
**PODSTAWY INŻYNIERII DIAGNOSTYKI OBRAZOWEJ
W MEDYCYNIE (PIDOM)**

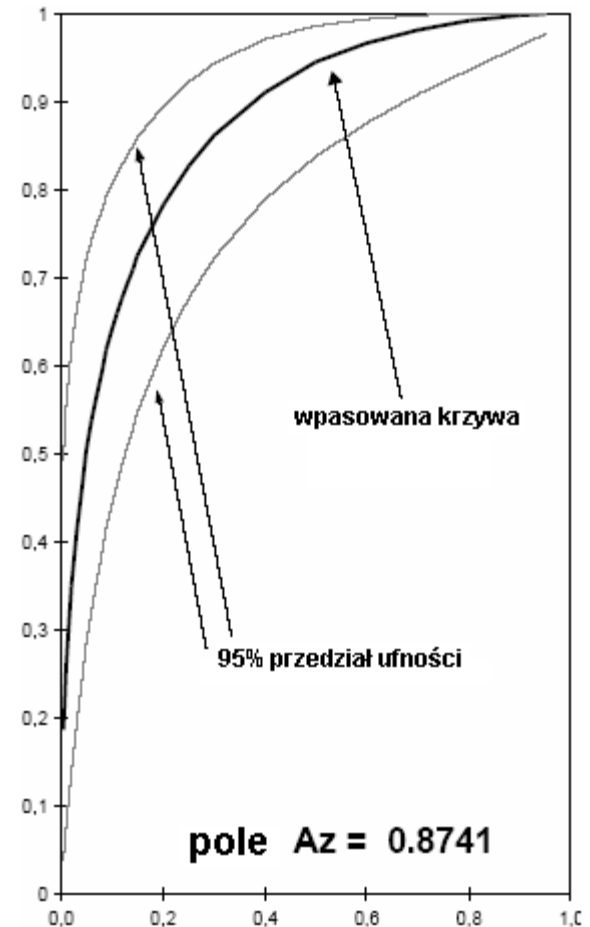
**WERYFIKACJA PROCEDUR
DIAGNOSTYCZNYCH**

Artur Przelaskowski
materiały do wykładu
(rysunki w dużej części ze źródeł internetowych oraz własnych)

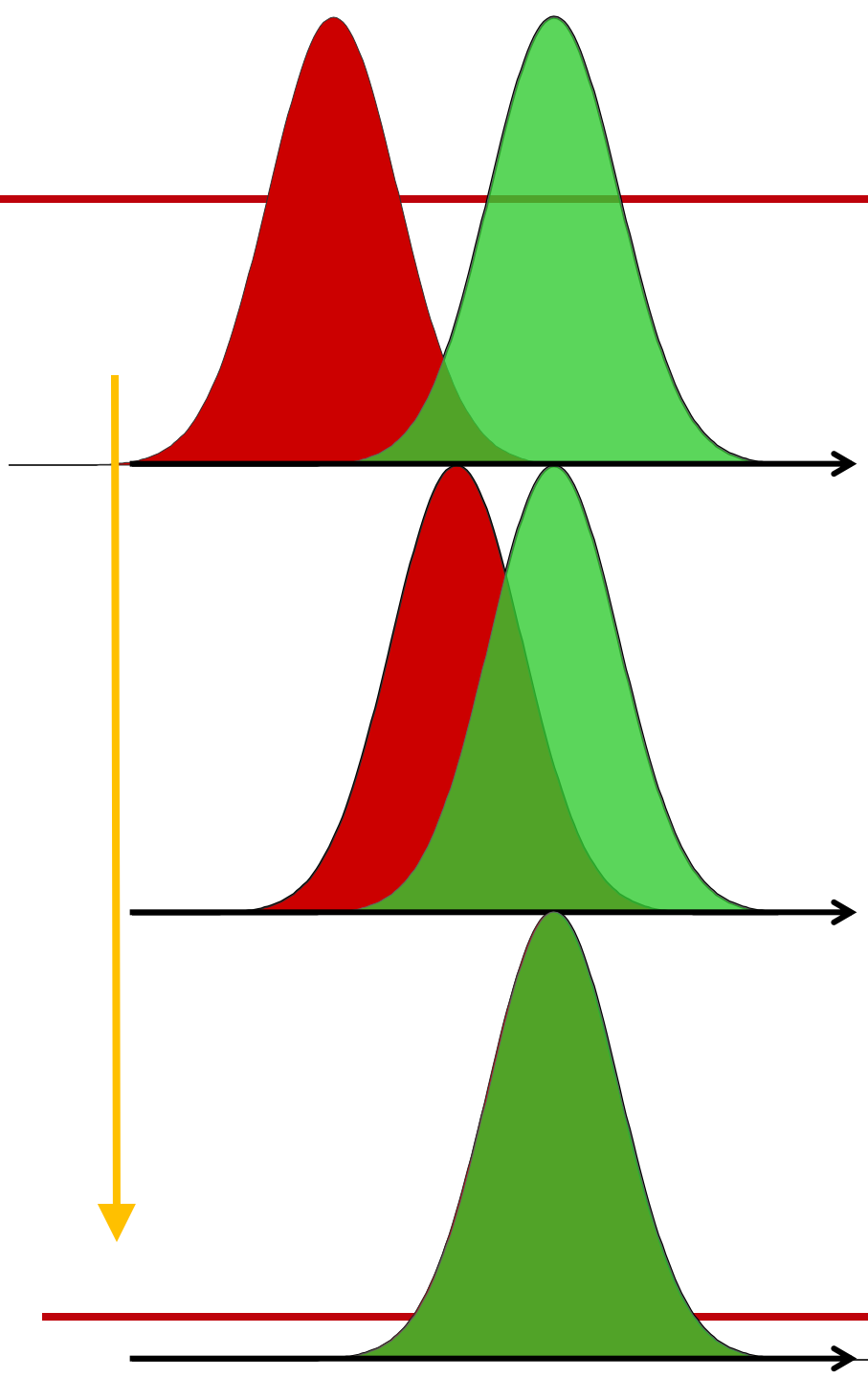
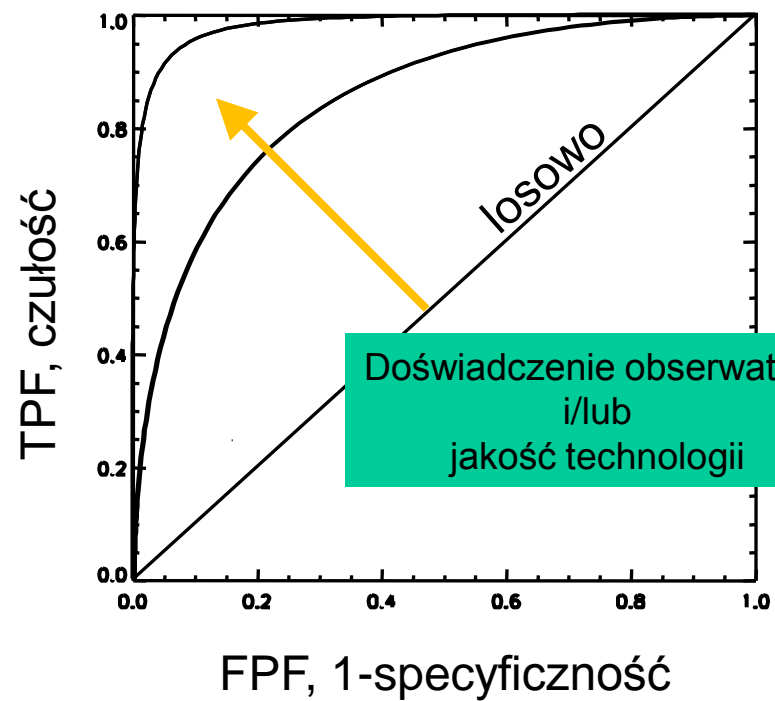
Test z krzywą ROC

Wymogi:

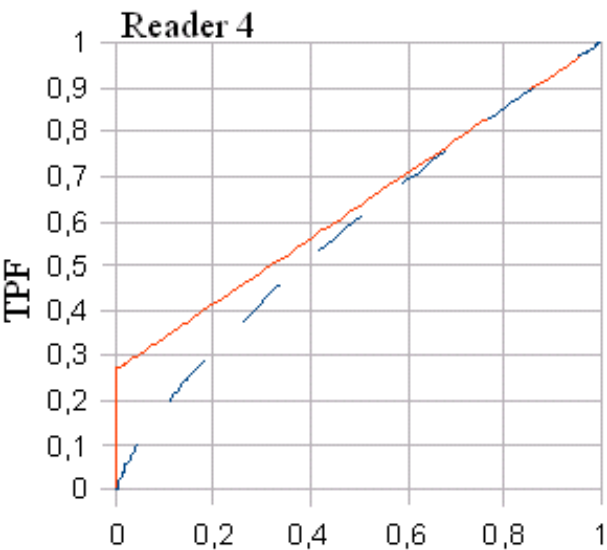
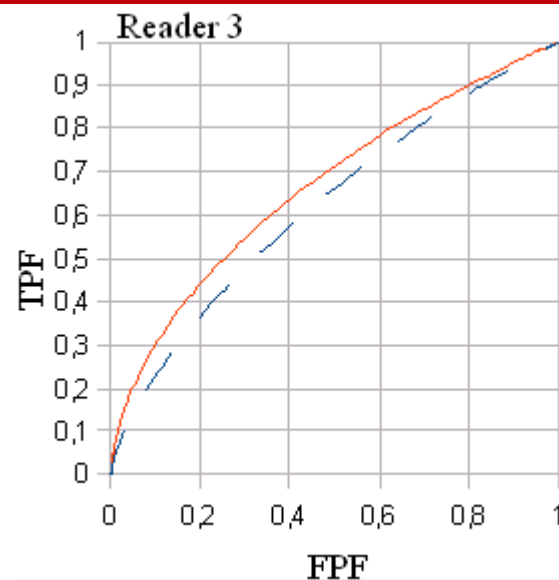
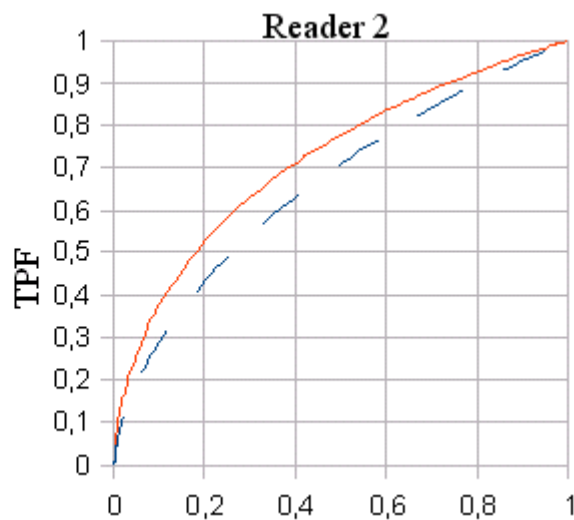
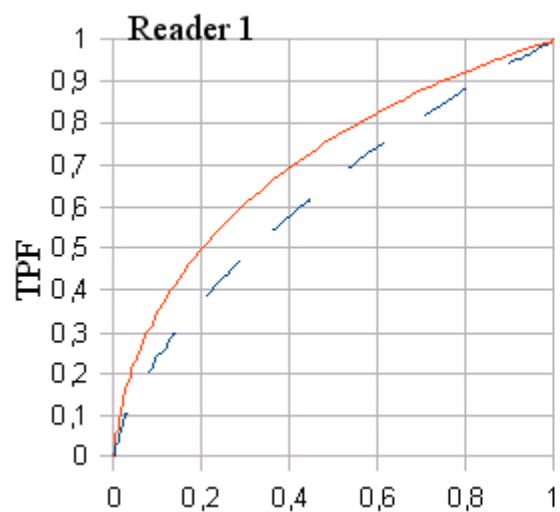
- dwie populacje danych: z patologią i bez
- opracowany 'złoty standard'
- wiarygodne uwarunkowania testu
- analiza różnicująca wyników testu: parametry badane to czułość, specyficzność, pole pod krzywą
- weryfikacja hipotez statystycznych: testy parametryczne 1W i 2W z poziomem istotności (jakościowe) lub z przedziałami ufności (ilościowe)
- odrzucanie hipotez albo brak podstaw do ich odrzucenia



krzywa ROC



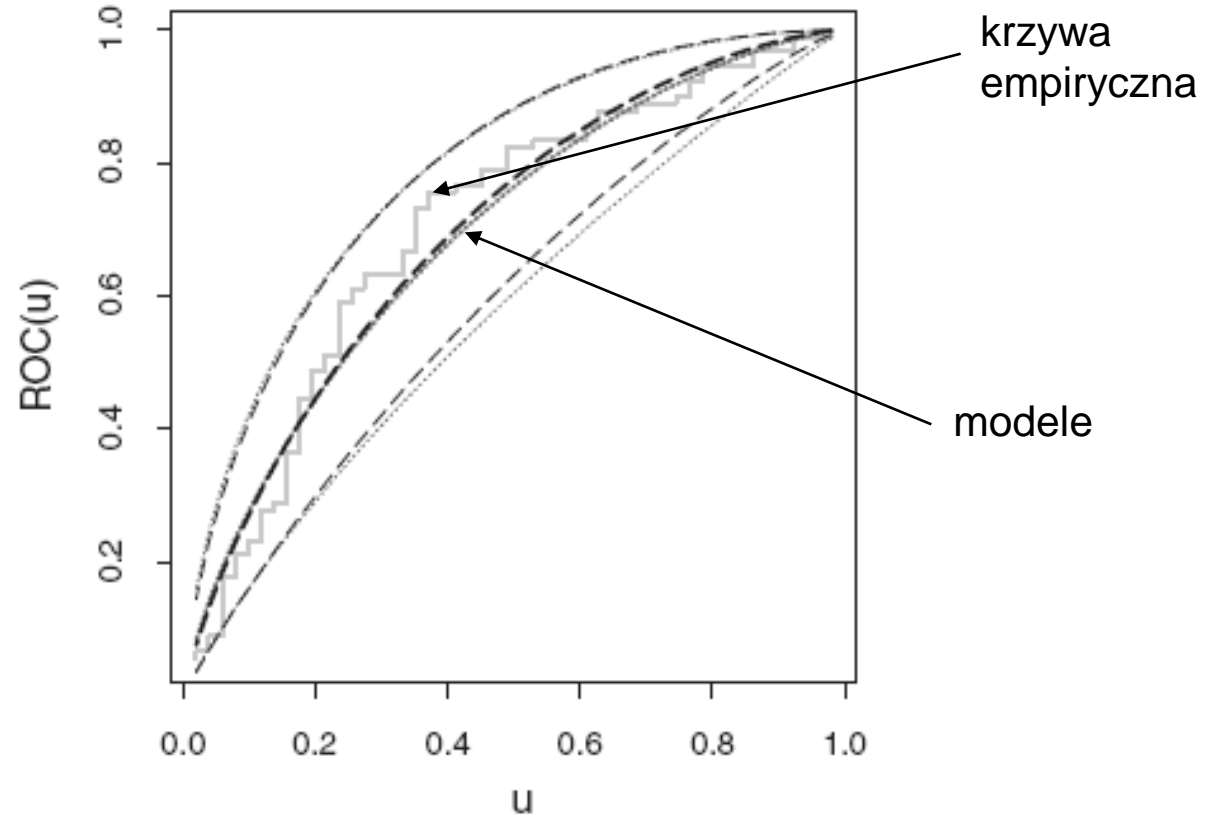
Przykładowa ocena



Reader	AUC		Sensitivity		Specificity		Accuracy		PVP	
	I	II	I	II	I	II	I	II	I	II
1	.633	.712	0.387	0.432	0.800	0.809	0.474	0.516	0.879	0.889
2	.666	.721	0.440	0.533	0.800	0.800	0.516	0.589	0.892	0.909
3	.625	0.693	0.533	0.773	0.650	0.650	0.558	0.747	0.851	0.892
4	.596	.640	0.200	0.280	0.900	1.000	0.347	0.432	0.882	1.000

