

# ZASTOSOWANIE ZAAWANSOWANYCH METOD MODELOWANIA STATYSTYCZNEGO W OCENIE PRAWDOPODOBIENSTWA ISTNIENIA RAKA JAJNIKA U KOBIET Z GUZAMI PRZYDATKOWYMI

Agata Barbara Smoleń

*Zakład Matematyki i Biostatystyki Medycznej, Akademia Medyczna, Lublin*

## 1 WPROWADZENIE

Nowotwory jajnika, pomimo intensywnych badań nad histogenezą, epidemiologią oraz wprowadzania nowych sposobów leczenia ciągle stanowią najtrudniejszy problem w onkologii ginekologicznej. Rak jajnika nazywany jest „cichym zabójcą” przede wszystkim ze względu na trudności diagnostyczne, które spowodowane są m.in. brakiem znanych prekursorów i łatwo dostępnego materiału biologicznego. Stosunkowo rzadkie występowanie i podstępny charakter rozwoju raka jajnika sugeruje konieczność wykrywania wszystkich wczesnych przypadków tej choroby [6, 14, 15]. Brak dolegliwości we wczesnej fazie choroby i wiarygodnych testów diagnostycznych powoduje że zdecydowana większość, bo ponad 70% kobiet ciągle jeszcze zgłasza się do lekarza już w zaawansowanym stadium klinicznym choroby. Szansa wyleczenia kobiety w tej fazie choroby nie przekracza 30% i od lat nie zmienia się istotnie pomimo wprowadzenia nowych sposobów diagnostyki biochemicznej i obrazowej oraz niezwykle kosztownego leczenia [2, 23]. W Polsce odnotowuje się corocznie ok. 3000 nowych zachorowań, a prawie 2000 kobiet umiera w tym czasie z powodu tego nowotworu [24]. Zestawione badanie ginekologiczne pozostaje w dalszym ciągu jedną z podstawowych metod wykrywania guzów jajnika. Pomimo zastosowania najnowocześniejszych metod obrazowania, w tym ultrasonografii z wykorzystaniem kolorowego i „power angio” Dopplera oraz oznaczeń szeregu nowych markerów biochemicznych czułość i specyficzność tych metod jest ciągle niewystarczająca. Ze względu na to, że pewność co do natury guza

przynosi dopiero badanie histopatologiczne, tradycyjnym sposobem weryfikacji diagnozy pozostaje w dalszym ciągu wykonanie laparotomii lub laparoskopii.

Istniejące liczne kontrowersje dotyczące przydatności badania klinicznego, oznaczeń markerów i nowoczesnych badań ultradźwiękowych we wstępnej charakterystyce guzów jajnika sugerują konieczność podjęcia prób innych, bardziej zaawansowanych metod analizy statystycznej danych klinicznych. Ocena jakie jest prawdopodobieństwo, że wykryty u badanej kobiety guz przydatkowy może być złośliwy jest niezwykle ważna, gdyż ma bezpośredni wpływ na rodzaj dalszego postępowania. Uzasadnione podejrzenie istnienia raka jajnika łączy się bowiem z koniecznością wykonania rozszerzonego zabiegu operacyjnego. Problem ten ma istotne znaczenie również ze względu na koszty leczenia oraz na potrzebę wykonywania operacji oszczędzających narząd rodny wtedy, gdy jest to tylko możliwe. Największą nadzieję na stworzenie modelu decyzyjnego, który będzie jeszcze przed operacją określał stopień zaawansowania raka jajnika daje ocena kombinacji testów i cech klinicznych, które zostaną następnie razem przeanalizowane w procesie przewidywania istnienia tej niebezpiecznej choroby. Wykorzystanie komputerowego modelowania w ocenie ryzyka istnienia raka jajnika w badanym guzie przydatkowym wydaje się być jednym z rozwiązań tego problemu.

## 2 CELE PRACY

Ze względu na przedstawione wyżej problemy głównym celem pracy było zbudowanie modelu predykcyjnego pozwalającego ocenić jaki jest wpływ statusu menopauzalnego kobiety, wyniku badania w ultrasonografii dopochwowej i dopplerowskiej oceny angiogenezy na dokładność przedoperacyjnego różnicowania guzów jajnika. Drugim celem było zbadanie czy kombinacja dwóch lub więcej z analizowanych parametrów może podwyższyć wartości prognostyczne pojedynczego testu stosowanego w różnicowaniu niezłośliwych i złośliwych guzów przydatków macicy. Kolejnym celem prowadzonych badań było prospektywne sprawdzenie skuteczności diagnostycznej uzyskanego modelu statystycznego.

## 3 MATERIAŁ I METODY

### 3.1 Pacjentki

Badania zostały przeprowadzone w oparciu o dane zebrane w Pracowni USG I Kliniki Ginekologii AM w Lublinie. Do analizy zakwalifikowano 307 kobiet z wykrytymi w badaniu ultrasonograficznym guzami przydatkowymi. Weryfikację rodzaju guza stanowiły wyniki pooperacyjnego badania histopatologicznego lub udokumentowana regresja zmiany w kontrolnych badaniach ultrasonograficznych. Guzy jajnika klasyfikowano zgodnie z kryteriami histologicznymi podanymi przez WHO [Serov i wsp. 1973]. Stadium zaawansowania klinicznego określano przy pomocy kryteriów podanych przez Międzynarodową Federację Ginekologów i Położników [FIGO, 1987]. Analizowano wyniki badania ultrasonograficznego z zastosowaniem kolorowego i spektralnego Dopplera. Dla każdej pacjentki oceniono następujące cechy kliniczne i ultrasonograficzne: wiek, status menopauzalny, obustronność i największy wymiar guza przydatkowego, jego objętość, obecność elementów litych, przegród  $>3\text{mm}$  i wyrosły brodawkowatych  $>3\text{mm}$ . W badaniu dopplerowskim analizowano subiektywną ocenę unaczynienia (wg Timmermanna i wsp. 1999) oraz zmierzone w naczyniach guza wartości indeksów pulsacji PI i oporu RI oraz największą prędkość skurczową PSV. Ponadto analizowano wyniki oznaczenia stężenia antygenu CA-125

w surowicy krwi, jako najczęściej oznaczanego markera raka jajnika.

### 3.2 Analiza statystyczna

Uzyskane wyniki badań poddano analizie statystycznej. Wartości analizowanych parametrów mierzonych w skali nominalnej scharakteryzowano przy pomocy liczebności i odsetka; w skali porządkowej przy pomocy mediany, modalnej i zakresu zmienności, natomiast mierzone w skali interwałowej z uwagi na skośny rozkład badanych parametrów scharakteryzowano przy pomocy mediany wraz z zakresem zmienności i z podaniem 25-go i 75-go percentyla. Do wykrycia istotności różnic między porównywanymi grupami guzów złośliwych i niezłośliwych użyto testu jednorodności  $\chi^2$  dla cech jakościowych oraz testu mediany lub U Manna-Whitney'a dla cech ilościowych.

