
Prof. dr hab. Zbislaw Tabor
Katedra Biocybernetyki i Inżynierii Biomedycznej
Wydział Elektrotechniki, Automatyki, Informatyki
i Inżynierii Biomedycznej
Akademia Górniczo-Hutnicza
ztabor@agh.edu.pl

Kraków, 23 maja 2022

PW WEiTI Kancelaria
wpłynęło dnia 31.05.22r.
numer

RECENZJA

rozprawy doktorskiej Pana mgra inż. Piotra Sobeckiego pt.:

**„THE USE OF DOMAIN KNOWLEDGE IN METHODS OF COMPUTER-AIDED PROSTATE
CANCER DIAGNOSIS”**

wykonanej pod kierunkiem promotora prof. dra hab. inż. Artura Przelaskowskiego

I. Ogólna charakterystyka podjętych przez Doktoranta problemów badawczych

W przedłożonej mi do oceny rozprawie Doktorant przedstawia założenia badawcze, metodykę oraz wyniki prac nad dwoma problemami wpisującymi się w obszar zastosowania metod informatycznych we wspomaganiu diagnostyki medycznej, w szczególności w diagnostyce raka prostaty. Pierwszym z rozwiązywanych problemów jest opracowanie narzędzi do binarnej klasyfikacji zmian uwidocznionych w prostacie w multiparametrycznym badaniu rezonansu magnetycznego (RM). Narzędzia takie przyporządkowują zobrażowanej zmianie jedną z dwóch kategorii: zmiana klinicznie istotna/zmiana klinicznie nieistotna. Drugim rozwiązywanym problemem jest opracowanie metodyki oraz implementacja aplikacji (z interfejsem webowym) do przygotowywania ustrukturyzowanych opisów radiologicznych, z przykładowym zastosowaniem w opisie multiparametrycznego badania RM prostaty.

W rozwiązaniu obu problemów wykorzystywana jest wiedza dziedzinowa tzn. rekomendowany schemat postępowania diagnostycznego w medycznej diagnostyce obrazowej raka prostaty PI-RADS 2.1. Schemat PI-RADS 2.1 definiuje procedury, które pozwalają przypisać zmianom patologicznym uwidocznionym w badaniu RM ocenę w skali od 1 do 5, w oparciu o cechy obrazowe tych zmian. O ile ocena w skali PI-RADS może być interpretowana jako stopień prawdopodobieństwa istotności klinicznej, to sama istotność kliniczna jest

zmienną binarną, której wartość może być wiarygodnie ustalona w zasadzie tylko w badaniu histopatologicznym.

Ocena klinicznej istotności zmiany w skali PI-RADS determinuje sposób postępowania z pacjentem i dlatego opracowanie narzędzi informatycznych pozwalających na dokładną predykcję istotności klinicznej zmian patologicznych prostaty może mieć ogromne, pozytywne znaczenie dla systemu ochrony zdrowia. Badania nad takimi narzędziami znajdują się wciąż na froncie badań nad zastosowaniami uczenia maszynowego w medycznej diagnostyce obrazowej i dlatego uważam pierwszy z podjętych przez Doktoranta tematów badawczych jako ważny, mający istotny pierwiastek naukowy i wpisujący się w dyscyplinę informatyki technicznej i telekomunikacji, w szczególności w obszar wykorzystania uczenia maszynowego w diagnostyce medycznej.

Drugi z podjętych przez Doktoranta problemów, tzn. opracowanie metodyki rozwoju systemów raportowania strukturalnego, według której zestaw elementów w strukturze raportu oraz zbiór reguł wspomagających decyzje diagnostyczne definiowane są w oparciu o wiedzę dziedzinową i procesy standaryzacyjne, również uważam za ważny i wpisujący się w dyscyplinę informatyki technicznej i telekomunikacji. Opracowana metodyka jest przez Doktoranta użyta w implementacji aplikacji do wspomagania przygotowywania ustrukturyzowanych opisów radiologicznych miltiparametrycznego RM prostaty. Zaproponowany przez Doktoranta koncept teoretyczny jest więc wykorzystany w praktycznym problemie, co także uważam za ważne osiągnięcie, ponieważ aplikacja taka – co pokazuje Doktorant - przyczynia się do zwiększenia wydajności pracy lekarza radiologa oraz obiektywizacji pracy diagnostycznej. Ustrukturyzowane opisy mogą być również znacznie łatwiej wykorzystane np. w zadaniach tagowania obrazów w celu wykorzystania tych obrazów w projektowaniu narzędzi wspierających procesy diagnostyczne.