Binormalny model krzywej ROC

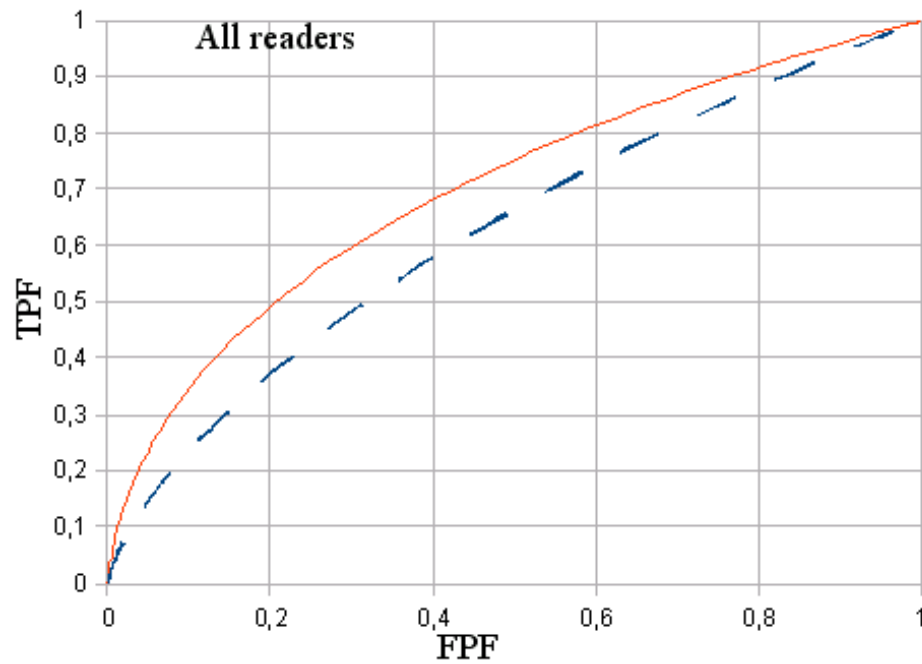
CBM - Contaminated Binormal Model



$$ROC(u) = \Phi \left\{ a + b\Phi^{-1}(u) \right\}$$

Metoda: DBM (Dorfman-Berbaum-Metz) MRMC (*Multi Reader Multi Case*)

Przykładowa ocena



*poziom
istotności*

DBM MRMC analysis	AUC		p-value	
	I	II	CBM	TRAP
All test cases (95)	.630	.686	.0026	.0074
Cases without movement artifacts (85)	.615	.677	.0011	.0086
Cases without significant asymmetry (89)	.639	.687	.0067	.0118
Cases without scarring (59)	.653	.739	.1502	.0005

Przykład realizacji testu udarowego

Liczebność przypadków w grupach testowych

Grupa	ogółem	kobiety	mężczyźni	% kobiet	% mężczyzn	śr. wiek [lat]	odch. std. [lat]
Wszyscy	123	56	67	45,5	54,5	69,9	14,30
Wzorcowa	92	38	54	41,3	58,7	68,9	14,37
Czas ICT	103	46	57	44,6	55,4	69,7	14,28

- **wzorcowa**, spełniająca oba warunki – I CT < 6h od wystąpienia objawów neurologicznych i II CT < 7 dni (84 przypadki)
- **I CT** – pacjentów, u których pierwsze badanie przeprowadzono w fazie nadostrej udaru, a badanie kontrolne przeprowadzono w trakcie tej samej hospitalizacji (dodano przypadki badań kontrolnych wykonanych po czasie dłuższym niż 7 dni od pierwszego badania) spełniająca pierwszy warunek – I CT < 6h od objawów (czas I CT) – 96 przypadków

Formularz oceny tradycyjnej

Protokół oceny badań wyjściowych

Inicjały:				
Specyfikacja sprzętu (rodzaj monitora/konsoli, itp.):				

Osoba nadzorująca:	
Data i godzina rozpoczęcia:	
Data i godzina zakończenia:	

Legenda

Format określania lokalizacji (lokalizacja warstw + położenia (godzinowo))

Lokalizację określa się poprzez podanie numeru warstwy oraz położenia obszaru hipodensyjnego na tej warstwie przyjmując analogię do tarczy zegarowej
przykładowo: 4-6,7 oznacza lokalizację w warstwach 4-6 na godzinie 7

Określanie obecności udaru

skala ocen 1 - 5 gdzie: 1 - definitywny brak udaru; 2 - raczej nie ma udaru; 3 - możliwy udar; 4 - raczej jest udar; 5 - zdecydowany udar

Dane identyfikujące badanie			Ocena widoczności udaru i przetwarzania					Uwagi
nr badania	nazwa	wiek	obecność udaru (skala 1-5)	lokalizacja udaru (lokalizacja warstw + położenia (godzinowo))	pewność decyzji (zakres 0-100%)			
3	case3	49						
4	case4	82						
7	case7	55						
12	case12	68						

Formularz oceny wspomaganej

Protokół badań (Stroke Monitor)

Inicjały:	
Specyfikacja sprzętu (rodzaj monitora/konsoli, itp.):	
Osoba nadzorująca:	
Data i godzina rozpoczęcia:	
Data i godzina zakończenia:	

Legenda

Format określania lokalizacji (lokalizacja warstw + położenia (godzinowo))

Lokalizację określa się poprzez podanie numeru warstwy oraz położenia obszaru hipodensyjnego na tej warstwie przyjmując analogię do tarczy zegarowej
przykładowo: 4-6,7 oznacza lokalizację w warstwach 4-6 na godzinie 7

Określanie obecności udaru

skala ocen 1 - 5 gdzie: 1 - definitywny brak udaru; 2 - raczej nie ma udaru; 3 - możliwy udar; 4 - raczej jest udar; 5 - zdecydowany udar

Ocena segmentacji struktur nadmiarowych (bruzdy, blizny, itp.)

skala ocen 1 - 3 gdzie: 1 - segmentacja słaba; 2 - segmentacja dobra; 3 - segmentacja bardzo dobra

Preferowana wizualizacja (formy prezentacji przetworzonych obrazów)

najlepsza forma wizualizacji efektów wspomagania (zgodnie z numeracją na ekranie monitora): 2 - segm; 3 - MU-SE; 4 - MU-DE; 5 - MU-PP; 6 - MU-DD

Dane identyfikujące badanie			Ocena widoczności udaru i przetwarzania					Uwagi
nr badania	nazwa	wiek	obecność udaru (skala 1-5)	lokalizacja udaru (lokalizacja warstw + położenia (godzinowo))	pewność decyzji (zakres 0-100%)	ocena segmentacji (skala 1-3)	preferowana wizualizacja (2,3,4,5,6)	uwagi
3	case3	49						
4	case4	82						
7	case7	55						
12	case12	68						

Test z porównaniem wartości średniej - założenia

prosty test ze statystyką U

Mamy: dwie duże, niezależne próby pobrane z populacji niekoniecznie normalnych, o nieznanymi wartościami średnich m_1 i m_2 i o nieznanymi, lecz równymi wariancjach σ_1^2 i σ_2^2 . Hipoteza: $H_0: m_1 = m_2$. Statystyka:

$$U = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}},$$

której rozkład przy prawdziwości hipotezy H_0 jest asymptotycznie normalny $N(0, 1)$,

n_1, n_2 - liczebności prób.

Weryfikacja hipotezy

Reguła postępowania jest następująca:

- przybliżamy wartości średnich i wariancji poprzez $\bar{x}_1, \bar{x}_2, s_1^2, s_2^2$,
- ustalamy poziom istotności α
- rozpatrujemy, którą z hipotez alternatywnych należy wziąć pod uwagę:

$$H_1: m_1 \neq m_2, \quad H_2: m_1 > m_2, \quad H_3: m_1 < m_2$$

- jeśli wybieramy H_1 , to stosujemy test dwustronny i **odrzucaamy hipotezę** H_0 na korzyść hipotezy H_1 , gdy dla obliczonej wartości

spełniona jest nierówność

$$|u_0| > u_\alpha,$$

$$u_0 = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

gdzie u_α jest wartością statystyki U wyznaczoną z tablicy rozkładu normalnego, dla

$$\text{której } P(|U| \geq u_\alpha) = \alpha.$$

Jeśli hipotezą alternatywną względem hipotezy H_0 jest hipoteza H_2 to stosujemy test prawostronny i odrzucaamy H_0 na korzyść H_2 jeśli $u_0 > u_\alpha$ itd.