Dokładność testu została określana przez porównanie pomiędzy wynikami testu a pooperacyjnym wynikiem histopatologicznym, który stwierdza rzeczywistą sytuację chorobową. Zastosowano podstawowe miary dokładności testu tj: czułość (sensitivity - SENS), specyficzność (specifity - SPEC), dokładność (accuracy - ACC), dodatnią wartość predykcyjną (positive predictive value - PPV) oraz ujemną wartość predykcyjną (negative predictive value - NPV). W sytuacjach, kiedy wartości testu diagnostycznego były ciągłe lub miały wiele kategorii obliczono wartości predykcyjne dla najlepiej różnicujących charakter guza jajnika wartości granicznych. W przeprowadzonych obliczeniach zmiany czułości i specyficzności testów przy przesuwaniu wartości granicznej przedstawiono w formie krzywych ROC („Receiver Operating Characteristic Curves”) oraz wykorzystano dodatkowe informacje porównując pola pod krzywą ROC [10], które skonstruowano przy pomocy programu GraphROC for Windows (GraphROC Software, Finlandia).

W dalszej części pracy dokonano oceny czy i które z wybranych parametrów mają istotny statystycznie wpływ na prawdopodobieństwo istnienia nowotworu złośliwego lub niezłośliwego. W celu znalezienia najlepszej kombinacji cech istotnie wpływających na istnienie procesu złośliwego i obliczenie indywidualnego prawdopodobieństwa istnienia nowotworu jajnika zastosowano wstępującą analizę regresji logistycznej

z estymacją quasi-Newtona. Wartości przewidywane, które z uwagi na zastosowany model logistyczny należą do przedziału  $<0,1>$  traktowano jako prawdopodobieństwo istnienia procesu złośliwego bądź niezłośliwego. Model ten przedstawiał się następująco:  $p = \frac{1}{1 + e^{-z}}$ ,

gdzie  $e$  jest stałą matematyczną, granicą ciągu  $\lim_{n \rightarrow +\infty} (1 + \frac{1}{n})^n = e = 2,71828\dots$ , a  $z$  jest kombinacją

liniową zmiennych  $x_i$  i ich estymatorów  $b_i$  zawartych w modelu:  $z = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n$ .

W zastosowanym modelu wartości prawdopodobieństwa  $\leq 0,5$  klasyfikowały nowotwór jako niezłośliwy, a przypadki z wartościami predykcyjnymi  $> 0,5$  jako nowotwór złośliwy.

Do obliczenia prawdopodobieństwa istnienia nowotworu złośliwego na podstawie danych ultrasonograficznych zastosowano również sztuczne sieci neuronowe. Użyto klasyfikacji wzorcowej, której zadaniem było przypisanie każdego badanego przypadku do jednej z dwóch klas (guz złośliwy bądź niezłośliwy). Zadanie to było realizowane przy pomocy sieci liniowych, perceptronów wielowarstwowych, sieci o radialnych funkcjach bazowych oraz probabilistycznych sieci neuronowych. Dostępny zbiór danych podzielono w sposób losowy na trzy zbiory: uczący, walidacyjny i testowy. Podstawową metodą uczenia sieci wielowarstwowych był algorytm wstecznej propagacji błędów z uwagi na kierunek przepływu informacji o błędzie. Użyto funkcji logistycznej do transformacji każdego z węzłów w warstwie ukrytej i wyjściowej. Ta nieliniowa funkcja przedstawia się

następująco:  $Tf(\sum_{i=1}^n w_i x_i) = \frac{1}{1 + e^{-\sum_{i=1}^n w_i x_i}}$ , gdzie

$Tf(\sum_{i=1}^n w_i x_i)$  jest wartością wyjściową aktywującą neuron „ $i$ ”, a  $w_i x_i$  jest wartością wejściową neuronu do sieci ( $w_i$ -wagi;  $x_i$ -pojedyncze wartości zmiennej) [17, 18].

Bazę danych wykonano przy pomocy programu Excel XP for Windows, natomiast analizy statystyczne przeprowadzono z użyciem oprogramowania komputerowego *STATISTICA* (StatSoft, Polska) i *STATISTICA Neural Networks* (StatSoft, Polska). W pracy przyjęto 5%

ryzyko błędu wnioskowania i związany z nim poziom istotności  $p < 0,05$  wskazujący na istnienie istotnych statystycznie różnic.

#### 4 WYNIKI

Do analizy zakwalifikowano 307 kobiet w wieku od 13 do 84 lat (średnia 42,71 lata, mediana 43 lata). Wśród nich było 228 (74,27%) kobiet z niezłośliwymi guzami przydatkowymi oraz 79 (25,73%) kobiet z nowotworami złośliwymi. Średnia wieku w grupie kobiet z guzami złośliwymi wynosiła  $51,47 \pm 14,48$  lat, natomiast w grupie kobiet z guzami niezłośliwymi  $39,68 \pm 14,46$  lat. Porównanie przy pomocy testu U Manna-Whitney'a wykazało, że różnica ta jest istotna statystycznie ( $Z=5,94$ ,  $p < 0,000001$ ). Odsetek pacjentek po menopauzie w grupie chorych z guzami niezłośliwymi wynosił 15,79% (36 z 228). W grupie kobiet z nowotworami złośliwymi odsetek ten wynosił 48,1% (38 z 79). Porównanie przy pomocy testu jednorodności  $\chi^2$  wykazało istotną statystycznie różnicę w częstości występowania nowotworów złośliwych w zależności od statusu menopauzalnego ( $\chi^2=33,48$ ,  $p < 0,00001$ ). W badanej grupie 307 kobiet 77 z nich tj. 25,08% miało obustronne guzy przydatkowe. Znacznie częściej guzy te występowały w grupie kobiet z nowotworem złośliwym ( $\chi^2=57,54$ ,  $p < 0,00001$ ). Wśród pacjentek z rakiem jajnika stopień zaawansowania klinicznego wg FIGO przedstawiał się następująco: 21 chorych było w I stopniu, w II stopniu było 10 pacjentek, pozostałe były w III stopniu (44 chore) lub w IV stopniu (4 chore) stopniu zaawansowania.

Charakterystykę morfologii ultrasonograficznej badanych guzów przydatkowych w grupie 307 pacjentek przedstawia tabela 1.

Przydatność diagnostyczną poszczególnych testów stosowanych pojedynczo dla wartości progowej najlepiej dyskryminującej guzy złośliwe i niezłośliwe przedstawiono w tabeli 2.

