

Teresa PODSIADŁY-MARCZYKOWSKA
Instytut Biocybernetyki i Inżynierii Biomedycznej PAN
Rafał ZAWIŚLAK
Politechnika Łódzka, Instytut Automatyki

OCENA NARZĘDZIA DO WSPOMAGANIA INTERPRETACJI MAMMOGRAMÓW OPARTEGO NA DZIEDZINOWEJ ONTOLOGII WIEDZY

Streszczenie. W artykule przedstawiono założenia, przebieg oraz wyniki testu, którego podstawowym celem była eksperymentalna weryfikacja hipotezy mówiącej, że interpretacja mammogramu przy użyciu edytora MammoEdit (narzędzia bazującego na własnej ontologii, zawierającego wbudowaną wiedzę z dziedziny mammografii) podnosi efektywność diagnostyczną radiologa. Ocenie poddany został również interfejs graficzny edytora i opracowane zasady wizualizacji, mającej na celu sygnalizowanie radiologowi wartości cech zmian istotnych dla właściwej interpretacji ich charakteru.

Słowa kluczowe: interpretacja mammogramu, inżynieria biomedyczna, ontologia mammografii

EXPERIMENTAL EVALUATION OF THE MAMMOGRAM INTERPRETATION AIDING TOOL SUPPORTED BY THE DOMAIN KNOWLEDGE ONTOLOGY

Summary. The article presents assumptions, design, and results of a test whose primary aim was the experimental verification of a hypothesis stating that the use of the MammoEdit editor – a tool that uses its own ontology and the embedded knowledge of mammography – increases the effectiveness of a radiologist diagnosis. The graphical user interface of the editor was similarly assessed, as well as the rules for visualization which assists the radiologist in the interpretation of the lesions' character.

Keywords: mammogram interpretation, biomedical engineering, mammographic ontology

1. Wstęp

Dokonując przeglądu stosowanych obecnie komputerowych metod wspomaganie diagnostyki raka sutka, można stwierdzić występowanie różnicy w stopniu zaawansowania komputerowych metod wspomaganie detekcji oraz interpretacji zmian mammograficznych. Metody detekcji doczekały się już rozwiązań na tyle dojrzałych, żeby zaowocować, przynajmniej dla pewnych typów zmian, zastosowaniami komercyjnymi, podczas gdy metody interpretacji tego etapu jeszcze nie osiągnęły.

Niniejszy artykuł prezentuje testy kliniczne narzędzia MammoEdit (ME), wypełniającego lukę w istniejących rozwiązaniach problemu wspomaganie interpretacji mammogramów, ściśle dopasowanego do potrzeb i wymagań radiologa.

Interpretację zmiany mammograficznej przez radiologa można przedstawić jako proces składający się z dwóch etapów. Wynikiem pierwszego etapu jest opis zmiany, wynikiem końcowym to przyporządkowanie zmianie odpowiedniej kategorii diagnostycznej. Opisane w literaturze systemy wspomaganie interpretacji mammogramów, typu CADx czy CAD-CBIR, końcową ocenę zmiany wspomagają w sposób pośredni, a stosowany w obu typach systemów opis cech zmian nie jest wyrażany w kategoriach semantycznych, zrozumiałych dla radiologa.

Podczas prac nad projektem edytora MammoEdit (ME), użyta została stworzona specjalnie do tego celu ontologia mammografii. Została ona wykorzystana jako zbiór częściowych wymagań projektowych dla interfejsu użytkownika i bazy danych przechowującej opisy badań. Taka rola ontologii mammografii w projekcie edytora ME była zgodna ze wskazaniem literaturowymi, dotyczącymi zastosowań ontologii w systemach informatycznych [1].

2. Interpretacja mammogramu

Mammografia jest powszechnie uznawana za najtrudniejszą w radiologii modalność, jeśli chodzi o interpretację diagnostyczną. Źródłem ograniczeń interpretacji badań w diagnostyce obrazowej są najczęściej:

- specyfika anatomii i fizjologii badanego narządu;
- obraz struktur anatomicznych badanego narządu i patologii w wybranej technice obrazowania;
- specyfika typów i morfologii zmian patologicznych w danej technice obrazowania;
- inne zewnętrzne w stosunku do medycyny i techniki obrazowania uwarunkowania pracy lekarza (organizacja i warunki pracy, poziom techniczny sprzętu itp.).

Często spotykany w życiu splot wszystkich tych czynników stanowi trudne do pokonania ograniczenie interpretacji badań mammograficznych.

Interpretację obrazów w mammografii można podzielić na 3 podstawowe etapy:

1. Detekcję – wymaga ona umiejętności: wizualnej inspekcji mammogramu, detekcji podejrzanego obszaru, odróżnienia rzeczywistej zmiany od wariantu normy, oceny wizualnych cech zmian.
2. Opis – niezbędna jest znajomość typów zmian i pełnego zestawu ich cech oraz umiejętność diagnostycznej interpretacji wartości cech w kontekście danego typu zmiany.
3. Diagnoza – wymaga umiejętności oceny kategorii BI-RADS zmiany (tzn. oszacowania prawdopodobieństwa jej złośliwości) na podstawie konfiguracji wartości jej cech w kontekście informacji z rekordu medycznego pacjentki.