Rozkład normalny dla wartości poziomu istotności

Wartości u_α , dla których $P(|U| > u_\alpha) = 1 - \frac{2}{\sqrt{2\pi}} \int_0^{u_\alpha} e^{-\frac{1}{2}x^2} dx$

α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,575 829	2,326 348	2,170 090	2,053 749	1,959 964	1,880 794	1,811 911	1,750 686	1,695 398
0,1	1,644 854	1,598 193	1,554 774	1,514 102	1,475 791	1,439 521	1,405 072	1,372 204	1,340 755	1,310 579
0,2	1,281 552	1,253 565	1,226 528	1,200 359	1,174 987	1,150 349	1,126 391	1,103 063	1,080 319	1,058 122
0,3	1,036 433	1,015 222	0,994 458	0,974 114	0,954 165	0,934 589	0,915 365	0,896 473	0,877 896	0,859 617
0,4	0,841 621	0,823 894	,806 421	,789 192	,772 193	,755 415	,738 847	,722 479	,706 303	,690 309
0,5	,674 490	,658 838	,643 345	,628 006	,612 813	,597 760	,582 841	,568 051	,553 385	,538 836
0,6	,524 401	,510 073	,495 850	,481 727	,467 699	,453 762	,439 913	,426 148	,412 463	,398 855
0,7	,385 320	,371 856	,358 459	,345 125	,331 853	,318 639	,305 481	,292 375	,279 319	,266 311
0,8	,253 347	,240 426	,227 545	,214 702	,201 893	,189 118	,176 374	,163 658	,150 969	,138 304
0,9	,125 661	,113 039	,100 434	,087 845	,075 270	,062 707	,050 154	,037 608	,025 069	,012 533

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
u	3,290 53	3,890 59	4,417 17	4,891 64	5,326 72	5,730 73	6,109 41

Testy o innych założeniach

t-Studenta

W tych zastosowaniach można użyć następującego testu:

weryfikujemy hipotezę, że średnie dwóch małych prób nie różnią się istotnie, czyli

$H_0: m_1 = m_2$. Przyjmujemy, że próby są niezależne, a populacje w przybliżeniu

normalne o nieznanym, lecz równym wariancjach (niestety niezbyt spełnione w tym

przypadku). Podstawą testu jest statystyka:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Ma ona rozkład t-Studenta o $n_1 + n_2 - 2$ stopniach swobody

Można też stosować inne testy, np. F-Snedecora dotyczący weryfikacji hipotezy o równości wariancji empirycznych rozkładów, itp.

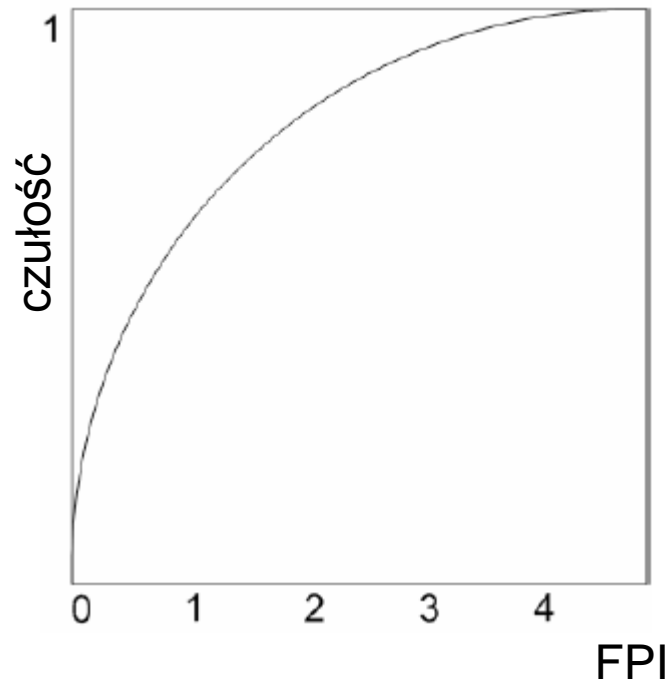
Rozkład t StudentaWartości t_α , dla których $P(|t| > t_\alpha) = \alpha$

$r \backslash a$	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,001
1	0,158	0,325	0,510	0,727	1,000	1,376	1,963	3,078	6,314	12,706	31,821	63,657	636,619
2	,142	,289	,445	,617	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	31,598
3	,137	,277	,424	,584	,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	12,941
4	,134	,271	,414	,569	,741	,941	1,190	1,533	2,132	2,776	3,747	4,604	8,610
5	,132	,267	,408	,559	,727	,920	1,156	1,476	2,015	2,571	3,365	4,032	6,859
6	,131	,265	,404	,553	,718	,906	1,134	1,440	1,943	2,447	3,143	3,707	5,959
7	,130	,263	,402	,549	,711	,896	1,119	1,415	1,895	2,365	2,998	3,499	5,405
8	,130	,262	,399	,546	,706	,889	1,108	1,397	1,860	2,306	2,896	3,355	5,041
9	,129	,261	,398	,543	,703	,883	1,100	1,383	1,833	2,262	2,821	3,250	4,781
10	,129	,260	,397	,542	,700	,879	1,093	1,372	1,812	2,228	2,764	3,169	4,587
11	,129	,260	,396	,540	,697	,876	1,088	1,363	1,796	2,201	2,718	3,106	4,437
12	,128	,259	,395	,539	,695	,873	1,083	1,356	1,782	2,179	2,681	3,055	4,318
13	,128	,259	,394	,538	,694	,870	1,079	1,350	1,771	2,160	2,650	3,012	4,221
14	,128	,258	,393	,537	,692	,868	1,076	1,345	1,761	2,145	2,624	2,977	4,140
15	,128	,258	,393	,536	,691	,866	1,074	1,341	1,753	2,131	2,602	2,947	4,073
16	,128	,258	,392	,535	,690	,865	1,071	1,337	1,746	2,120	2,583	2,921	4,015
17	,128	,257	,392	,534	,689	,863	1,069	1,333	1,740	2,110	2,567	2,898	3,965
18	,127	,257	,392	,534	,688	,862	1,067	1,330	1,734	2,101	2,552	2,878	3,922
19	,127	,257	,391	,533	,688	,861	1,066	1,328	1,729	2,093	2,539	2,861	3,883
20	,127	,257	,391	,533	,687	,860	1,064	1,325	1,725	2,086	2,528	2,845	3,850
21	,127	,257	,391	,532	,686	,859	1,063	1,323	1,721	2,080	2,518	2,831	3,819
22	,127	,256	,390	,532	,686	,858	1,061	1,321	1,717	2,074	2,508	2,819	3,792
23	,127	,256	,390	,532	,685	,858	1,060	1,319	1,714	2,069	2,500	2,807	3,767
24	,127	,256	,390	,531	,685	,857	1,059	1,318	1,711	2,064	2,492	2,797	3,745
25	,127	,256	,390	,531	,684	,856	1,058	1,316	1,708	2,060	2,485	2,787	3,725
26	,127	,256	,390	,531	,684	,856	1,058	1,315	1,706	2,056	2,479	2,779	3,707

WERYFIKACJA PROCEDUR

Ocena efektywności detekcji

- Czułość = TP (z poprawną lokalizacją)/liczba obiektów (patologii, zmian)
- FPI (*false positives per image*) = FP/liczba obrazów
- Krzywa FROC (*free-response receiver operating characteristic*)