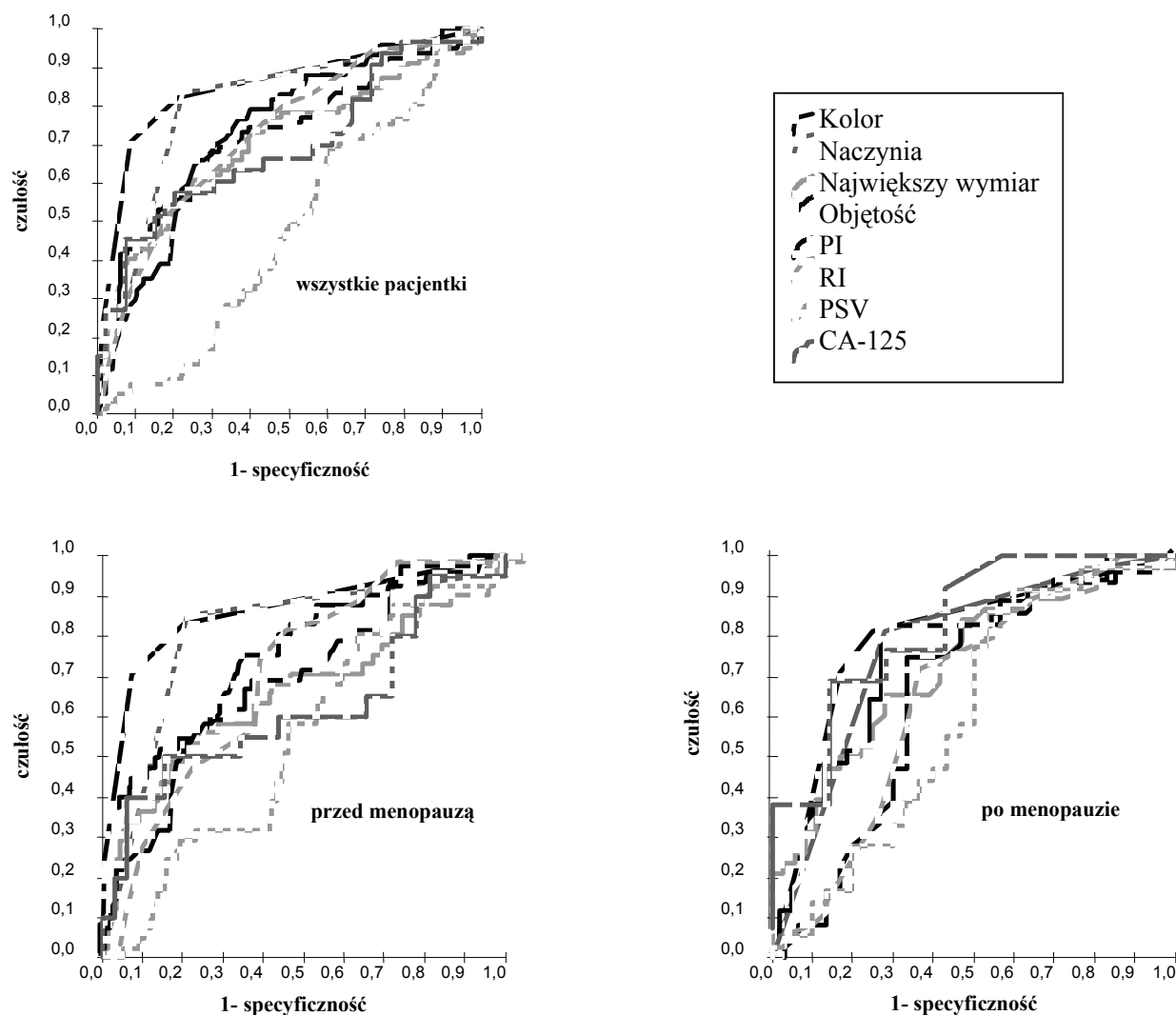
Najwyższą dokładność testu diagnostycznego wynoszącą 85,89% stwierdzono dla dopplerowskiej oceny cechy „Kolor” równej „2”. U kobiet przed menopauzą najwyższą dokładność diagnostyczną wynoszącą 88,84% miała wartość progowa cechy „Kolor” równa „2”. W grupie kobiet po menopauzie najwyższą dokładność testu wynoszącą 78,38% stwierdzono dla cechy „Kolor” równej „1”.

Tabela 1. Wartości prognostyczne kryteriów morfologicznych przydatnych w różnicowaniu guzów przydatkowych.

Cecha	SENS	95% CI	SPEC	95% CI	ACC	PPV	NPV
Obustronność	56,96	45,29-68,09	85,97	80,75-90,21	78,5	58,44	85,22
Elementy lite	83,54	73,44-90,98	50,44	43,75-57,11	58,96	36,87	89,84
Przegrody	91,14	82,51-96,41	60,96	54,3-67,34	68,73	44,72	95,21
Wyrośla brodawkowate	74,68	63,6-83,83	77,19	71,18-82,48	76,55	53,15	89,8

Tabela 2. Wartości prognostyczne poszczególnych testów w badanej grupie 307 kobiet.

Parametr (optymalna wartość graniczna)	SENS	95% CI	SPEC	95% CI	ACC	PPV	NPV
Kolor (2)	70,87	59,54-80,6	91,23	86,76-94,57	85,89	73,68	90,04
Naczynia (1)	83,54	73,44-90,98	77,63	71,65-82,87	79,15	56,41	93,16
Najw. wym. (119mm)	39,24	28,41-50,91	93,42	89,36-96,28	79,48	67,39	81,61
Objętość (527ml)	41,77	32,38-51,68	93,86	90,51-96,23	80,46	70,21	82,31
PI (0,53)	24,68	15,22-35,87	94,09	89,87-96,92	75,00	61,29	76,71
RI (0,41)	31,17	21,06-42,79	92,61	88,08-95,82	75,71	61,54	78,01
PSV (4,6cm/s)	0,48	0,00-4,68	99,51	97,21-100	72,14	0,48	72,40
CA-125 (101,68 U/ml)	45,45	28,00-63,76	92,31	78,89-98,52	70,83	83,33	66,67



Rys. 1. Porównanie krzywych ROC dla zastosowanych testów diagnostycznych ocenianych wśród kobiet bez uwzględnienia statusu menopauzalnego oraz u pacjentek przed i po menopauzie.

Ze względu na to, że a priori nie wiadomo który z dwóch testów jest skuteczniejszy, jedną z możliwych metod różnicowania jest równoległa wzrokowa ocena przebiegu krzywych ROC. Skonstruowane krzywe ROC dla obliczonych różnych wartości progowych zastosowanych testów diagnostycznych przedstawiono na powyższych rysunkach.

Rysunek 1. przedstawia porównanie wartości prognostycznych jednocześnie dla wszystkich analizowanych testów w grupach kobiet bez uwzględniania statusu menopauzalnego oraz w podgrupach pacjentek przed i po menopauzie. Na każdym z wykresów na osi Oy przedstawiono czułość, czyli odsetek wyników prawdziwie pozytywnych, a na osi Ox 1-specyficzność, czyli odsetek wyników fałszywie pozytywnych.