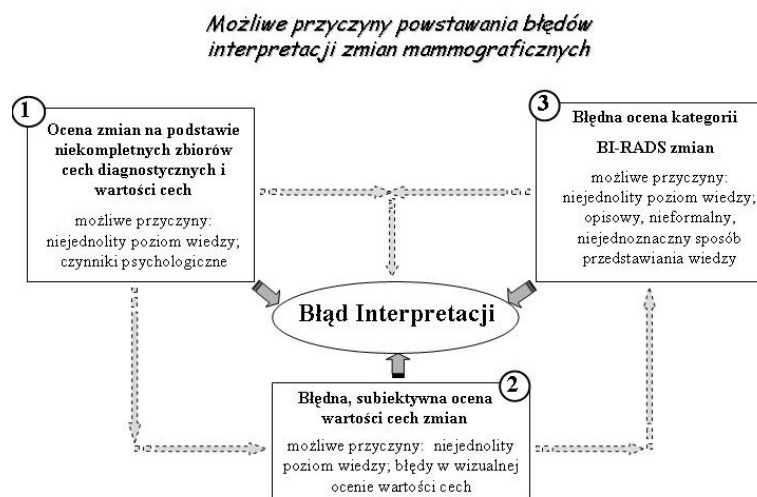
Celem systemu BI-RADS [2] (ang. *Breast Imaging Reporting and Database System*), rekomendowanego przez American College of Radiology, były systematyzacja i standaryzacja terminologii dla opisu i interpretacji badania mammograficznego. Radiolodzy są najczęściej zgodni w swoich ocenach kategorii diagnostycznych BI-RADS ($Kappa=0,73$), kiedy mają do czynienia ze zmianami o cechach typowych dla radiologicznego obrazu raka sutka (skupiska mikrozwpanień o bardzo wysokim prawdopodobieństwie złośliwości). Zgodność ocen dla zmian kategorii 4 BI-RADS¹ jest znacznie niższa ($Kappa=0,28$) [3], szczególnie dla jej podkategorii 4a, 4b i 4c, dla których wartości statystyki Kappa wyniosły odpowiednio 0,14, 0,16 i 0,26. Co prawda terminy zawarte w systemie BI-RADS mają określone znaczenie diagnostyczne, ale brak kompletnych definicji zmian zawierających i podkreślających znaczenie wczesnych i subtelnych oznak procesów nowotworowych (z wyjątkiem odcinkowego charakteru zarysu guzów), utrudnia efektywne wspomaganie interpretacji zmian. Opisowe, niekompletne i nieprecyzyjne, odwołujące się do wiedzy i doświadczenia radiologa oraz obrazów zmian mammograficznych w innych modalnościach zalecenia systemu, dotyczące szacowania kategorii BI-RADS zmian, pozostawiają duży margines dowolności interpretacji, co stawia w szczególnie niekorzystnym położeniu lekarzy dopiero zdobywających doświadczenie. Ponadto, system BI-RADS, wyrażony w formie słownika kontrolowanego, nie jest czytelny dla maszyny i nie pozwala na powtarzalną, automatyczną ocenę kategorii BI-RADS zmian.

2.1. Przyczyny błędnego opisu mammogramów i sposób ich eliminacji

Sztywny podział interpretacji mammogramu na następujące po sobie etapy jest jedynie przybliżeniem rzeczywistego procesu, w którym zadania często są wykonywane w znacznym stopniu równolegle, a ich wyniki wpływają na siebie. Analiza procesu interpretacji mammogramów, prowadząca do zrozumienia źródeł błędów radiologa, jest jednym z możliwych

¹ Są to zmiany podejrzanane, o niejednoznacznej konfiguracji wartości cech, bez klasycznych oznak złośliwości [2]

punktów wyjścia do realizacji skutecznych narzędzi wspomaganie. Analiza doniesień literaturowych, dotyczących procesu interpretacji zmian mammograficznych, zorientowana na poszukiwanie przyczyn błędów popełnianych przez radiologów, pozwoliła na wyłonienie trzech głównych źródeł błędnej interpretacji guzów i mikrozwapnień.



Rys. 1. Przyczyny błędów interpretacji zmian mammograficznych
Fig. 1. Reasons for mammographic interpretation errors

Analizując powyższe przyczyny powstawania błędów oraz uwzględniając wyniki publikacji [4, 5], można sformułować ogólne stwierdzenie, że potencjalnym źródłem błędów interpretacji mammogramów jest niejednolity poziom wiedzy i umiejętności diagnostycznych radiologów oraz ich subiektywizm.

W celu minimalizacji błędnej interpretacji mammogramów można zaproponować następujące metody wspomaganie dla każdej z przyczyn błędów:

1. Niekompletny zbiór cech i wartości cech zmian:

- wprowadzić bazujące na wiedzy dziedzinowej, wyczerpujące w świetle wiedzy medycznej, jednoznaczne definicje pojęć modelujących opis podstawowych zmian mammograficznych, uwzględniające wczesne oznaki złośliwości;
- zaproponować standardowy protokół opisu i oceny badania, zawierający wyżej opisane, wyczerpujące definicje zmian.

2. Błąd wizualnej oceny wartości cechy:

- obiektywizacja oceny wartości cechy przy użyciu deskryptorów matematycznych;
- dobór deskryptorów matematycznych zmian może wykorzystywać bazujące na wiedzy dziedzinowej metody uczenia maszyn.

3. Błąd oceny kategorii BI-RADS zmiany:

- stworzyć kompletne, bazujące na wiedzy dziedzinowej, formalne definicje pojęć modelujących kategorii BI-RADS dla podstawowych zmian mammograficznych;

- wykorzystanie formalnych definicji pojęć modelujących kategorie BI-RADS i mechanizmu wnioskowania w ontologiach do wspomaganie oceny kategorii BI-RADS zmian mammograficznych.