- Krzywa AFROC (alternative FROC) z trafnością = FP/liczba obrazów bez patologii

Przykład: ocena wydajności detekcji MammoViewera

Patologia dobrze
wykryta



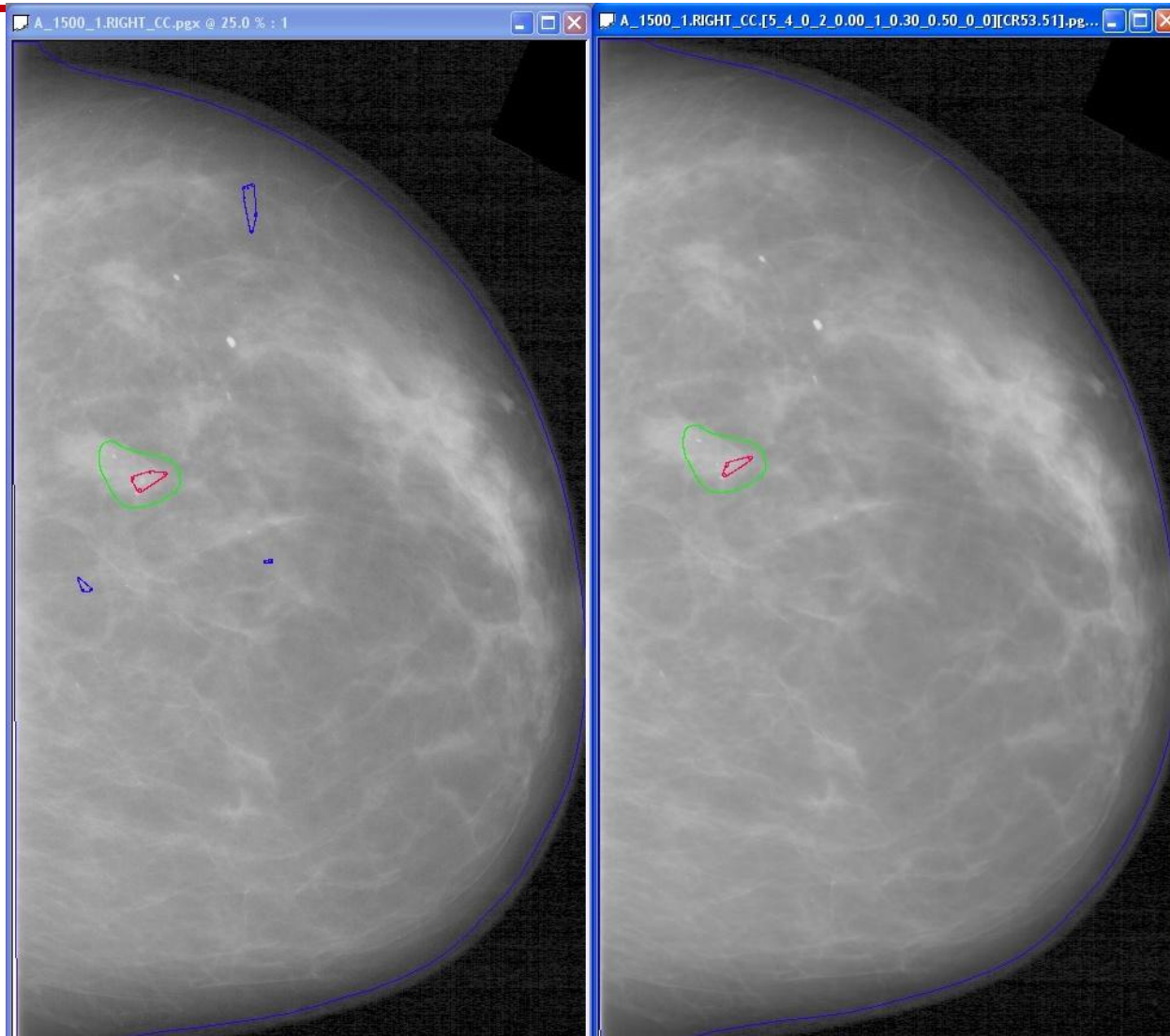
Czułość: około 90%

Średnia liczba fałszywych
wskazań na obraz to 1.6



Fałszywe
wskazania

Poprawa wskazań detekcji

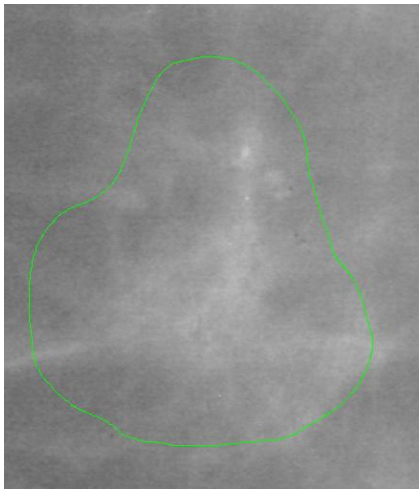
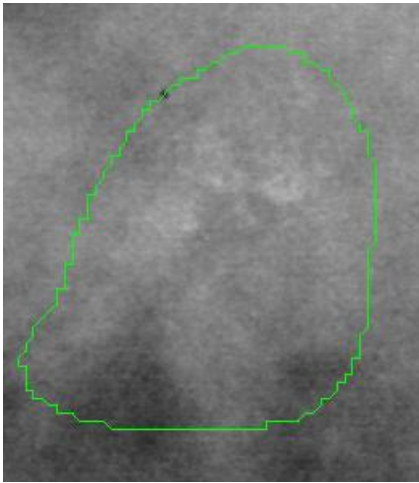


oryginał

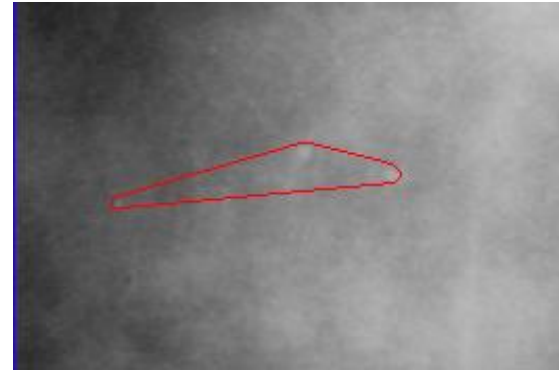
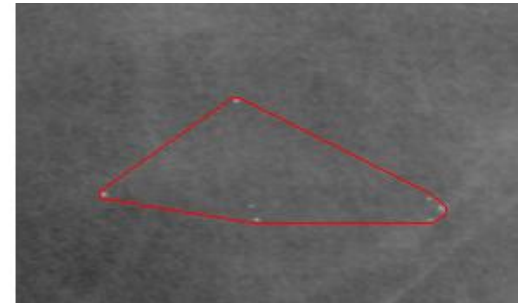
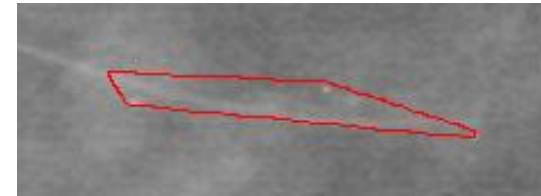
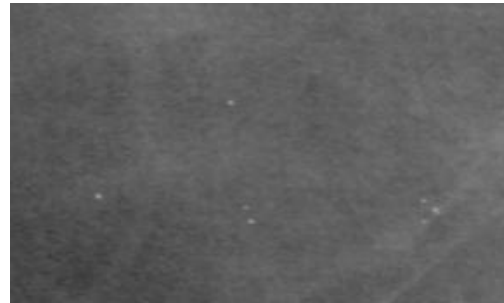
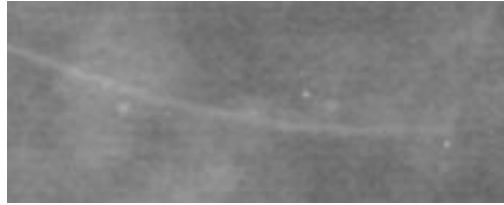
efekt przetwarzania

Błędy wskazań

Fałszywie
ujemne:

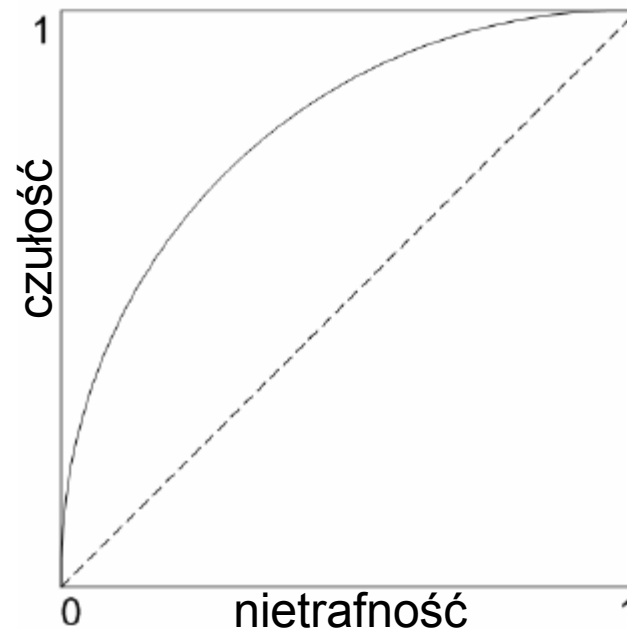


Fałszywie dodatnie:

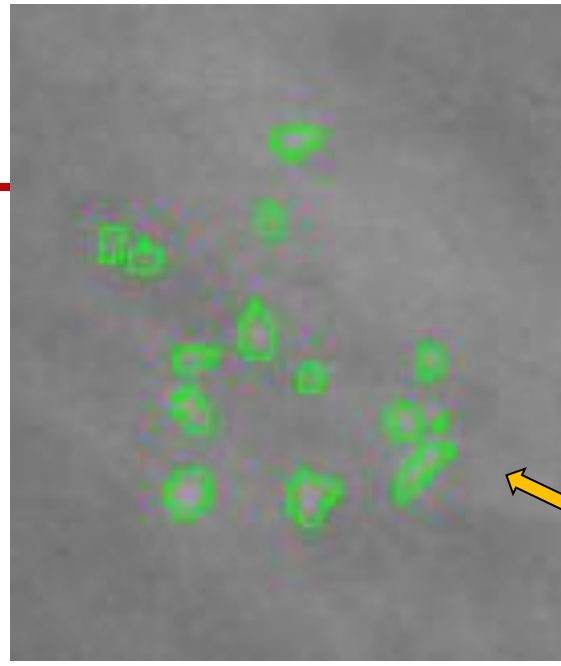
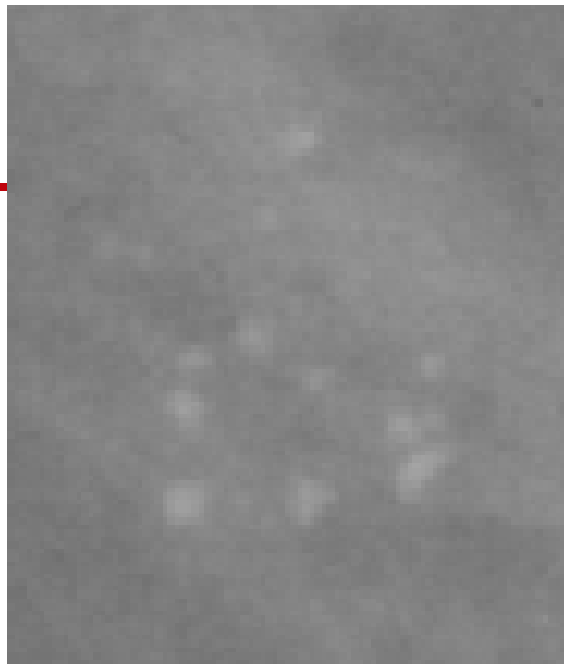


Ocena efektywności diagnozy – dotyczy zmian łagodnych i złośliwych (LROC)

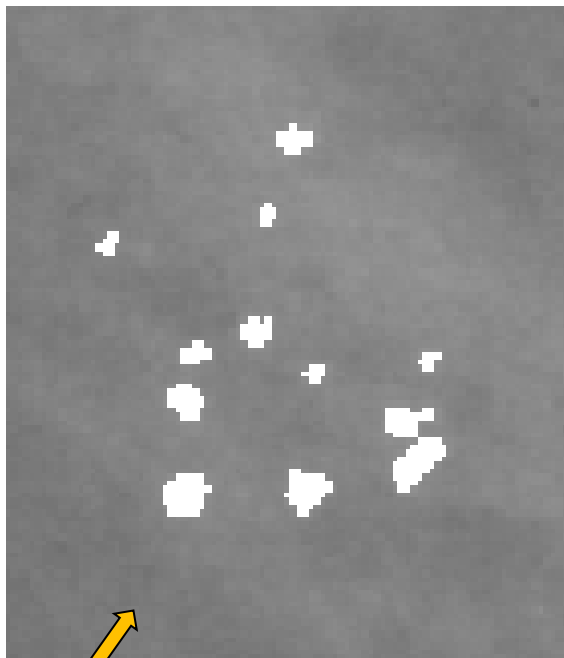
- Czułość = TP (z poprawną lokalizacją)/liczba zmian złośliwych
- Nietrafność = FP/liczba zmian łagodnych
- Krzywa LROC (*localization receiver operating characteristic*)



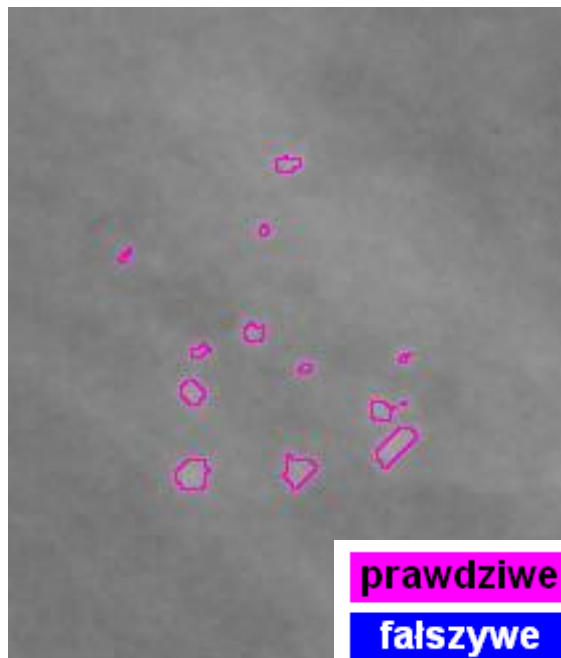
Weryfikacja detekcji i diagnoza pojedynczych mikrozwapnień



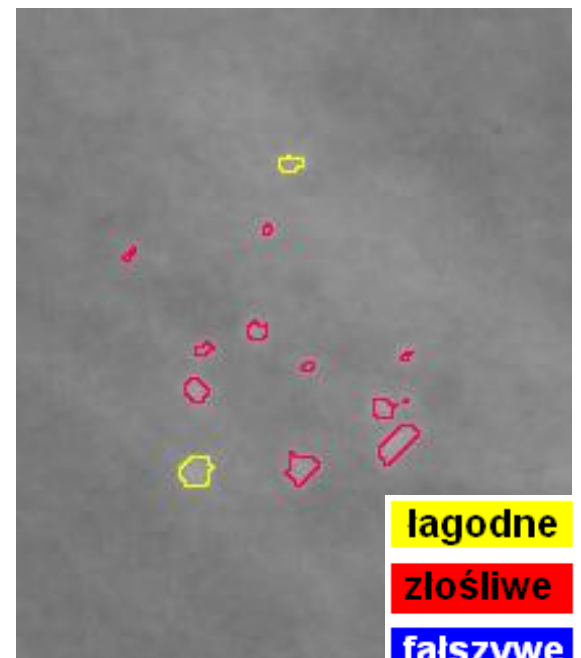
obrysy lekarza



automatyczna detekcja



prawdziwe
fałszywe



łagodne
złośliwe
fałszywe

klasyfikacja w kategoriach diagnostycznych

OGRANICZENIA ROC

Ograniczenia ROC

- konieczność zamiany normalnego trybu diagnozowania w praktyce klinicznej na wyrażenie opinii w pewnej skali ocen.
 - ponieważ technika ROC została stworzona przy założeniu Gaussowskiego rozkładu szumów w zbiorze analizowanych danych, jej stosowanie do oceny danych obrazowych o zazwyczaj nie-Gaussowskim charakterze nasuwa pewne wątpliwości (istnieją pewne metody redukcji błędów wynikających z tych Gaussowskich założeń).
 - wiele praktycznych zadań diagnostycznych, jakie stoją przed specjalistami, nie sprowadza się do decyzji dwupoziomowej: tak lub nie; w niektórych patologiach występuje kilka nieprawidłowości w różnych fragmentach obrazu, a proces decyzyjny jest dużo bardziej złożony.
-

Modyfikacje ROC

- Specjaliści obserwują obrazy i zaznaczają obecność pewnych anomalii np. powiększonych węzłów chłonnych w obrazie CT klatki piersiowej, lub też obecność guzków w płucach, przy czym liczba anomalii jest różna w poszczególnych obrazach testowych.
- Warstwa decyzyjna zostaje rozszerzona na kilka, czy nawet kilkanaście poziomów.
- Analiza tak otrzymanych wyników przeprowadzana jest przy pomocy dwu parametrów: czułości i przewidywanej wartości pozytywnej PVP (*predictive value positive*) -
$$\frac{N_{pp}}{N_{pp} + N_{fp}}$$
.