Tabele 3, 4, 5 przedstawiają analizę pola powierzchni pod krzywymi ROC dla ocenianych parametrów z uwzględnieniem wielkości błędu standardowego (SE) oraz 95% przedziału ufności (CI).

Tabela 3. Porównanie wartości diagnostycznej ocenianych parametrów w oparciu o pole powierzchni pod krzywą ROC dla wszystkich pacjentek.

Badany parametr	Pole pod krzywą (AUROC)	SE	95% CI
Kolor	0,8500	0,0286	0,7928-0,9072
Naczynia	0,7987	0,0304	0,7379-0,8595
Największy wymiar	0,7057	0,0383	0,6291-0,7823
Objętość	0,7226	0,0359	0,6508-0,7944
PI	0,7379	0,0323	0,6733-0,8025
RI	0,7270	0,0337	0,6596-0,7944
PSV	0,4721	0,0368	0,3985-0,5457
CA 125	0,6904	0,0645	0,5614-0,8194

Tabela 4. Porównanie wartości diagnostycznej ocenianych parametrów w oparciu o pole powierzchni pod krzywą ROC w podgrupie pacjentek przed menopauzą.

Badany parametr	Pole pod krzywą (AUROC)	SE	95% CI
Kolor	0,8617	0,0381	0,7855-0,9379
Naczynia	0,8062	0,0408	0,7246-0,8878
Największy wymiar	0,6557	0,0554	0,5449-0,7665
Objętość	0,6926	0,0501	0,5924-0,7928
PI	0,7309	0,0416	0,6477-0,8141
RI	0,7338	0,0421	0,6496-0,818
PSV	0,5464	0,0475	0,4514-0,6414
CA 125	0,6203	0,0872	0,4459-0,7947

Tabela 5. Porównanie wartości diagnostycznej ocenianych parametrów w oparciu o pole powierzchni pod krzywą ROC w podgrupie pacjentek po menopauzie.

Badany parametr	Pole pod krzywą (AUROC)	SE	95% CI
Kolor	0,8015	0,0539	0,6937-0,9093
Naczynia	0,7774	0,0559	0,6656-0,8892
Największy wymiar	0,7288	0,0586	0,6116-0,846
Objętość	0,7310	0,0606	0,6098-0,8522
PI	0,6662	0,0712	0,5238-0,8086
RI	0,6569	0,0710	0,5149-0,7989
PSV	0,6046	0,0740	0,4566-0,7526
CA 125	0,7297	0,0896	0,5505-0,9089

Analiza krzywych ROC wskazuje, że u kobiet bez uwzględniania ich statusu menopauzalnego najlepsze wartości prognostyczne miały: ocena cechy „Kolor” i lokalizacji naczyń, a następnie pomiary objętości, indeksu pulsacji PI i oporu RI. Dodatkowe wyróżnienie kobiet po menopauzie wskazuje, że wartość diagnostyczna wzrastała w tej grupie w przypadku oznaczenia parametrów: „Kolor”, lokalizacja naczyń i objętość. W podgrupie kobiet przed menopauzą najwyższą dokładność diagnostyczną miały wyniki koloru, unaczynienia, indeksu pulsacji i oporu.

#### 4.1 Wyniki analizy regresji logistycznej

Zastosowanie wstępującej analizy regresji logistycznej z estymacją quasi-Newtona przy wykorzystaniu wszystkich parametrów diagnostycznych w badanej grupie kobiet pozwoliło stwierdzić, że spośród analizowanych kilkunastu zmiennych tylko 6 miało istotny statystycznie wpływ ( $p < 0,05$ ) na konstruowany model i w związku z tym zostały uwzględnione w równaniu regresji. Zmienne te to: wiek, guz jedno- lub obustronny, objętość, przegrody, wyrośla brodawkowate oraz cecha „Kolor”. Model ten opisany poprzez uwzględniane zmienne przedstawia się następująco:

prawdopodobieństwo istnienia procesu złośliwego  $P(X) = \frac{1}{1 + e^{-z}}$ , gdzie  $e$  jest stałą matematyczną, a

$$\log it P = z = -9,15 + 0,045 \text{ wiek} + 0,84 \text{ obust} + 1,66 \text{ obj.} + 1,57 \text{ sept} + 1,23 \text{ pap} + 1,09 \text{ kolor}$$

(1,13)      (0,01)      (0,44)      (0,54)      (0,51)  
(0,41)      (0,21)

Dodatnie wartości parametrów odpowiadające badanym zmiennym wskazują, że ich wzrost zwiększa prawdopodobieństwo istnienia nowotworu złośliwego w badanym guzie.

Najwyższa czułość i specyficzność przy założonym 50% ryzyku istnienia nowotworu złośliwego dla uzyskanego modelu predykcyjnego wynosiły odpowiednio 74,68% (CI 63,6%-83,83%) i 94,74% (CI 90,46%-96,96%).

W modelu tym 20 z 79 przypadków nowotworów złośliwych błędnie przypisano prawdopodobieństwo mniejsze niż 50%. Jednocześnie 12 z 228 przypadków guzów niezłośliwych zostało błędnie zaklasyfikowanych jako złośliwe z prawdopodobieństwem większym niż 50%. Pozytywna i negatywna wartość predykcyjna wynosiły odpowiednio 83,1% i 91,53%.

Analogiczną analizę przeprowadzono w podgrupach kobiet przed menopauzą i po menopauzie. Otrzymany model dla kobiet przed menopauzą przedstawiał się w następującej postaci logitowej:

$$\log itP = z = -9,74 + 0,05 \text{ wiek} + 2,1 \text{obj} + 1,18 \text{sept} + 1,14 \text{pap} + 1,35 \text{ kolor}$$

(1,6) (0,03) (0,77) (0,61) (0,52) (0,3)

Najwyższe czułość i specyficzność przy założonym 50% ryzyku istnienia nowotworu złośliwego u kobiet przed menopauzą dla tego modelu wynosiły odpowiednio 65,85% (CI 52,81%-85,28%) i 96,88% (CI 93,28%-98,87%). W modelu tym 14 z 41 przypadków nowotworów złośliwych błędnie przypisano prawdopodobieństwo mniejsze niż 50%. Jednocześnie 6 z 192 przypadków guzów niezłośliwych zostało błędnie zaklasyfikowanych jako złośliwe z prawdopodobieństwem większym niż 50%. Pozytywna i negatywna wartość predykcyjna wynosiły odpowiednio 81,82% i 93%.

