stosowanych statystycznych metod analizy wielowymiarowych danych opisujących pacjentki z chorobą nowotworową (*Adv Clin Exp Med* 2005, 14, 5, 973–979).

Słowa kluczowe: sztuczne sieci neuronowe, rak sutka, nawrót choroby.

Breast cancer has become the commonest malignant disease causing the death of women in the European Community and its increasing incidence in all Western countries is observed [1–3]. The therapeutic strategy depends on the prediction of recurrence and response to the therapy. However, a reliable prediction is extremely difficult because of the lack of a single prognostic parameter or identified combinations of them [3, 4].

Several factors have been recommended to help to prognosticate both overall survival and recurrence-free interval in patients with breast tumors. However considered separately those factors usually appear disputable. The best-known example has been the patient age. It has been demonstrated eventually (in statistical terms) that age under 35 was an independent prognostic factor for an unfavorable outcome [5, 6]. On the other hand, age over 50 years (post-menopausal women) was shown in several studies to provide a better prognosis [7]. It could not be recommended, however, to correlate linearly survival with age or to prognosticate of the recurrence and response to therapy of an individual patient based on age alone.

In various types of cancer the prognostic indexes have been proposed which are derived by the multiple regression analysis of a number of patient, disease and treatment parameters [8, 9]. For instance, when used simultaneously in a seven-variable regression equation significant as the prognostic factors for survival in small-cell lung cancer appeared the levels of bicarbonate, alanine transaminase, alkaline phosphatase, sodium, potassium, urea and uric acid, together with erythrocyte sedimentation rate and patient age [8].

The fundamental problem with multiple regression analysis is that the parameters (independent variables) considered simultaneously can not be mutually related, *i.e.*, they should be orthogonal [10]. It is difficult to find a representative (and sufficiently large for statistical reasons) set of treatment parameters which would be orthogonal. Therefore, prognostic indexes derived by means of multiple regression analysis are of a rather limited reliability.

The artificial neural network (ANN) analysis is a method of data analysis, which is to emulate the brain's way of working. Neural nets exhibit the way in which arrays of neurons probably function in biological learning and memory. ANNs differ from classical computer programs in that they "learn" from a set of examples rather than are programmed to get the right answer. The information is encoded in the strength of the network's "synaptic" connections [11, 12].

In chemistry and related fields of research interest in neural-network computing has been noted

since 1986 [13, 14]. ANNs found application in compound classification, modeling of structure-activity relationships, identification of potential drug targets and the localization of structural and functional features of biopolymers [15]. Also, the suitability of neural networks to describe the strong correlation between the carcinogenicity of polycyclic aromatic hydrocarbons and their structural parameters from ^{13}C -NMR was reported [16]. A good performance of ANN in predicting bioactivity classes based on physicochemical parameters of agents was demonstrated for dihydrofolate reductase inhibitors [17, 18]. Antitumor activity predicting potency of ANNs was also illustrated [11]. Those authors analyzed the patterns of activity against a panel of 60 malignant cell lines. Given six possible classes of drug mechanism, the ANNs missed the correct category for only 12 out of 141 agents studied (8.5 percent). ANNs were proposed as decision support systems in dentistry [19], in urology [20–22], and to assess HIV/AIDS-related health performance [23]. ANNs are a sophisticated tool for exploration of the complex data sets. Because of their ability to mimic a number of relationships, they are used to process the clinical data, too. Confirmation of that fact can be found in literature, including ANNs's usefulness in oncology and studies on breast cancer [24–28].

The presumed work was aimed at application of ANNs as convenient and reliable prognostics in breast cancer. Authors' former publication [29] demonstrated the use of Principal Component Analysis (PCA) for this purpose. By using ANNs one can exploit all the types of information on patient, disease and treatment making use in a single analysis of the variables ranging from sociological to genetic ones. In this project the approach has been tested on the material available for 228 breast cancer patients after mastectomy that were treated and observed in the Chemotherapy Unit, Oncology Center of Wielkopolska (Wielkopolskie Centrum Onkologii) in Poznan, Poland. Mastectomy was done in 1990/1991 in the Surgery Unit of the same institution. Patient observation was carried on till 1996/1997.

Material and Methods

Patients

Data on 228 patients with breast cancer were retrospectively collected and analyzed. The variables considered in this study are presented in Table 1. They were converted with the use of Minimax method. It enabled to scale those data into the 0-1 range [30]. A total number of 16 variables we-

re subjected to ANN analysis. The final data matrix subjected to ANN analysis was 228 patients \times 16 variables. Table 2 presents the data for six selected patients: with and without recurrence within a 5-year period after mastectomy.