- Czułość i PVP wykreśla się osobno – łącznie dla wszystkich testowanych grup badań
 - Porównanie czułości i PVP dla różnych stopni kompresji przeprowadzono za pomocą testu t-Studenta z wykorzystaniem rozkładu permutacji dwuelementowych (dla danych, które nie mają charakteru gaussowskiego) – test Behrensa-Fishera
-

Przykłady: czułość

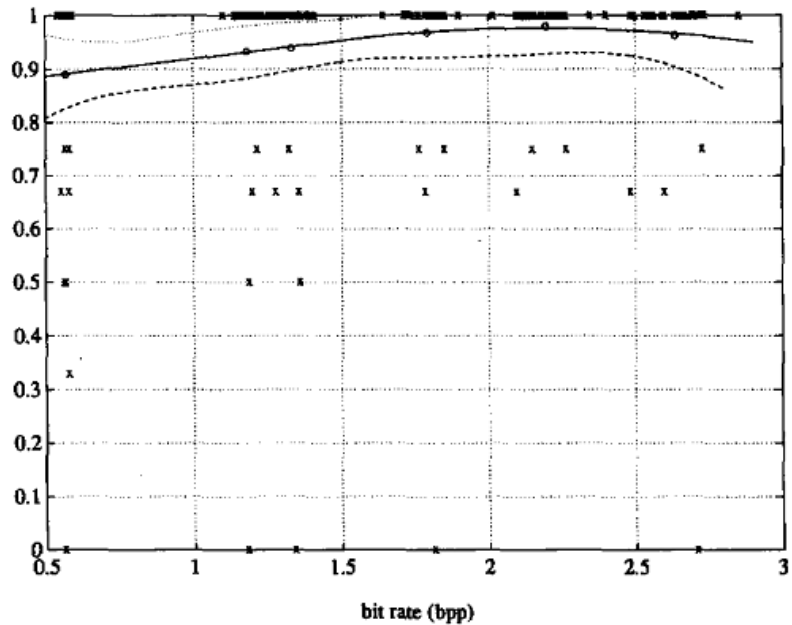


Fig. 8. Lung sensitivity: rms = 0.177.

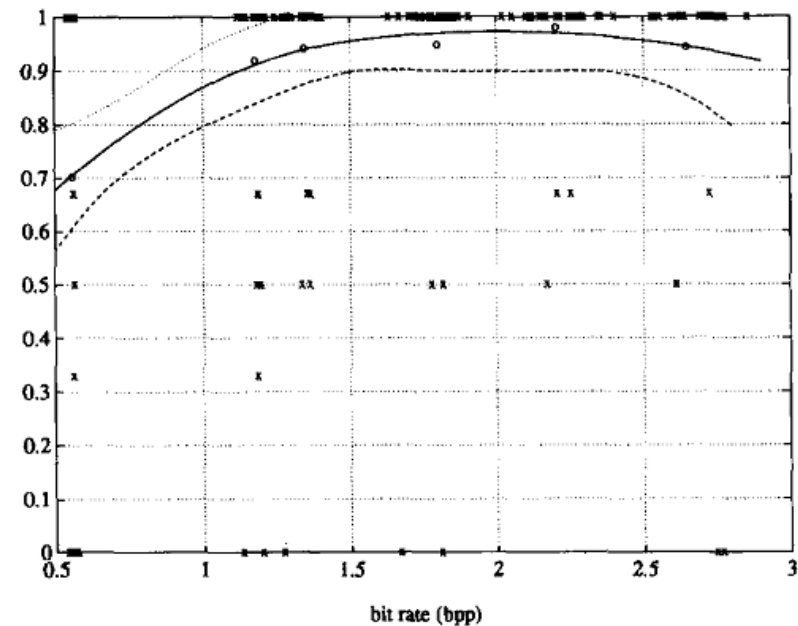


Fig. 10. Mediastinum sensitivity: rms = 0.243.

P. Cosman, R. Gray, R. Olshen Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy. PROCEEDINGS OF THE IEEE, VOL. 82, NO. 6, JUNE 1994

Przykłady: PVP

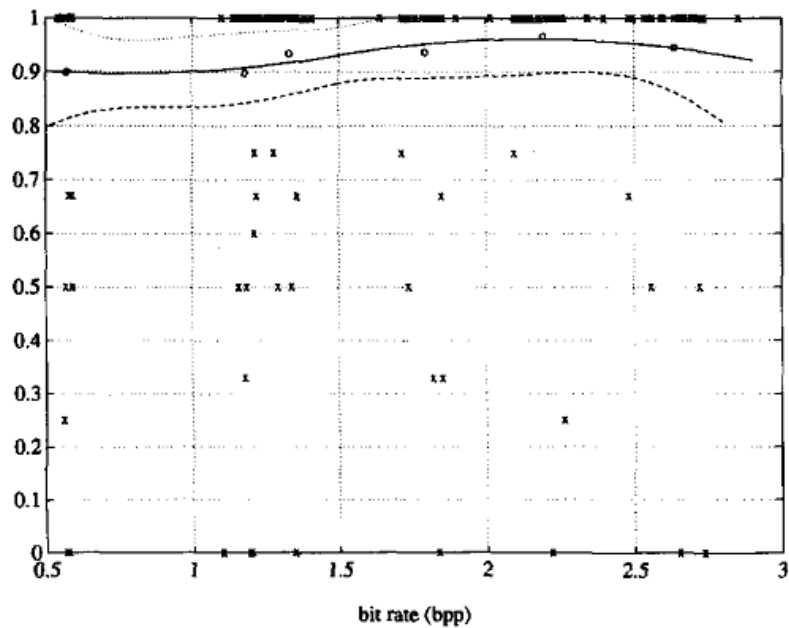


Fig. 9. Lung PVP: rms = 0.215.

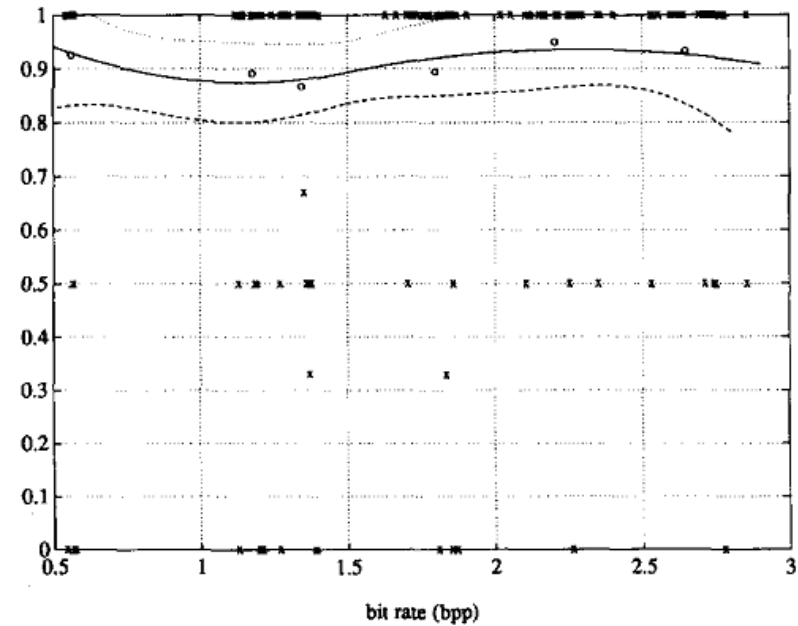


Fig. 11. Mediastinum PVP: rms = 0.245.