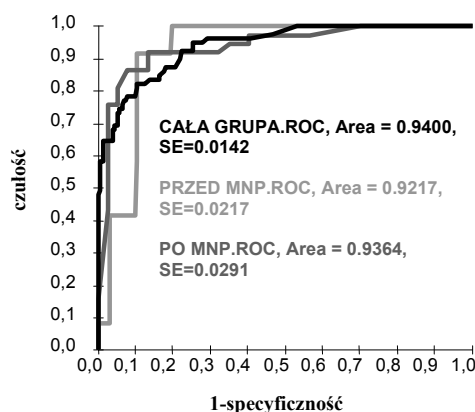
Model uzyskany w wyniku analizy regresji logistycznej w podgrupie kobiet po menopauzie przedstawiał się w następującej postaci logitowej

$$\log itP = z = 4,21 + 3,82 \text{ obust.} + 3,86 \text{ sept} + 1,97 \text{ PI} + 0,14 \text{ PSV}$$

(1,41) (1,19) (1,26) (0,99) (0,06)

Najwyższe czułość i specyficzność przy założonym 50% ryzyku istnienia nowotworu złośliwego u kobiet po menopauzie dla tego modelu wynosiły odpowiednio 84,21% (CI 71,03%-95,59%) i 94,44% (CI 77,84%-98,44%). W modelu tym 6 z 38 przypadków nowotworów złośliwych błędnie przypisano prawdopodobieństwo mniejsze niż 50%. Jednocześnie 2 z 36 przypadków guzów niezłośliwych zostało

błędnie zaklasyfikowanych jako złośliwe z prawdopodobieństwem większym niż 50%. Pozytywna i negatywna wartość predykcyjna wynosiły odpowiednio 94,12% i 85%.

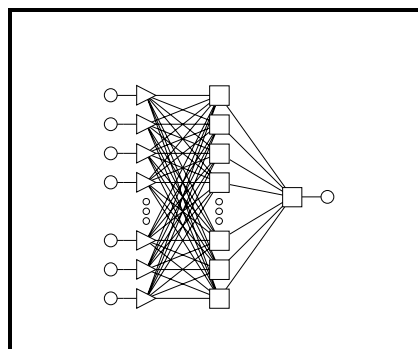


Rys. 2. Porównanie krzywych ROC dla otrzymanych modeli.

Analiza pól pod krzywymi ROC wykazała brak istotnych statystycznie różnic w dokładności diagnostycznej uzyskanych modeli dla wszystkich pacjentek i dla podgrup z uwzględnieniem statusu menopauzalnego (wszystkie pacjentki w porównaniu do kobiet przed menopauzą  $p=0,48$ , wszystkie pacjentki w porównaniu do kobiet po menopauzie  $p=0,91$ , przed menopauzą w porównaniu do kobiet po menopauzie  $p=0,68$ ). Wartość pola powierzchni pod krzywą ROC była największa dla całej badanej grupy, w związku z czym uznano ten model za najlepszy z uzyskanych.

#### 4.2 Sieci trójwarstwowe

Najlepszą siecią z wszystkich kolejnych szkoleń okazała się trójwarstwowa sieć MLP o 13 neuronach wejściowych i 11 w warstwach ukrytych. Średni błąd kwadratowy wynosił 0,09 dla zbioru szkoleniowego i 0,006 dla zbioru testowego.



Rys. 3. Model otrzymanej sieci.

Strukturę drugiej otrzymanej sieci przedstawia powyższy rysunek.

Wartości statystyk klasyfikacyjnych dla otrzymanego modelu przedstawia tabela 6.

Tabela 6. Statystyki klasyfikacyjne otrzymanego modelu.

	niezłosl	złoslwy	niezłosl	złoslwy	niezłosl	złoslwy
Total	188	62	26	14	14	3
Correct	188	60	26	14	13	0
Wrong	0	2	0	0	1	3
Unknown	0	0	0	0	0	0
niezłosl	188	2	26	0	13	3
złoslwy	0	60	0	14	1	0

Przy założonym prawdopodobieństwie istnienia guza złośliwego większym niż 75% sieć pozwoliła na obliczenie czułości i specyficzność wynoszącej odpowiednio 96,77% i 100% w zbiorze szkoleniowym. W zbiorze testowym i walidacyjnym zastosowanie tego poziomu decyzyjnego dało czułość i specyficzność o wartości odpowiednio 82,35% i 97,5%.

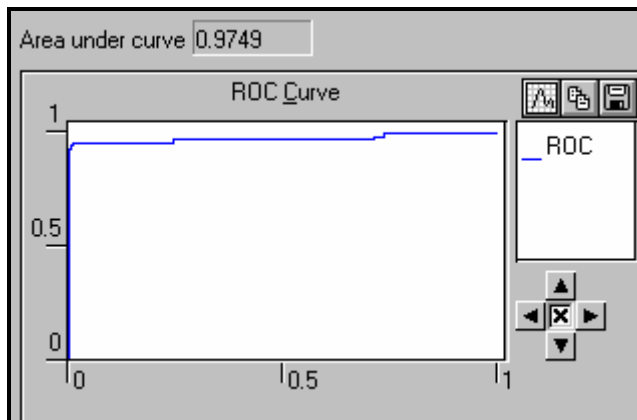
Tabela 7. Analiza wrażliwości modelu.

	WIEK	MNP	OBUSTR	NAJWYM	OBJ	LITE	SEPT	PAP	VESS
Rank	7	10	5	11	12	1	6	4	3
Error	0.2482956	0.1867146	0.2778268	0.1499226	0.1434936	0.3231484	0.2694086	0.2974988	0.3058666
Rank	2.775936	2.087463	3.106094	1.676129	1.604253	3.612788	3.011979	3.326026	3.419578
Rank	4	6	2	13	10	3	5	9	1
Error	0.2500589	0.2423114	0.2770176	0.003286	0.1581879	0.2606987	0.2481211	0.1906836	0.3455031
Rank	44.91248	43.52096	49.75446	0.5901409	28.41174	46.82346	44.56444	34.24822	62.05499

KOLOR	PI	RI	VMAX
2	13	9	8
0.3078977	0.1390837	0.1885894	0.1948706
3.442285	1.554951	2.108423	2.178646
8	12	11	7
0.2057935	0.1457522	0.1581751	0.2072023
36.96208	26.1782	28.40945	37.2151

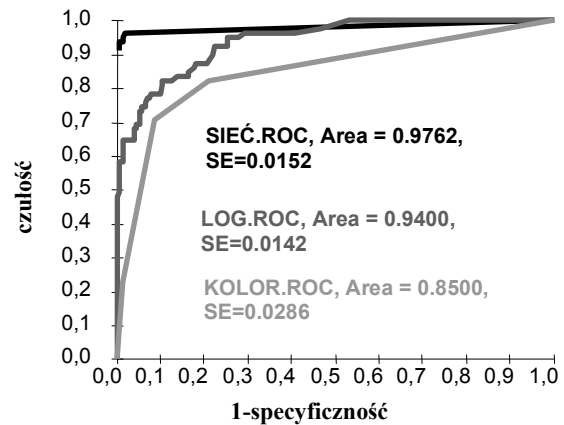
Przykład oceny skuteczności uczenia sieci przedstawia rysunek 4.