ANN Analysis

Artificial neural networks (ANN) were run on a personal computer using Statistica Neural Networks software (StatSoft, Tulsa, OK, USA).

Artificial neural network based on multilayer perceptron consisting of 16 artificial neurons in input layer, 5 ones in hidden layer and 1 neuron in

output layer was used. The architecture of the model utilized is depicted in Figure 1. 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 168 patients, validating set with 30 patients and testing set with 30 patients. Learning of the ANN was executed with learning coefficient set at 0.1 and momentum set at 0.3. Data from the learning set were presented in randomized manner during the learning process. The changes in RMS error were recorded for the training and validating set during the learning process. For further considerations one takes the arti-

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

Tabela 1. Zmienne rozważane w analizie metodą sztucznych sieci neuronowych

Variable No. (Numer zmiennej)	Variable Name (Nazwa zmiennej)	Variable No. (Numer zmiennej)	Variable Name (Nazwa zmiennej)
1	Age (years) 1. < 30 2. 31–50 3. 51–60 4. > 60	10	Malignancy: Bloom's degree 1. I 2. II 3. III
2	Menopause 1. Before 2. During 3. After	11	Surgery 1. Halsted's mastectomy 2. Patey's mastectomy 3. Mastectomy
3	Hormonal activity (years) 1. < 10 2. 11–20 3. 21–30 4. 31–40 5. > 40	12	Adjuvant therapy 1. None 2. Radiotherapy 3. Chemotherapy 4. Hormonotherapy 5. Radiotherapy + Chemotherapy 6. Radiotherapy + Hormonotherapy 7. Chemotherapy + Hormonotherapy 8. Radiotherapy + Chemotherapy + + Hormonotherapy
4	Number of births 1. 0 2. 1 3. 2 4. 3 5. > 3	13	Adjuvant radiotherapy 1. None 2. Scar and lymph nodes 3. Peripheral lymph nodes
5	Breast cancer in I. and II. Line 1. Yes 2. No	14	Neoadjuvant chemotherapy 1. Yes 2. No
6	Tumor size 1. Data missing 2. < 40 mm 3. > 40 mm	15	Type of first line chemotherapy 1. None 2. CMF 3. With anthracyclines
7	Positive lymph nodes 1. No 2. 1–3 3. 4–8 4. > 8	16	Type of adjuvant hormonotherapy 1. None 2. Tamoxifen
8	Infiltration of node capsule 1. Yes 2. No	17	Category 1. Recurrence 2. No recurrence
9	Arrests in microvessels 1. Yes 2. No		

ficial neural network characterized by the smallest RMS error with regard to validating set of data. In the case of this network learning was completed in 50 epochs by Back Propagation (BP) method and 247 epochs by Conjugate Gradient Descent (CGD) method.

Results and Discussion

Figure 1 presents the architecture of the ANN model used for predictions of patients with and without recurrence within a 5-year period post mastectomy, based on the input data from the training, validating and testing data sets, respectively.

In Table 3 classification statistics are collected for the training, validating and testing sets. For cases in the training set, the model of the trained neural network classified correctly 144 patients out of the total of 168 (117 without recurrence and 51 with recurrence). It means that correctness of the classification was at the level of 85.7%. Similarly high classification ratio was observed also in the case of patients in the validating and testing sets. Here, classification correctness was 90% and 86.7%, respectively. The confirmation of the classification correctness with the use of the applied neural network model are high values of the areas

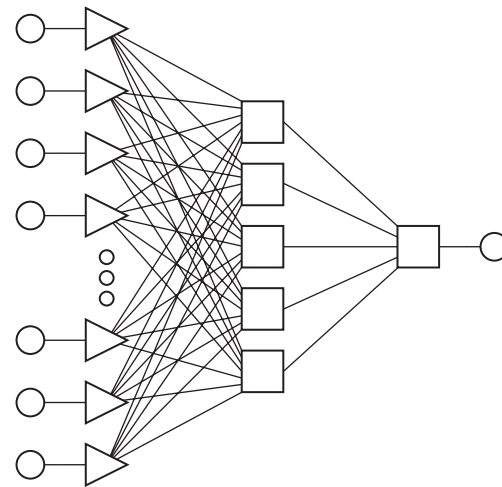


Fig. 1. Architecture of artificial neural network used for predictions of breast cancer recurrence within a 5-year period post mastectomy