P. Cosman, R. Gray, R. Olshen Evaluating Quality of Compressed Medical Images: SNR, Subjective Rating, and Diagnostic Accuracy. PROCEEDINGS OF THE IEEE, VOL. 82, NO. 6, JUNE 1994

Test B-F

- specjalista ocenia N obrazów należących do dwóch poziomów **A** i **B**
- obrazy te należą do dziewięciu grup: bez patologii, z jedną patologią, z dwoma patologiami, ..., z ośmioma patologiami; przez oznaczymy liczbę obrazów i -tej grupy, a niech reprezentuje różnicę wartości czułości (lub PVP) dla j -tego obrazu i -tej grupy oglądanego na dwóch poziomach jakości
- niech $\bar{\Delta}_i$ będzie średnią różnicą wartości czułości (lub PVP) i -tej grupy

$$\bar{\Delta}_i = \frac{1}{N_i} \sum_j \Delta_{i,j}$$

- wariancję zmian wartości prób dla i -tej grupy definiujemy jako

$$s_i^2 = \frac{1}{N_i - 1} \sum_j (\Delta_{i,j} - \bar{\Delta}_i)^2$$

- statystyka t Behrensa-Fishera

$$t_{BF} = \frac{\sum_i \bar{\Delta}_i}{\sqrt{\sum_i \frac{s_i^2}{N_i}}}$$

- poziom istotności (k - liczba wartości, które przekraczają wartość 'klasyczną', przy inaczniej ustalonym składzie obu grup i wymieszaniu obrazów, z różnych poziomów w każdej grupie testowej)

$$\alpha = \frac{(k+1)}{2^N}$$

Metod bez ROC

- określenie złotego standardu: zgodny, osobisty, niezależny i osobny
- obiektywna diagnoza powstaje na podstawie analizę obrazów analogowych, względem których ocenia się obrazy cyfrowe (także cyfrowy oryginał)
- nowy protokół oceny jakości obrazów, w którym lekarze wyrażają swe opinie w kategoriach jak najbardziej diagnostycznych
- analiza statystyczna bez krzywych ROC:

do zapisu wyników testu wykorzystano tablicę 2×2 postaci

III	dobrze	źle
dobrze	N(1,1)	N(1,2)
źle	N(2,1)	N(2,2)

gdzie I - może oznaczać obraz oryginalny analogowy a II - oryginał cyfrowy lub też I oryginał cyfrowy, II - cyfrowy kompresowany. Decyzje **dobrze** znaczą zgodnie ze złotym standardem, a **źle** - niezgodnie.

Przykładowe oceny

Betts B.J., Li J., Cosman P.C., Gray R.M. et al.: Image quality in digital mammography. Revision of final report to the Army Medical Research and Materiel Command, Compression and Classification of Digital Mammograms for Storage, Transmission, and Computer Aided Screening, September 1998.

II \ I	R	W
R	7	2
W	1	2

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	6	4
W	3	5

C/B

II \ I	R	W
R	14	2
W	2	8

BX

(A) Analog vs. Digital Original

II \ I	R	W
R	6	3
W	1	2

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	8	2
W	2	6

C/B

II \ I	R	W
R	14	2
W	1	9

BX

(B) Analog vs. Digital Lossy Compressed: 1.75 bpp

II \ I	R	W
R	6	3
W	0	3

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	6	4
W	2	6

C/B

II \ I	R	W
R	12	3
W	4	6

BX

(C) Analog vs. Digital Lossy Compressed: 0.4 bpp

II \ I	R	W
R	4	4
W	0	3

RTS

II \ I	R	W
R	0	0
W	0	1

F/U

II \ I	R	W
R	3	7
W	4	4

C/B

II \ I	R	W
R	11	4
W	4	6

BX

(D) Analog vs. Digital Lossy Compressed: 0.15 bpp

Table 9: Agreement 2×2 tables for radiologist A.

RTS - przypadkowy, negatywny, lub łagodny do powtórnego badania.

F/U - prawdopodobnie łagodny ale wymagający sześciomiesięcznej obserwacji

C/B - potrzebne dodatkowe badania

BX - biopsja

Analiza statystyczna

Jeśli uzyskane tablice nie są diagonalne, to do oceny statystyki uzyskanych wyników stosuje się test McNemara:

porównanie $N(1,2)$ i $(N(1,2)+N(2,1))/2$ z pomocą rozkładu dwumianowego o parametrach $N(1,2)+N(2,1)$ i 0.5

$$P(N(1,2) = k \mid N(1,2) + N(2,1) = n) = \binom{n}{k} 2^{-n}; \quad k = 0, \dots, n$$

lub Fishera:

liczymy statystykę $(N(1,2) - N(2,1))^2 / (N(1,2) + N(2,1))$, która ma rozkład chi-kwadrat z jednym stopniem swobody

Dodatkowo tablice zgodności decyzji

	RTS	F/U	C/B	BX
RTS	12	0	5	0
F/U	0	0	0	0
C/B	3	0	12	6
BX	0	0	2	17

RTS - przypadkowy, negatywny, lub łagodny do powtórnego badania,

F/U - prawdopodobnie łagodny ale wymagający sześciomiesięcznej obserwacji

C/B - potrzebne dodatkowe badania

BX - biopsja

Przykładowe oceny

	RTS	F/U	C/B	BX
RTS	11	0	5	1
F/U	0	0	0	0
C/B	3	0	11	7
BX	2	0	2	15

	RTS	F/U	C/B	BX
RTS	4	0	0	0
F/U	0	0	0	1
C/B	3	0	3	3
BX	1	0	7	35

	RTS	F/U	C/B	BX
RTS	8	0	6	1
F/U	0	0	0	0
C/B	1	0	10	1
BX	0	0	7	23

A: Analog versus Digital

	RTS	F/U	C/B	BX
RTS	11	0	6	0
F/U	0	0	0	0
C/B	2	0	15	4
BX	1	0	2	16

	RTS	F/U	C/B	BX
RTS	2	1	0	1
F/U	0	1	0	0
C/B	3	1	3	2
BX	1	0	4	37

	RTS	F/U	C/B	BX
RTS	11	0	4	0
F/U	0	0	0	0
C/B	1	1	8	2
BX	1	0	5	24

B: Analog versus Lossy Compressed Digital: 1.75 bpp

	RTS	F/U	C/B	BX
RTS	9	0	6	2
F/U	0	0	0	0
C/B	1	0	10	10
BX	1	0	2	15

	RTS	F/U	C/B	BX
RTS	1	0	2	1
F/U	0	0	0	1
C/B	2	0	2	5
BX	2	0	5	36

	RTS	F/U	C/B	BX
RTS	7	0	7	1
F/U	0	0	0	0
C/B	2	0	8	2
BX	1	0	4	25

C: Analog versus Lossy Compressed Digital: 0.4 bpp

	RTS	F/U	C/B	BX
RTS	8	0	7	1
F/U	0	0	0	0
C/B	3	1	9	8
BX	1	0	6	11

	RTS	F/U	C/B	BX
RTS	3	1	0	0
F/U	0	1	0	0
C/B	3	0	3	2
BX	1	1	5	35

	RTS	F/U	C/B	BX
RTS	7	0	7	0
F/U	0	0	0	0
C/B	0	0	9	3
BX	0	0	9	20

D: Analog versus Lossy Compressed Digital: 0.15 bpp

Radiologist A

Radiologist B

Radiologist C

Table 10: Radiologist agreement tables, Stanford judges.

Betts B.J., Li J., Cosman P.C., Gray R.M. et al.: Image quality in digital mammography. Revision of final report to the Army Medical Research and Materiel Command, Compression and Classification of Digital Mammograms for Storage, Transmission, and Computer Aided Screening, September 1998.