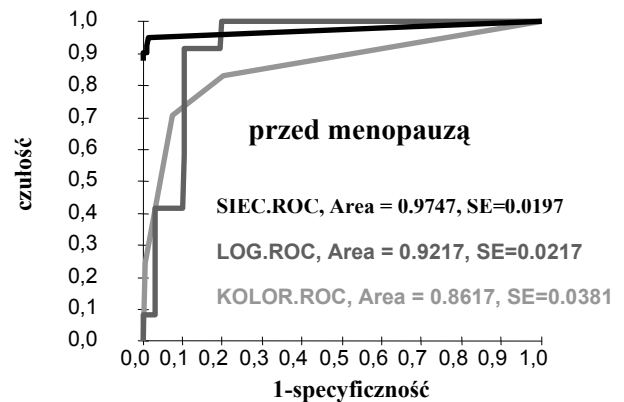
Rys. 4. Wynik oceny pola pod krzywą ROC.

Porównanie dokładności diagnostycznej otrzymanych modeli i najlepiej różnicującego

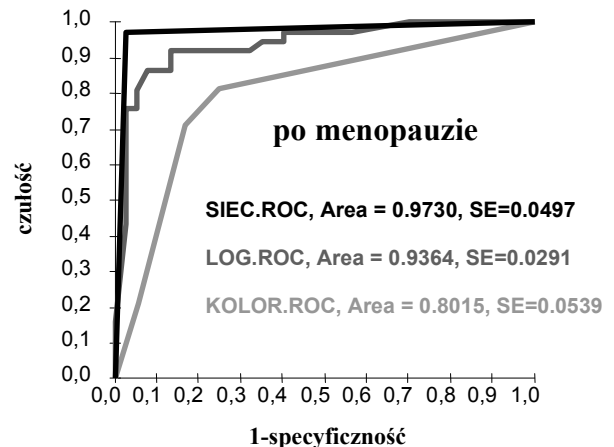
parametru w całej grupie pacjentek oraz w podgrupach z uwzględnieniem statusu menopauzalnego przedstawiają poniższe rysunki.



Rys. 5. Porównanie dokładności diagnostycznej sieci, regresji logistycznej i najlepiej różnicującego parametru w całej grupie pacjentek.



Rys. 6. Porównanie dokładności diagnostycznej sieci, regresji logistycznej i najlepiej różnicującego parametru w podgrupie kobiet przed menopauzą.

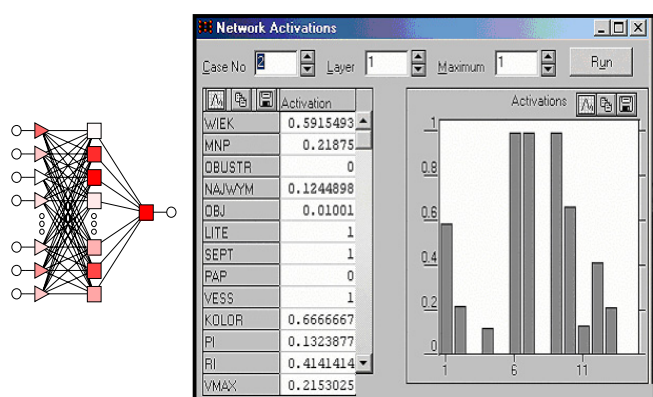


Rys. 7. Porównanie dokładności diagnostycznej sieci, regresji logistycznej i najlepiej różnicującego parametru w podgrupie kobiet po menopauzie.

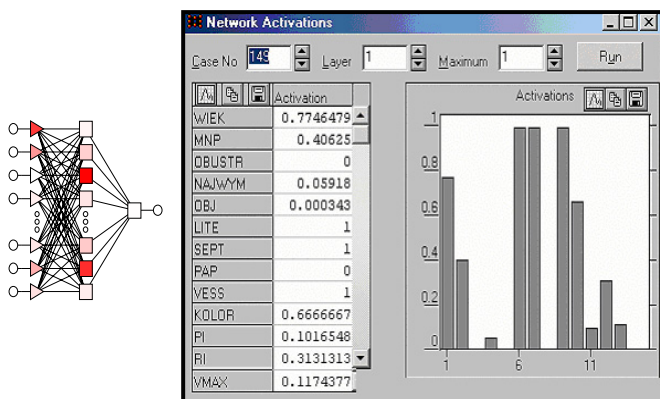
Wykorzystanie sieci jako modelu prognostycznego pozwoliło skuteczniej niż model regresji logistycznej i ilościowa ocena cechy „Kolor” przewidywać charakter badanego guza jajnika.

#### 4.3 Wybrane przykłady analizy i testowania sieci neuronowej

Poziom aktywacji sieci przedstawia graficzna i numeryczna prezentacja wartości wyjściowych neuronów w wybranej warstwie.



Rys. 8. Przypadek A – nowotwór złośliwy ( $p=0,9995523$ ). Wynik zgodny z wynikiem histopatologicznym.



Rys. 9. Przypadek B – nowotwór niezłośliwy ( $p=3,976e-05$ ). Wynik zgodny z wynikiem histopatologicznym.

#### 4.4 Prospektywna ocena przy użyciu modelu o najwyższej wartości diagnostycznej

Prospektywną ocenę na zbiorze liczącym 62 przypadki (11 złośliwych i 51 niezłośliwych) przeprowadzono z zastosowaniem sieci z zastosowaniem sieci o największej wartości diagnostycznej i zdolności do generalizacji. Przy testowaniu sieć popełniła 2 błędy w podgrupie nowotworów złośliwych 3 błędy w podgrupie nowotworów niezłośliwych. Wartości

predykcyjne: czułość i specyficzność wynosiły odpowiednio 81,82% i 94,12%.

## 5 DYSKUSJA

Większość zjawisk biologicznych w tym w szczególności takich jak procesy prowadzące do rozwoju nowotworu złośliwego jest złożona i z punktu widzenia matematyka nieliniowa w swej naturze. Modelowanie nieliniowe może być zrealizowane za pomocą zaawansowanych technik statystycznych takich jak analiza regresji logistycznej oraz sztucznych sieci neuronowych. W ostatnim dziesięcioleciu w dużej mierze dzięki rozwojowi techniki komputerowej metody te zostały zaakceptowane i obecnie są coraz szerzej stosowanymi sposobami prognozowania i klasyfikowania zjawisk medycznych.