Uwzględniając powyższe, należałoby założyć, że redukcję błędów interpretacji można uzyskać, opisując mammogram za pomocą standaryzowanego protokołu zawierającego kompletne definicje zmian, uwzględniające wczesne, subtelne oznaki złośliwości oraz sygnalizację w trakcie opisu wartości cech wskazujących na charakter zmiany (potencjalnie złośliwy lub łagodny). Zaproponowana metoda redukcji błędów polega na wspomaganie wiedzy radiologa na podstawie specjalistycznych narzędzi informatycznych (edytor MammoEdit). Ponadto, projekt narzędzia realizującego powyższy protokół powinien bazować na formalnie wyrażonej, dziedzinowej wiedzy medycznej – ontologii mammografii.

3. Ontologia mammografii

Pierwsza ontologia mammografii była wyrażona za pomocą ramowej reprezentacji wiedzy. Model zawierał klasy reprezentujące wzorce opisu, ale bez subtelnych cech złośliwości, i klas reprezentujących opis rzeczywistych zmian. Możliwości wnioskowania w tak skonstruowanej ramowej reprezentacji wiedzy były również bardzo ograniczone. Kolejnym krokiem było zatem przeniesienie tego modelu do reprezentacji wyrażonej w języku OWL. Ostateczna wersja ontologii zawiera między innymi klasy modelujące: wyczerpujące, standardowe wzorce opisu podstawowych zmian mammograficznych, wzbogacone o subtelne cechy złośliwości, klasy reprezentujące opis rzeczywistych zmian, wykonany na podstawie uzupełnionych wzorców, oraz klasy modelujące kategorie diagnostyczne systemu BI-RADS.

Najważniejszą różnicą pomiędzy pierwszą a ostatnią wersją ontologii było jednak zapełnienie luki szczegółowości w zaleceniach systemu BI-RADS, dotyczących kategorii diagnostycznych zmian. Luka szczegółowości to znane w biomedycynie zjawisko formułowania kryteriów za pomocą pojęć niejednoznacznych lub zagregowanych. To właśnie owa luka szczegółowości prowadzi do tak dużego stopnia niejednoznaczności opisu. Podczas konstruowania ontologii istnienie luki szczegółowości oznacza w uproszczeniu brak wyjściowych danych koniecznych do wyrażenia ograniczeń nałożonych na dozwolone wartości właściwości klas modelujących kategorie diagnostyczne systemu. Brak tych klas w modelu w sposób oczywisty uniemożliwia zastosowanie mechanizmu klasyfikacji ontologii do oceny kategorii diagnostycznych rzeczywistych zmian mammograficznych.

4. Edytor MammoEdit – narzędzie wspomaganie opisu i interpretacji zmian mammograficznych

Typowy system informatyczny składa się z trzech podstawowych komponentów: interfejsu użytkownika, bazy danych oraz aplikacji, której zadaniem jest wykonanie założonych funkcji systemu. W przypadku edytora ME stworzona ontologia mammografii została użyta jako zbiór częściowych wymagań projektowych dla interfejsu użytkownika i bazy danych przechowującej opisy badań.

Głównym założeniem projektowym było to, że opis mammogramu powinien odbywać się w trybie graficznym, przy użyciu komponentów typu menu, zaś tekst opisu, tzw. epikryza, powinna być generowana automatycznie. Wybór trybu graficznego wydawał się oczywisty, radiolodzy przyzwyczajeni są do obrazów, a piktograficzny sposób zapisu informacji umożliwia zarówno ich szybkie wprowadzenie, jak i natychmiastową wizualizację. Na potrzeby edytora stworzonych zostało ponad 300 piktogramów, obrazujących w sposób sugestywny, ale przede wszystkim jednoznaczny, najważniejsze cechy opisywanych zmian. Biblioteka obiektów graficznych (projekty przycisków, ikon, piktogramów itp.) powstała już po opracowaniu modelu ontologicznego i jest bezpośrednią konsekwencją jego zawartości. Również układ wzajemnych zależności wprowadzanych/prezentowanych danych został oparty na opracowanej ontologii. Pozwoliło to w efekcie na uzyskanie czytelności i jednoznaczności interfejsu oraz uniemożliwiło wprowadzanie danych sprzecznych (wykluczających się). Szczegółowy opis edytora można znaleźć w [1, 6]. Poniżej przytaczamy podstawowe wymagania funkcjonalne, jakie postawione zostały przed interfejsem.

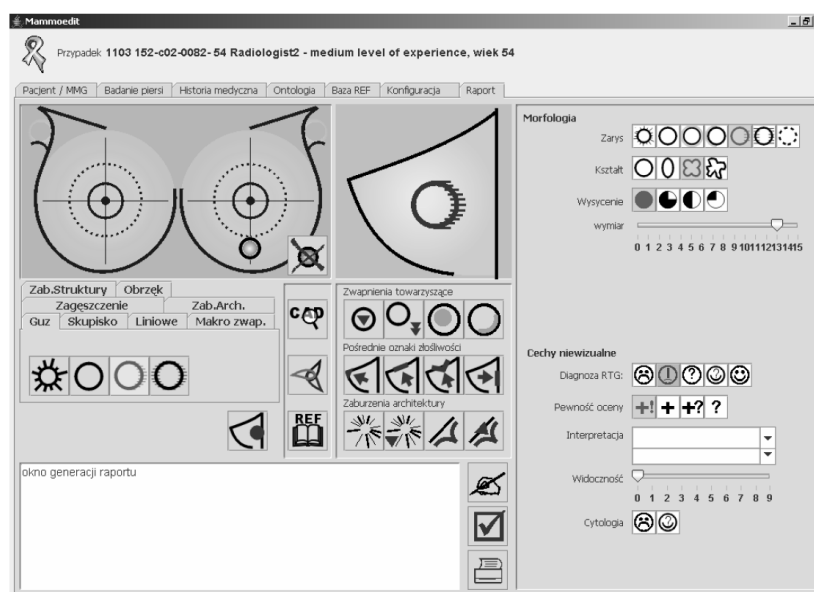
Zapewnienie użytkownikowi (lekarzowi):