Ryc. 1. Architektura sztucznej sieci neuronowej użytej do przewidywania wystąpienia nawrotu choroby nowotworowej w ciągu pięciu lat od wykonania mastektomii

under the ROC curves. They are as follows: 0.9128 for the training set, 0.9689 for the validating set and 0.8750 for the testing set (the area under the ROC curve of the “ideal” classifier equals 1). Receiver Operating Characteristic curves (ROC) for

Table 2. Variables considered in analysis by artificial neural network and their values for six exemplary patients

Tabela 2. Zmienne rozważane w analizie metodą sztucznych sieci neuronowych i ich wartości dla 6 przykładowych pacjentów

Variable No. (Numer zmiennej)	Variable name (Nazwa zmiennej)	Variable value for patient (Wartość zmiennej dla pacjenta)					
		No. 1	No. 2	No. 3	No. 4	No. 5	No. 6
1	Age (years)	3	3	1	2	3	2
2	Menopause	3	3	1	1	3	1
3	Hormonal activity (years)	4	4	1	3	4	3
4	Number of births	3	1	3	1	4	1
5	Breast cancer in I and II line	2	2	2	2	2	2
6	Tumor size	2	2	3	3	3	1
7	Positive lymph nodes	1	1	3	1	4	4
8	Infiltration of node capsule	2	2	1	2	2	2
9	Arrests in microvessels	2	2	2	2	2	2
10	Malignancy: Bloom's degree	2	3	3	3	2	3
11	Surgery	3	3	3	2	2	3
12	Adjuvant therapy	4	4	5	3	7	5
13	Adjuvant radiotherapy	1	1	2	1	2	2
14	Neoadjuvant chemotherapy	2	2	2	1	1	2
15	Type of first line chemotherapy	1	1	2	2	2	2
16	Type of adjuvant hormonotherapy	2	2	1	1	2	1
17	Category	No recur- rence	No recur- rence	No recur- rence	Recur- rence	Recur- rence	Recur- rence