Wartość omawianych metod diagnostycznych jest ograniczona ze względu na ich niskie relatywnie wartości predykcyjne. Obliczenie czułości, specyficzności oraz dodatnich i ujemnych wartości predykcyjnych pozwala na właściwą ocenę przydatności testu [4, 5, 12, 16]. Wartości predykcyjne powszechnie stosowanych metod wczesnego rozpoznawania raka jajnika pozostają niezadowolające ponieważ nie dają dostatecznej informacji dotyczącej oceny rzeczywistego prawdopodobieństwa istnienia guza złośliwego u każdej chorej. Żadna z pojedynczo stosowanych metod diagnostycznych nie nadaje się aktualnie do populacyjnych badań przesiewowych tej potencjalnie śmiertelnej choroby [13, 22]. Indywidualne testy dla określenia stadium klinicznego raka jajnika zawodzą w kwestii dokładności i użyteczności praktycznej. Pomimo stosowania nowoczesnych metod diagnostycznych potrzeba udoskonalania nieinwazyjnych metod wykrywania i oceny charakteru guza przydatkowego doprowadziła do prób zastosowania zaawansowanych metod analizy statystycznej dostępnych danych klinicznych i ultrasonograficznych.

Kryteria, które odróżniają zdecydowanie użyteczny test od takiego o wątpliwej wartości są generalnie dobrze ustalone i zostały wcielone do zasad krytycznej oceny literatury medycznej [3]. Zgodnie z tymi zasadami nowe sposoby modelowania statystycznego jak regresja logistyczna czy sieci neuronowe są preferowane w stosunku do tradycyjnych pojęć czułości i specyficzności w ocenie klinicznej wartości testu [7, 11, 21].

Dzieje się tak dlatego, że klinicyści są generalnie bardziej zainteresowani określeniem prawdopodobieństwa obecności lub braku choroby, gdy wynik testu jest znany jako pozytywny lub negatywny. Podczas gdy czułość i specyficzność nie podają wprost tej informacji, wskaźnik prawdopodobieństwa określa o ile dany wynik testu zwiększy lub obniży prawdopodobieństwo zaistnienia zdarzenia, które ma przewidzieć ten test. Poprzez modelowanie komputerowe określić można rzeczywiste niezależne parametry, które następnie mogą być wykorzystywane w różnych kombinacjach i równocześnie do oceny istnienia procesu złośliwego. Wyniki badań opublikowanych ostatnio są zachęcające i sugerują wysoką przydatność prognostyczną zaawansowanych metod statystycznych w ocenie dostępnych danych. Idea wykorzystania modeli matematycznych w celu obliczenia indywidualnego prawdopodobieństwa istnienia złośliwego raka jajnika w guzach przydatkowych wydaje się być bardzo atrakcyjna [20]. Model matematyczny mógłby bowiem być alternatywą do prostego analizowania cech guza pochodzących z powszechnie dostępnego badania ultrasonograficznego w skali szarości i z badań dopplerowskich.

### 5.1 Analiza regresji logistycznej

W celu stworzenia algorytmu matematycznego, który pozwoliłby obliczyć indywidualne prawdopodobieństwo istnienia nowotworu złośliwego dla chorej z guzem jajnika w badanej grupie 307 pacjentek wykorzystano analizę regresji logistycznej. Jako narzędzie klasyfikacyjne ma ona szereg niewątpliwych zalet. Po pierwsze model bada możliwy związek rodzaju guza z licznymi, a nie tylko z jednym wskaźnikiem. Po drugie podczas konstrukcji modelu zostaną wybrane tylko najistotniejsze parametry prognostyczne uwzględnione w równaniu regresji. Po trzecie wskaźniki wybierane przez dany model jako istotne będą miały właściwe wyodrębnione współczynniki wagowe, które będą odzwierciedlać liczbowo udział każdego wskaźnika w procesie przewidywania istnienia procesu złośliwego [9]. Po czwarte jeśli dowolne zmienne ciągłe np. PI, RI są ujęte w równaniu regresji, to wtedy dla każdej chorej znaczenie tych wskaźników będzie uwzględnione w ostatecznym wyniku podającym ryzyko istnienia nowotworu złośliwego. Dokładność oceny

prawdopodobieństwa w modelu regresji logistycznej okazała się być większa niż dla oceny opartej wyłącznie na kryteriach morfologicznych lub dopplerowskich. Wartość tego modelu musi być jednak dalej testowana prospektywnie w nowej, znacznie większej grupie chorych. Istotne jest włączenie do modelu możliwie jak największej liczby chorych z wczesnymi stadiami klinicznymi raka jajnika [11, 19].

### 5.2 Sieci neuronowe

Głównym celem przedstawionej pracy była próba wykorzystania sztucznych sieci neuronowych w ocenie prawdopodobieństwa istnienia nowotworu złośliwego u kobiet z guzami jajnika. Jedną z największych zalet tej metody modelowania statystycznego w naukach biomedycznych jest zdolność do odnajdywania skomplikowanych zależności pomiędzy danymi klinicznymi i laboratoryjnymi. Powiązania te najczęściej są trudne do wykrycia, czasem nawet niemożliwe do znalezienia w inny sposób [1, 8].

W przedstawionej analizie własnych danych wykorzystano komputerowo generowane sieci neuronowe z których wybrano 5 o najwyższych wartościach prognostycznych. Wśród nich najlepsza okazała się być trójwarstwowa sieć perceptronowa o 13 neuronach w warstwie wejściowej i 11 w warstwie ukrytej. Przy prawdopodobieństwie większym niż 75% obliczona czułość i specyficzność modelu predykcyjnego wynosiły w zbiorze uczącym odpowiednio 96,77% i 100%. W zbiorach testowym i walidacyjnym wartości predykcyjne wynosiły 82,35% i 97,5%. Wartość pola powierzchni pod krzywą ROC dla tej sieci była najwyższa i wynosiła 0,9749.

### 5.3 Podsumowanie

Opisane wyniki badań własnych i przedstawionych w dostępnym piśmiennictwie wskazują, że jednym z przyszłych kierunków poprawy wyników wczesnej diagnostyki raka jajnika może być obok sonoangiografii i oznaczania markerów biochemicznych zastosowanie zaawansowanych metod nieliniowej estymacji. W modelu tym poszukiwane będą korelacje zestawu znanych parametrów klinicznych i relatywnie prostych badań ultrasonograficznych z istnieniem lub brakiem procesu złośliwego w guzie jajnika. Po zastosowaniu odpowiednich metod

„uczenia”, wynik obliczeń sieci można interpretować jako prawdopodobieństwo istnienia guza złośliwego jajnika na podstawie szeregu danych klinicznych, w tym ocenie ultrasonograficznej morfologii i unaczynienia guza w badaniu ultrasonograficznym z pulsacyjnym i kolorowym Dopplerem.