- łatwego i szybkiego dostępu do wybranego zakresu edytowanych informacji;
- wzoru opisu uwzględniającego:
 - diagnostycznie istotne cechy zmian w mmg;
 - diagnostycznie istotne wartości cech zmian w mmg;
 - sygnalizację w trakcie opisu tych wartości cech zmian, które wskazują na ich potencjalnie podejrzany charakter;
 - uporządkowanie hierarchii treści opisu oraz przestrzeganie ustalonego protokołu opisu.

Przykładowy widok ekranu edytora ilustruje rys. 2.

W celu systematyzacji prac projektowych nad interfejsem proces opisu zmian podzielony został na następujące, ogólne czynności radiologa:

- wskazanie lokalizacji zmiany,
- wybór typu i kategorii morfologicznej zmiany,
- opis wartości cech wybranej zmiany wraz z oceną kategorii BI-RADS.



Rys. 2. Przykładowy widok okna edytora MammoEdit

Fig. 2. A sample of the MammoEdit's window

Każda z czynności ogólnych składa się z wielu pojedynczych funkcji, realizowanych przy użyciu zaprojektowanych i zaimplementowanych graficznych menu. Opis zmian przy użyciu edytora MammoEdit polega na wykonaniu sekwencji następujących czynności:

1. Wskazanie lokalizacji zmiany przez wybór elementu reprezentującego schemat lokalizacji zmian w sutku (efektem czynności jest wizualizacja wskazanego położenia zmiany).
2. Wybór typu i podstawowej kategorii morfologicznej zmiany (efektem czynności są aktywacja odpowiedniej definicji opisu zmiany oraz wyświetlenie graficznego symbolu wybranej zmiany).
3. Opis zmiany wykonywany przez przyporządkowanie odpowiednich wartości poszczególnym ich cechom (efektem opisu jest wyświetlenie wybranych elementów oraz sygnalizacja prawdopodobieństwa złośliwości zmian, szacowana na podstawie ontologii wybranej cechy; jeśli wskazano zmiany towarzyszące, ich symbole graficzne również są wyświetlane).
4. Ocena kategorii BI-RADS zmiany na podstawie konfiguracji wartości cech.
5. Automatyczna generacja tekstu opisu badania, bazującą na wskazanych przez radiologa wartościach cech zmiany.

5. Ocena narzędzia wspomagania interpretacji mammogramów – testy edytora ME

Podstawowym celem testu była eksperymentalna weryfikacja hipotezy mówiącej, że interpretacja mammogramu przy użyciu edytora, systematyzującego proces opisu oraz sygnali-

zującego radiologowi wartości cech zmian, istotne dla właściwej interpretacji ich charakteru, podnosi efektywność diagnostyczną radiologa.

Test polegał na dwukrotnej interpretacji tego samego zbioru mammogramów (pochodzących z bazy DDSM), po raz pierwszy bez wspomagania, po raz drugi przy użyciu testowanego narzędzia. Próba testowa zawierała mammografie negatywne i badania z patologiami. Ocenę ME przeprowadzono na zbiorze badań zawierających zmiany trudne, głównie zmiany o kategorii 4 BI-RADS. W teście wzięli udział radiolodzy o różnym stopniu doświadczenia zawodowego (stażysta, specjalista i ekspert). Przyjętą miarą wpływu ME na pracę radiologów była różnica między efektywnością interpretacji mammogramów z jego zastosowaniem i bez. Efektywność radiologów oceniono przy użyciu AUC (ang. *Area Under Curve*), podstawowego wskaźnika analizy ROC.

5.1. Analiza ROC

Początki analizy ROC (ang. *Receiver Operating Characteristic analysis*) są datowane na okres II wojny światowej, kiedy metoda ta służyła do oceny dokładności różnicowania sygnału i szumu w detekcji obiektów, przy użyciu radaru [7]. Obecnie obszar zastosowania analizy ROC znacznie się poszerzył i obejmuje między innymi analizę wyników różnego rodzaju testów diagnostycznych w medycynie, bioinformatyce oraz ocenę wyników klasyfikacji [7, 8]. Test diagnostyczny najczęściej wykonuje się w celu podziału pacjentów (przypadków testowych) na dwie grupy: chorych i zdrowych. Każdą z powyższych grup należy podzielić, w zależności od rzeczywistego stanu pacjenta i wyniku jego testu, na dwie podgrupy. Można to przedstawić przy użyciu tablicy kontyngencji o wymiarach 2×2 .