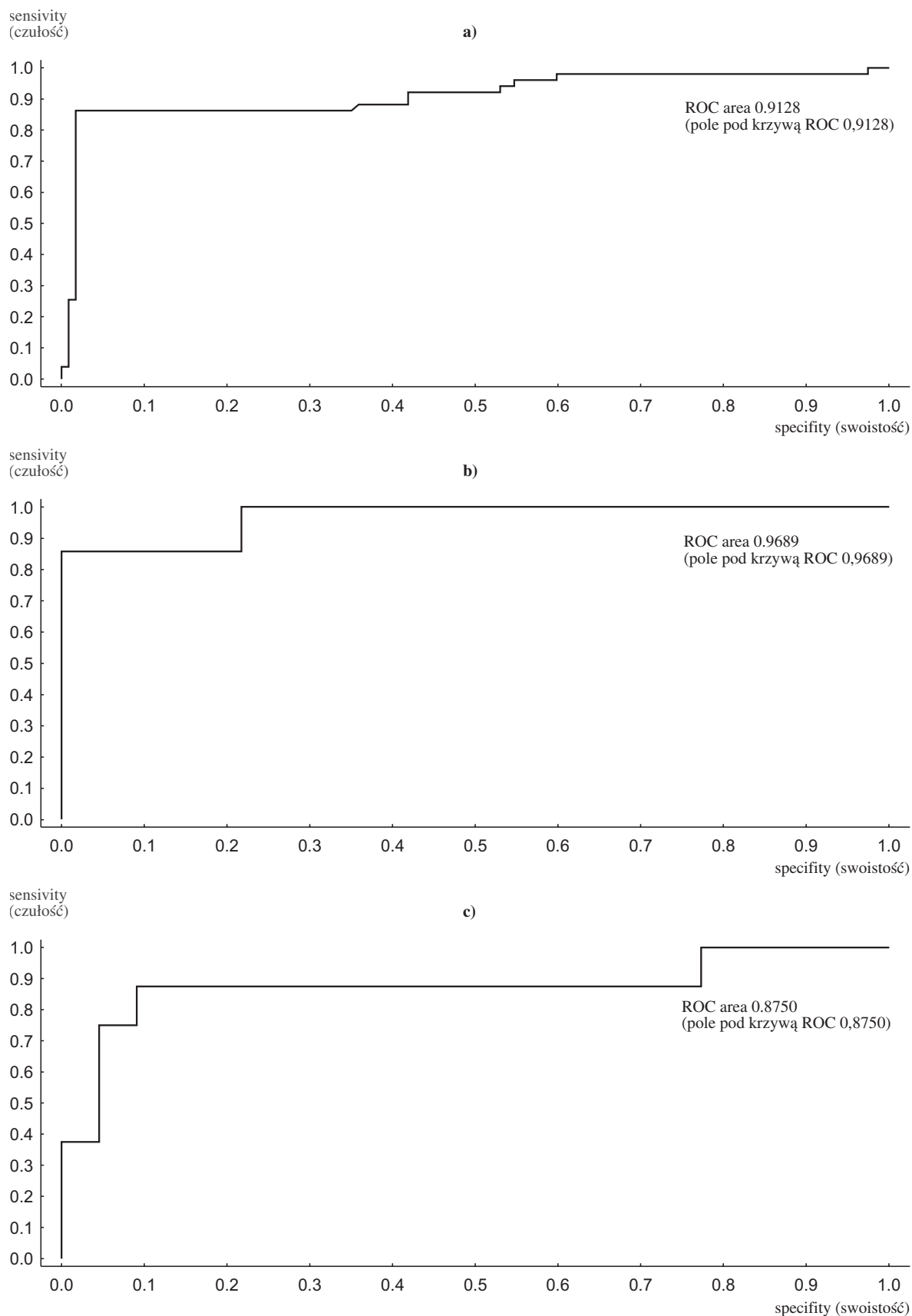


Fig. 2. ROC curves for: a) training set, b) validating set, c) testing set

Ryc. 2. Krzywe ROC dla zbiorów: a) trenującego, b) walidacyjnego, c) testującego

Table 3. Classification statistics**Tabela 3.** Statystyki klasyfikacyjne

	Training set (Zbiór uczący)		Validating set (Zbiór walidacyjny)		Testing set (Zbiór testujący)	
	no recurrence (brak wznowy)	recurrence (wznowa)	no recurrence (brak wznowy)	recurrence (wznowa)	no recurrence (brak wznowy)	recurrence (wznowa)
All together (Razem)	117	51	23	7	22	8
Right (Poprawnie)	100	44	21	6	19	7
Wrong (Błędnie)	17	7	2	1	3	1
No recurrence (Brak wznowy)	0	0	0	0	0	0
Recurrence						

the learning, validating and testing set is shown in Figures 2a, 2b and 2c, respectively.

With the use of the proposed method it was possible to differentiate the patients in the testing group as those with and without recurrence of breast cancer after a period of 5 years without significant error. The prognostic potency of ANN regarding the set of test patients is excellent and proves good choice and shape of the network proposed. Almost all patients (except one) in the testing set of the total number of 30 have been correctly classified, which means that one is able to predict the recurrence of breast cancer for the patients after mastectomy within 5 years after the treatment utilizing the variables used with very high probability.

The conducted research provides information, that the trained model based on the artificial neural networks is able to classify automatically the

individual patient characterized by the presented data into one of the classes. Moreover, the selected input data can be modified during the ANN analysis. For example, it is possible to exchange the kind of radiotherapy or chemotherapy and then observe the response of the trained network. Hence, the use of artificial neural networks in the analysis of clinical data is not limited only to treat it as a valuable tool for predictions of the recurrence. ANNs can be considered also as a prognostic method for the evaluation of patient response on the modification of treatment strategy. In the end, it must be additionally emphasized that ANNs enable to analyze efficiently the multivariable set of data at the same time, including intercorrelating data which is the obstacle in the case of regression methods and data for which restrictions in the application of classical statistical tests occurs.

Acknowledgement

The work was supported by the Polish State Committee for Scientific Research Project for years 2004–2007 no. 3 T11E 042 26.