## 6 WNIOSKI

Analiza wyników przedstawionych w pracy pozwala na wyciągnięcie następujących wniosków:

- ◆ model prognostyczny zbudowany przy zastosowaniu sztucznej sieci neuronowej wykorzystującej dane kliniczne i ultrasonograficzne w celu przedoperacyjnego różnicowania guzów złośliwych i niezłośliwych jajnika u kobiet charakteryzuje się wyższą czułością i specyficznoscią niż stosowane pojedynczo testy diagnostyczne:
  - maksymalny wymiar, objętość i ocena morfologii guza
  - ocena przepływu naczyniowego w ultrasonografii z kolorowym Dopplerem
  - sonoangiografia i subiektywna ocena cechy „Kolor”
  - ocena stężenia markera CA-125
- ◆ prospektywna weryfikacja uzyskanego modelu predykcyjnego w nowej grupie chorych z guzami jajnika miała niższą czułość i specyficznosc niż model oceniany w grupie retrospektywnej. Sugeruje to konieczność zwiększenia liczby badanych kobiet z różnymi rodzajami guzów jajnika, tak by sieć mogła być jeszcze skuteczniejsza w diagnostyce różnicowej guzów przydatkowych
- ◆ wyniki badań wskazują, że korzystne może okazać się w przyszłości dołączenie do modelu predykcyjnego innych, nie badanych jeszcze parametrów. Wykorzystanie ultrasonografii trójwymiarowej, tomografii rezonansu magnetycznego lub badanie stężenia nowych markerów u chorych z guzami przydatkowymi może pozwolić na jeszcze bardziej precyzyjną diagnostykę różnicową nowotworów jajnika.

## BIBLIOGRAFIA

- 1) Biagiotti R, Desii C, Vanzi E, Gacci G. Predicting Ovarian Malignancy: Application of Artificial Neural Networks to Transvaginal and Color Doppler Flow US. *Radiology* 1999; 210: 399-403.
- 2) Boring CC, Squires TS i Tsong T. Cancer statistics, 1994. *CA Cancer J. Clin.* 1994; 44: 7-26.
- 3) Chock C, Irwig L, Berry G, Glasziou P. Comparing Dichotomous Screening Tests When Individuals Negative on Both Tests Are Not Verified. *J. Clin. Epidemiol.* 1997; 50: 1211-1217.
- 4) Czekierdowski A, Żrubek H, Sikorski M i wsp. Przydatność kliniczna oceny poziomu antygenu CA-125 i badania ultrasonograficznego z kolorowym Dopplerem w diagnostyce raka jajnika. *Gin Pol* 1996; 67: 553-56.
- 5) Czekierdowski A, Buczko-Świetlicka J, Czekierdowska D i Żrubek H. Diagnostyka różnicowa nowotworów jajnika przy zastosowaniu ultrasonografii z kolorowym Dopplerem i oznaczeń markerów CA 19.9 i CA-125. *Nowotwory* 1997; 47: 35-42.
- 6) Czekierdowski A. Badania nad angiogenezą w łagodnych i złośliwych guzach jajnika przy zastosowaniu ultrasonografii z kolorowym Dopplerem oraz oceny stężeń śródbłonkowego czynnika wzrostu i wybranych markerów nowotworowych: CA-125, CA-19.9, CA-72.4 i VEGF. Rozprawa habilitacyjna. AM w Lublinie 2000.
- 7) Czekierdowski A, Bednarek W, Rogowska W, Kotarski J. „Artificial neural networks in sonographic characterization of adnexal tumors: a preliminary study”. 11th Congress on Ultrasound in Obstetrics and Gynecology, Melbourne, 23-28.10.2001. *Suppl. Ultrasound Obstet Gynecol* 2001:CEG-13; 8.
- 8) Czekierdowski A, Smoleń A, Jędrych M. Sztuczne sieci neuronowe: nowe możliwości zastosowania zaawansowanych metod analizy statystycznej w medycynie perinatalnej i ginekologii. *Kliniczna Perinatologia i Ginekologia* 2002; 36:47-58.
- 9) Gore SM i Altman DG. „Metody oceny - wiele zmiennych” W: „Statystyka w praktyce lekarskiej” Wyd. Naukowe PWN Warszawa 1997: s.95-99.
- 10) Hanley JA i McNeil B. A method of comparing of the areas under the receiver operating characteristic curve derived from the same cases. *Radiology* 1983; 148: 839-43.
- 11) Smoleń A, Czekierdowski A, Daniłó J, Kraczkowski J. Sonoangiografia i zastosowanie logistycznej analizy regresji w przedoperacyjnej charakterystyce guzów jajnika. *Gin. Pol.* 2002; 11:1053-60.
- 12) Smoleń A, Czekierdowski A, Bednarek W, Kotarski J. Zastosowanie analizy krzywych ROC w przedoperacyjnym różnicowaniu guzów przydatkowych u kobiet. *Annales* 2002 (w druku).
- 13) Sadowski K. Badania ultrasonograficzne i ocena CA-125 we wczesnym wykrywaniu i diagnozie raka jajnika. *Klin Gin Perinatol* 2001, *Supl.* 22:54-58.
- 14) Spaczyński M. Red. Onkologia ginekologiczna. Wyd. Med. Urban&Partner, Wrocław 1997:208-20.
- 15) Spaczyński M, Spaczyński R. Rak jajnika - profilaktyka, diagnostyka, wczesne wykrywanie. *Gin Prakt* 2000; 10: 9-15.
- 16) Szperek D, Sajdak S, Obrębowska A i Spaczyński M. Wartość prognostyczna ultrasonografii dopochwowej z kolorowym Dopplerem w różnicowaniu charakteru złośliwości guzów jajnika. *Gin.Pol.* 1995; 66: 97-102.
- 17) Tadeusiewicz R, Lula P. Sieci neuronowe, Wyd. StatSoft Kraków 2001.
- 18) Tadeusiewicz R. Elementarne wprowadzenie do techniki sieci neuronowych z przykładowymi programami

mi. Akademicka Oficyna Wydawnicza PLJ Warszawa 1999: s. 1-51.

- 19) Tailor A, Jurkovic D, Bourne TH, Collins WP i Campbell S. Sonographic prediction of malignancy in adnexal masses using multivariate logistic regression analysis. *Ultrasound Obstet Gynecol* 1997; 10: 41-47.
- 20) Timmerman D, Schwärzler P, Collins WP, Claerhout F, Coenen M, Amant F, Vergote I, Bourne TH. Subjective assessment of adnexal masses with the use of ultrasonography: an analysis of interobserver variability and experience. *Ultrasound Obstet Gynecol* 1999; 13: 11- 16.
- 21) Timmerman D, Verrels H, Bourne TH, De Moor B, Collins WP, Vergote I, Vandewalle J. Artificial neural network models for the preoperative discrimination between malignant and benign adnexal masses. *Ultrasound Obstet Gynecol* 1999; 13: 17- 25.
- 22) Valentin L, Sladkievicius P, Marsal K. Limited contribution of Doppler velocimetry to the differential diagnosis of extrauterine pelvic tumors. *Obstet Gynecol* 1994; 83: 425-33.
- 23) VanNagell JR, Gallion HH, Pavlik EJ. Ovarian Cancer screening. *Cancer* 1995; 76:2086-91.
- 24) Zatoński W i Tyczyński J. Red. Nowotwory złośliwe w Polsce w roku 1993. Krajowy Rejestr Nowotworów, Centrum Onkologii Warszawa 1996; 21:23.