Tabela 1

Podział testu w zależności od stanu pacjenta oraz wyniku testu

Wynik testu	Pacjent	
	Chory	Zdrowy
dodatni	prawdziwie dodatni (<i>TP</i>)	fałszywie dodatni (<i>FP</i>)
ujemny	fałszywie ujemny (<i>FN</i>)	prawdziwie ujemny (<i>TN</i>)

Powyższy podział stanowi podstawę definicji podstawowych miar dokładności testu: czułości (*SE*) i specyficzności (*SP*) [9]. Czułość to zdolność testu do wykrycia choroby wśród pacjentów chorych, czyli prawdopodobieństwo pozytywnego wyniku testu wśród pacjentów chorych $SE = TP / (TP + FN)$. Specyficzność (*SP*) to zdolność testu do wykluczenia występowania choroby dla pacjentów zdrowych, tzn. jest to prawdopodobieństwo negatywnego dla nich wyniku testu $SP = TN / (TN + FP)$.

Wiele ocen w diagnostyce obrazowej bazuje na skalach interwałowych i dlatego zawsze pojawia się kwestia właściwej oceny punktu odcięcia. Liczebności *TP*, *FN*, *FP* i *TN* ściśle zależą od wartości kryterium odcięcia odróżniającego stan uznawany za normę od patologii.

Zmiana tej wartości ma istotny wpływ na ocenę czułości i swoistości stosowanego testu. Wybór wartości kryterium odcięcia ma wpływ na frakcje TP , FN , FP i TN i w konsekwencji na interpretację wyniku testu. Ilościowa ocena wydajności testu diagnostycznego (ang. *diagnostic test performance*) wymaga uwzględnienia zmienności kryterium odcięcia w całym zakresie zmienności tego parametru. Temu celowi służą krzywe ROC, powstające w wyniku wykreślenia zależności Czułości testu (oś Y) w funkcji Specyficzność (oś X). Każdy punkt na krzywej ROC odpowiada pewnej wartości kryterium odcięcia i określonej parze (Czułość, Specyficzność). Dlatego całkowite pole zawarte pod krzywą ROC, oznaczane symbolem AUC, jest interpretowane jako miara efektywności testu [7, 8, 10, 11]. Zaletą oceny wydajności testu przy użyciu analizy ROC jest niezależność² od kryterium odcięcia, a więc eliminacja jego wpływu na oszacowanie czułości i specyficzności, oraz niezależność od użytych jednostek.

5.2. Miary dokładności testu diagnostycznego, metody estymacji krzywej ROC

Kolejnym krokiem w ocenie eksperymentów przy użyciu analizy ROC jest wybór metody estymacji krzywej ROC. Metody te są zazwyczaj dzielone na dwie grupy – nieparametryczne i parametryczne. Najprostsza metoda nieparametryczna polega na połączeniu punktów odpowiadających danym eksperymentalnym, rzutowaniu ich na oś X oraz numerycznym obliczeniu wartości pola pod wykresem na podstawie metody trapezów (inne spotykane w literaturze terminy na określenie tej metody to: metoda Wilcoxona lub metoda eksperymentalna). Zaletą tej metody jest jej prostota i brak założeń dotyczących populacji, z których pochodzą dane eksperymentalne, wadą zaniżone wskaźniki analizy ROC dla małych prób testowych i dyskretnych skal ocen [9, 12, 13]. Najczęściej używaną metodą parametryczną jest metoda binormalna. Zakłada ona, że wyniki testu pochodzą z dwóch populacji, jedna grupuje chorych, druga zdrowych. W obydwu populacjach rozkłady wyników testów są normalne. Krzywa ROC jest wyznaczana przez równanie:

$$y = \Phi(a + b \cdot \Phi^{-1}(x)), \quad (1)$$

gdzie Φ to dystrybucja standardowego rozkładu normalnego, Φ^{-1} to funkcja odwrotna do dystrybucji, $x \in (0,1)$.

Niestety metoda binormalna, podobnie jak eksperymentalna, zaniża wskaźniki analizy ROC dla małych prób testowych i dyskretnych skal ocen zawierających mniej niż 10 punktów.

Do testów edytora ME zdecydowano się wykorzystać analizę ROC z wieloma obserwatorami (*multireader Receiver Operating Characteristic analysis*). Model matematyczny metody uwzględnia typowe dla radiologii źródła zmienności w ocenie badań obrazowych oraz możliwe korelacje pomiędzy ocenami lekarzy w ramach tej samej techniki i możliwe korelacje

² Przy założeniu gaussowskiego rozkładu wyników testu w populacjach chorych i zdrowych.

ocen pomiędzy ocenianymi technikami. Porównanie efektywności różnych technik interpretacji badań obrazowych wymaga przeprowadzenia eksperymentu, w którym na próbę testową składają się:

1. Oceniane techniki – obrazowania, interpretacji lub algorytmów przetwarzania obrazów medycznych.
2. Oceniane przypadki (zbiór ten musi zawierać badania z patologiami i negatywne)[4].
3. Lekarze radiolodzy, interpretujący przypadki testowe przy użyciu badanych technik.

Tego typu eksperymenty, oznaczane w skrócie jako „technika x radiolog x przypadek”, bazują na tzw. projekcie macierzowym (ang. *factorial design*). Jeśli dodatkowo wymagane jest, aby każdy radiolog interpretował każde badanie przy użyciu każdej z porównywanych technik, to eksperyment jest określany jako „sparowany” (ang. *paired*) [14]. Podstawowe miary dokładności testu używane w analizie ROC z wieloma obserwatorami są takie same, jak w analizie ROC: AUC, czułość, specyficzność.