References

- [1] McGuire WL: Adjuvant therapy of node-negative breast cancer. *N Engl J Med* 1998, 320, 525–527.
- [2] O'Higgins N: Aspects of breast cancers. *Chirurgia* 1992, 118, 324–327.
- [3] von Kleist S: Prognostic factors in breast cancer: theoretical and clinical aspects. *Anticancer Res* 1996, 16, 3907–3912.
- [4] Dhingra K, Hortobagyi GN: Critical evaluation of prognostic factors. *Semin Oncol* 1996, 23, 436–445.
- [5] Bonnier P, Romain S, Charpin C, Lejeune C, Tubiana N, Martin PM, Piana L: Age as a prognostic factor in breast cancer: relationship to pathologic and biological features. *Int J Cancer* 1995, 62, 138–144.
- [6] Chung M, Chang HR, Bland KI, Wanebo HJ: Younger women with breast carcinoma have a poorer prognosis than older women. *Cancer* 1996, 77, 97–103.
- [7] Aaltoma S, Lipponen PK, Kosma VM: Prognostic factors after 5 years follow-up in female breast cancer. *Oncology* 1992, 49, 93–98.
- [8] Beyer J, Kramar A, Mandanas R, Linkesch W, Greinix A, Droz JP, Pico JL, Diehl A, Bokemeyer C, Schmoll HJ, Nichols CR, Einhorn LH, Siegert W: High-dose chemotherapy as salvage treatment in germ cell tumours: a multivariate analysis of prognostic variables. *J Clin Oncol* 1996, 14, 2638–2645.
- [9] Maestu I, Pastor M, Gomez-Codina J, Aparicio J, Oltra A, Herranz C, Montalar J, Munarriz B, Reyes G:

- 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, 547–553.
- [10] **Jolliffe IT**: Principal Component Analysis. Springer, New York 1986.
- [11] **Zupan J, Gasteiger J**: Neural Networks for Chemists. An Introduction. VCH, Weinheim 1993.
- [12] **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.
- [13] **Kaliszan R**: Structure and Retention in Chromatography. A Chemometric Approach. Harwood Academic, Amsterdam 1997.
- [14] **Schneider G, Wrede P**: Artificial networks for computer-based molecular design. *Prog Biophys Mol Biol* 1998, 70, 175–222.
- [15] **Isu Y, Nagashima U, Hosoya H, Aoyama T**: Development of neural network simulator for structure-activity correlation of molecules: Neco. *J Chem Software* 1994, 2, 76–95.
- [16] **Andrea TA, Kalayeh H**: Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. *J Med Chem* 1991, 34, 2824–2836.
- [17] **So S-S, 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, 3201–3207.
- [18] **Ajay A**: Unified framework for using neural networks to build QSARs. *J Med Chem* 1993, 36, 3565–3571.
- [19] **Brickley MR, Shepherd JP, Armstrong RA**: Neural networks: a new technique for development of decision support systems in dentistry. *J Dent* 1998, 26, 305–309.
- [20] **Snow PB, Rodvold DM, Brandt JM**: Artificial neural networks in clinical urology. *Urology* 1999, 54, 787–790.
- [21] **Wei JT, Tewari A**: Artificial neural networks in urology: pro. *Urology* 1999, 54, 945–948.
- [22] **Krongrad A, Lai S**: Artificial neural networks in urology: con. *Urology* 1999, 54, 949–951.
- [23] **Lee CW, Park J-A**: Assessment of HIV/AIDS-related health performance using an artificial neural network. *Inform Manage* 2001, 38, 231–238.
- [24] **Mian S, Ball G, Hornbuckle J, Holding F, Carmichael J, Ellis I, Ali S, Li G, McArdle S, Creaser C, Rees R**: 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, 1725–1737.
- [25] **Biganzoli E, Boracchi P, Coradini D, Grazia DM, Marubini E**: Prognosis in node-negative primary breast cancer: a neural network analysis of risk profiles using routinely assessed factors. *Ann Oncol* 2003, 14, 1484–1493.
- [26] **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.
- [27] **Jerez-Aragones JM, Gomez-Ruiz JA, Ramos-Jimenez G, Munoz-Perez J, Alba-Conejo E**: A combined neural network and decision trees model for prognosis of breast cancer relapse. *Artif Intell Med* 2003, 27, 45–63.
- [28] **Abbass HA**: An evolutionary artificial neural networks approach for breast cancer diagnosis. *Artif Intell Med* 2002, 25, 265–281.
- [29] **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.
- [30] www.statsoft.com.

Address for correspondence:

Adam Buciński
Cellegium Medicum
Department of Biopharmacy
Jagiellońska 13
85-067 Bydgoszcz
Poland
e-mail: kizbiofarmacji@cm.umk.pl

Received: 14.12.2004

Revised: 27.04.2005

Accepted: 27.04.2005

Praca wpłynęła do Redakcji: 14.12.2004 r.

Po recenzji: 27.04.2005 r.

Zaakceptowano do druku: 27.04.2005 r.