5.3. Skala oceny badanego problemu medycznego – model testu

Wybór skali, w której były oceniane wyniki testu, podyktowany był przede wszystkim przez kryteria merytoryczne rozpatrywanego problemu medycznego [11]. Do wyrażenia oceny wyników testu wybrana została przyporządkowywana przez lekarzy radiologów kategoria diagnostyczna systemu BI-RADS. Skala ta jest dobrze znana radiologom zarówno z literatury, jak i z praktyki. Liczby badań i obserwatorów w typowej próbie testowej dla badawczych eksperymentów MRMC ROC, zalecane przez uznanych autorów, wahają się w dość szerokim zakresie. Z reguły przyjmuje się jednak liczbę przypadków od 50 do 200 i obserwatorów od 2 do 10.

Warunkiem generalizacji wyników testu dla zbioru badań jest ich pochodzenie z różnych ośrodków medycznych, dla populacji lekarzy wymagany jest dodatkowo zróżnicowany poziom kompetencji zawodowych [14, 15]. Przeprowadzony test obejmował 80 mammogramów trudnych w interpretacji (20 negatywnych i 60 z patologiami), opisywanych przez trzy grupy radiologów (stażysta, specjalista, ekspert) za pomocą dwóch metod interpretacji bez i ze wspomaganie za pomocą edytora ME. Mammogramy pochodziły z ogólnodostępnej bazy danych DDSM³ (*Digital Database for Screening Mammography*). Biorąc pod uwagę zróżnicowanie pochodzenia badań i lekarzy oraz zdecydowanie niejednorodności poziom ich doświadczenia zawodowego, w ocenie testu użyto modelu: losowe badania/losowy obserwator.

Dodatkowe założenia, jakie zostały przyjęte podczas testu, niewyspecyfikowane przez samą analizę MRMC ROC, to:

³ Materiał w niej pochodzi z czterech ośrodków medycznych w Stanach Zjednoczonych (Massachusetts General Hospital, University of South Floryda, Sandia National Laboratories i Washington University School of Medicine).

- fizyczne warunki oceny mammogramów, wynikające ze specyfiki badań mammograficznych,
- dane do interpretacji mammogramów,
- odstęp czasu pomiędzy interpretacją tego samego badania bez i ze wspomaganie za pomocą edytora ME.

Ocena mammogramów odbywała się w zacienionym pomieszczeniu. Mammogramy prezentowano radiologom na monitorach medycznych, przy użyciu aplikacji MammoViewer, ograniczonej na potrzeby testu jedynie do funkcji zaawansowanej przeglądarki medycznej. Ocena mammogramów odbywała się w sesjach trwających od 1 do 3 – 4 godzin. Tempo, w jakim odbywała się ocena badań, i czas trwania sesji były dyktowane przez lekarzy, na ich życzenie robiono przerwy w pracy lub ją zawieszano. W czasie jednej sesji lekarz dokonywał tylko jednego typu oceny mammografów, tzn. bez lub z narzędziem wspomagającym. Zmiany oceniane były na podstawie zdjęć piersi w dwóch podstawowych projekcjach, bez dodatkowych projekcji diagnostycznych, wyników badania klinicznego, historii medycznej pacjentki czy poprzednich zdjęć do porównania. Przed pierwszym czytaniem mammogramów wyjaśniono znaczenie używanych skal, w czasie trwania testów opis skal był stale dostępny na stanowisku pracy. Minimalny odstęp czasu pomiędzy interpretacją tego samego badania bez i ze wspomaganie oszacowany został na podstawie danych literaturowych. Przyjmuje się, że powinien wynosić minimum 4 tygodnie. W przeprowadzonym eksperymencie wahał się on od 7 do 12 tygodni.

Do obliczenia wskaźników wydajności diagnostycznej radiologów użyta została metoda DBM oceny eksperymentów MRMC ROC [16, 17], zaś do estymacji wskaźników krzywych ROC metoda PropRoc [18], umożliwiającą wiarygodną estymację wskaźników analizy dla małych prób testowych i dyskretnych skal ocen.

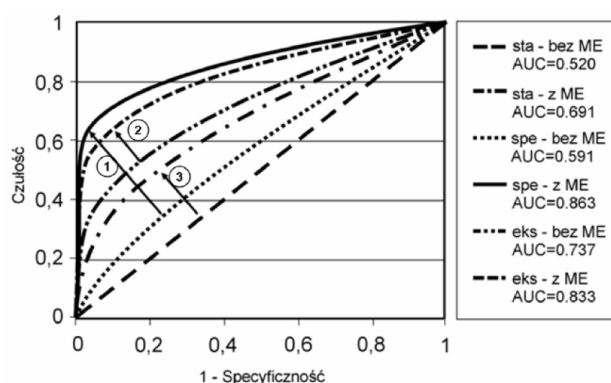
Zastosowanie edytora ME do wspomaganie interpretacji mammogramów podniosło średnią efektywność diagnostyczną (wzrost istotny statystycznie, poziom istotności mniejszy od 0,05). Szczegółowe wyniki przedstawia tabela 2.

Tabela 2

Wyniki oceny wpływu edytora ME na efektywność interpretacji mammogramów

	Wskaźniki Analizy MRMC ROC					
	AUC	AC %	SE %	SP %	SL	AUC(st)
bez edytora ME	0,616	50,3	26,0	96,1	0,0372	0,081
z edytorem ME	0,795	70,7	56,3	98,0		0,060

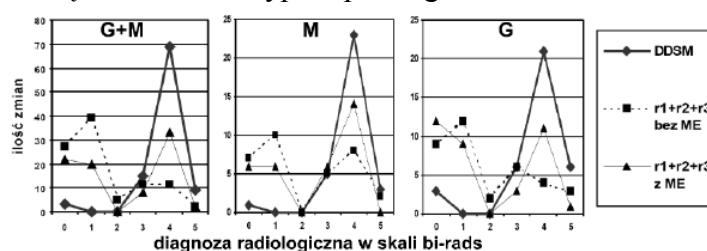
Wzrostowi ww. wskaźników towarzyszyło zmniejszenie odchylenia standardowego pola pod krzywą ROC.



Rys. 3. Wpływ ME na efektywność diagnostyczną radiologów
Fig. 3. ME's influence on the diagnostic effectiveness of the radiologists

Największy wzrost efektywności diagnostycznej zaobserwowano dla radiologa specjalisty, najmniejszy zaś dla radiologa eksperta. Świadczy to o złym wpływie rutyny na jakość oceny mammogramów.

Wpływ ME na zmienność diagnostycznych opinii radiologów oceniono również, porównując wartość statystyki Kappa, uzyskaną w I i II etapie testu. Porównywano zmienność opinii między parami radiologów oraz dla grupy wszystkich radiologów, osobno dla skupisk mikrozwapnień, guzów i łącznie dla obu typów patologii.



Rys. 4. Krzywe IZ-KD dla grupy radiologów (r1 – ekspert, r2 – specjalista, r3 – stażysta) dla danej kategorii diagnostycznej bez i z użyciem ME, dla porównania krzywa IZ-KD dla oceny zmian w DDSM

Fig. 4. IZ-KD curves for the group of radiologists (r1 – expert, r2 – specialist, r3 – intern) in given diagnostic category with and without use of ME

Użycie ME do interpretacji mammogramów zwiększyło spójność opinii diagnostycznych grupy radiologów dla obu typów zmian i każdego z typów patologii, większy wzrost spójności uzyskano dla skupisk mikrozwapnień. ME wpłynął też na wzrost spójności opinii we wszystkich parach obserwatorów i dla wszystkich grup zmian.

6. Podsumowanie

Zastosowanie edytora ME do wspomaganie interpretacji mammogramów podniosło średnią efektywność diagnostyczną grupy radiologów, wzrosły również inne wskaźniki analizy ROC. Pozwala to na stwierdzenie, że wstępnie w warunkach testowych potwierdzono hipote-

zę użyteczności ME jako narzędzia wspomagania interpretacji mammogramów i oceniono poziom reprezentacji wiedzy diagnostycznej w ontologii mammografii jako zadowalający. Pogłębiona analiza efektywności radiologów, w podziale na typy zmian, wykazała istnienie różnicy w efektywności wspomagania interpretacji guzów i mikrozwapnień (na niekorzyść guzów). Fakt ten oceniono jako potrzebę powtórzenia etapu pozyskiwania wiedzy diagnostycznej dotyczącej guzów.

Wstępna ocena logicznej poprawności metody budowy klas polegała na porównaniu kategorii BI-RADS wybranych przypadków testowych w bazie DDSM z oceną kategorii tych samych zmian uzyskaną przy wspomaganiu za pomocą edytora ME i oceną kategorii zmiany uzyskaną za pomocą klasyfikacji ontologii. W wyniku eksperymentu stwierdzono, że oceny kategorii BI-RADS zmian, uzyskane za pomocą klasyfikacji ontologii, są zgodne z ocenami w bazie DDSM, ponadto użycie edytora ME poprawiło zgodność oceny ze wzorcem nawet dla grupy radiologa eksperta.

Edytor MammoEdit wspomaga tylko fragment pracy radiologa, dlatego by móc go zastosować w praktyce klinicznej, należałoby jeszcze pozyskać wiedzę o matematycznych deskryptorach wszystkich wartości cech podstawowych typów zmian mammograficznych, guzów i skupisk mikrozwapnień, zintegrować go z aplikacją typu CAD, połączyć detekcję zmian z ich zweryfikowanym opisem, a następnie poszerzyć aplikację o funkcję współpracy z ontologią mammografii, w celu automatycznej weryfikacji kategorii BI-RADS zmian. Ostatni etap powinien mieć charakter implementacyjny, należy przeprowadzić testy, ocenić efektywność i uzyskać akceptację całości systemu przez użytkownika. Uzupełnieniem funkcjonalności zintegrowanej aplikacji są funkcje administracyjne (typu RIS ang. *Radiology Information System*), zarządzanie danymi (funkcje bazodanowe) i bezpieczeństwem oraz zintegrowanie ze stacją roboczą i archiwum typu PACS (ang. *Picture Archiving and Communication System*).

BIBLIOGRAFIA

1. Podsiadły-Marczykowska T., Zawisłak R.: *Dziedzinowa ontologia mammografii w projektowaniu edytora raportów. Bazy Danych: Struktury, Algorytmy, Metody*. WKŁ, Warszawa 2006.
2. American College of Radiology: *Breast imaging reporting and data system (BI-RADS)*. 4th ed. Reston, American College of Radiology, 2003.
3. Lazarus E., Mainiero M. B., Schepps B., Koelliker S. L., Livingston S. L.: *BI-RADS Lexicon for US and Mammography: Interobserver Variability and Positive Predictive Value*. *Radiology* 2006, Vol. 239, No. 2, s. 385÷391.

4. Wagner R. F., Beam C. A., Beiden S. V.: Reader Variability in Mammography and Its Implications for Expected Utility over the Population of Readers and Cases. *Medical Decision Making*, Vol. 24, No. 6, 2004, s. 561÷572.
5. Miglioretti D. L., Smith-Bindman R., Abraham L., Brenner R. J., Carney P. A.: Radiologist Characteristics Associated With Interpretive Performance of Diagnostic Mammography. *J Natl Cancer Inst* 2007, Vol. 99, 2007, s. 1854÷1863.
6. Przelaskowski A., Podsiadły-Marczykowska T., Koniński P., Wróblewska A.: Zintegrowany System wspomagania diagnostyki mamograficznej. *Elektronika*, nr 4, 2008, s. 122-127.
7. Swets J. A.: ROC analysis applied to the evaluation of medical imaging tests. *Invest Radiol*, Vol. 14, 1979, s. 109÷121.
8. Swets J. A., Pickett R.M.: Evaluation of diagnostic systems: methods from signal detection theory. Academic Press, New York 1982.
9. Zwig M. H., Campbell G.: Receiver-operating characteristics (ROC) plots: a fundamental evaluation tool in clinical medicine. *ClinChem*, No. 39(4), 1993, s. 561÷577.
10. Metz C. E.: ROC Methodology in Radiologic Imaging. *Invest Radiol*, No. 21, 1986, s. 720÷733.
11. Zhou X. H., Obuchowski N.A., Obuchowski D.M.: Statistical Methods in Diagnostic Medicine. John Wiley & Sons, New York 2002.
12. Hanley J. A., McNeil B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, No. 143, 1982, s. 29÷36.
13. Centor R. M., Schwartz J.: An evaluation of methods for estimating the area under the receiver operating characteristic (ROC) curve. *Med Decis Making*, No. 5, 1985, s. 149÷156.
14. Obuchowski N. A., Beiden S.V., Berbaum K.S., Stephen L., Hillis S.L.: Multireader, Multicase Receiver Operating Characteristic Analysis: An Empirical Comparison of Five Methods. *Acad. Radiol.*, No. 11, 2004, s. 980÷995.
15. Obuchowski N. A.: Fundamentals of Clinical Research for Radiologists. *AJR*, No. 184, 2005, s. 364÷372.
16. Dorfman D. D., Berbaum K. S., Metz C. E.: Receiver operating characteristic rating analysis: generalization to the population of readers and patients with the jackknife method. *Invest Radiol.*, No. 27, 1992, s. 723÷731.
17. Dorfman D. D., Berbaum K. S., Lenth R. V., Chen Y. F., Donaghy B. A.: Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. *Acad. Radiol.*, No. 5, 1998, s. 591÷602.
18. Pesce L. L., Metz C. E.: Reliable and computationally efficient maximum-likelihood estimation of "proper" binormal ROC curves. *Acad. Radiol.*, No. 14(7), 2007, s. 814÷829.

Recenzent: Dr inż. Paweł Kasprowski

Wpłynęło do Redakcji 31 stycznia 2011 r.

Abstract

Mammography is currently the most effective tool for early detection of breast cancer. It is still the only technique that can detect breast cancer in preinvasive stage. However, the diagnostic value of mammography is limited by significant and high, up to 25% rate of missed breast cancers. The missed malignant lesions only rarely are effect of negligence. Computer Aided Detection (CAD) systems aiming to reduce detection errors are now commercially available and are gaining increasing practical importance, but no serious attempts have been made to apply CAD systems for lesion diagnosis in practical clinical situations. The paper presents an assessment of diagnostic accuracy of MammoEdit (ME) – an ontology-based editor supporting mammograms description and interpretation. Radiologists are familiar with images, so the choice of graphical, iconic data presentation model for mammography report editor was obvious. Concepts necessary for mammographic lesions description were represented by graphical primitives – icons (Fig. 2.).

The aim of the tests was to compare the effectiveness of mammogram interpretation made in a routine way without aiding tools with the effectiveness of such interpretation made with our aiding tools. Tests were performed on mammograms from DDSM. The test sample for editor test consisted of 80 mammograms (20 negatives and 60 pathology case). For the assessment of the tested tool accuracy, the MRMC ROC studies were used. The use of MammoEdit in testing conditions resulted in increase of radiologist's diagnostic performance. Other measures of the diagnostic performance such as accuracy, sensitivity and specificity improved as well (Table 2, Fig. 3. and Fig. 4). From the analysis of the test results three major conclusions can be drawn. The first is that in testing conditions, MammoEdit, editor for description and interpretation of mammograms, showed a potential to improve radiologists diagnostic performance. Perhaps after proving its utility, the tool could be coupled with an ISAD system, helping to reliably describe and interpret CAD-cued lesions. The second conclusion is that there is a difference in achieved diagnostic performance improvement between microcalcification and masses. The last one is that the tool in question could also be used as an educational tool, but this issue needs further investigations.

Adresy

Teresa PODSIADŁY-MARCZYKOWSKA: Instytut Biocybernetyki i Inżynierii Biomedycznej PAN, ul. Trojdena 4, 02-109 Warszawa, tpodsiadly@ibib.waw.pl.

Rafał ZAWIŚLAK: Politechnika Łódzka, Instytut Automatyki, ul. Stefanowskiego 18/22, 90-912 Łódź, Polska, rafal.zawislak@p.lodz.pl.