II. Ogólna charakterystyka rozprawy

Na przedłożoną mi do oceny rozprawę składa się pięć rozdziałów. W rozdziale pierwszym jest przedstawiona wiedza dziedzinowa, dotycząca problematyki diagnostyki raka prostaty, w tym diagnostyki obrazowej. Doktorant omawia metody obrazowania medycznego wykorzystywane w diagnostyce obrazowej raka prostaty oraz wyjaśnia schemat postępowania diagnostycznego PI-RADS 2.1. Rozdział kończy się тезami pracy.

Rozdział drugi, mający formę artykułu naukowego, przedstawia wyniki prac nad narzędziami do klasyfikacji binarnej zmian patologicznych uwidocznionych w multiparametrycznym badaniu RM prostaty. Po części wstępnej przedstawiony jest materiał badawczy oraz metody wykorzystane do opracowania klasyfikatorów. W opisie metodyki Doktorant wyjaśnia, w jaki sposób wiedza dziedzinowa jest wykorzystana przy projektowaniu klasyfikatorów. W dalszej części raportowane są wyniki osiągnięte przez klasyfikatory opracowane przez Doktoranta - w tym celu Doktorant posługuje się polem pod krzywą ROC. Wyniki osiągnięte przez klasyfikatory są porównane z polem pod krzywą ROC skonstruowaną dla lekarzy radiologów. Rozdział kończy się dyskusją wyników oraz argumentacją, w jaki sposób wykorzystanie wiedzy dziedzinowej może wpłynąć korzystnie na proces klasyfikacji zmian zobrazowanych w badaniu RM.

W rozdziale trzecim są przedstawione są wyniki prac nad metodyką tworzenia ustrukturyzowanych opisów radiologicznych, z przykładowym zastosowaniem do opisu multiparametrycznego badania RM prostaty. We współpracy z końcowymi użytkownikami (lekarzami radiologami) jest opracowany leksykon terminów radiologicznych używanych w schemacie diagnostycznym PI-RADS 2.1 oraz zbiór reguł używanych do przypisywania zmianom oceny w skali PI-RADS. Leksykon i reguły są użyte do zbudowania systemu regułowego a następnie do zaimplementowania aplikacji wspomagającej procesy decyzyjne w diagnostyce. Aplikacja została użyta w środowisku laboratoryjnym i klinicznym do oceny „inter-reader agreement” i „intra-reader agreement”. Rozdział opisuje prace o charakterze teoretycznym (metodyka tworzenia systemów wspierających decyzje diagnostyczne), prace o charakterze inżynierskim (przygotowanie projektu i implementacja aplikacji) oraz wyniki badań statystycznych.

W rozdziale czwartym Doktorant wymienia najważniejsze osiągnięcia przeprowadzonych badań, które za Doktorantem przepisują poniżej:

1. „We have demonstrated that domain knowledge can be effectively applied to construct and improve the machine learning models of PCa diagnosis”.
2. „We have shown that the PI-RADS algorithm can be represented using subnetworks integrated with routing on multimodel CNNs, and tailored fitness function that favours specific modalities, which are more efficient in the diagnosis of lesions, depending on their locations”.
3. Dzięki wykorzystaniu aplikacji do tworzenia ustrukturyzowanych opisów „it is possible to identify the descriptors that characterise poor intra- and inter-reader

agreement; these could potentially benefit from redefinition in radiological lexicons or from integration of automatically quantified image features”.

Wymienione powyżej osiągnięcia są powtórzone w konkluzjach pracy, zawartych w rozdziale 5, po którym następuje bibliografia, zawierająca 137 pozycji, z których trzy (pozycje [91] i [92] to artykuły konferencyjne, a pozycja [96] to artykuł w recenzowanym czasopiśmie) są wieloautorskimi publikacjami w których Doktorant jest pierwszy autorem.

Układ rozprawy uważam za prawidłowy, a bibliografię za wystarczającą.

III. Formalna ocena rozprawy

W rozdziale 2 rozprawy opisane są prace badawcze nad metodami klasyfikacji zmian patologicznych w multiparametrycznym badaniu RM prostaty z wykorzystaniem wiedzy dziedzinowej (schemat PI-RADS 2.1). Zgodnie z cytowaniami umieszczonymi w tekście rozprawy, prace te były opisane również w wieloautorskich publikacjach [91], [92] i [96], w których Doktorant był pierwszym, ale nie jedynym autorem. Podobnie rzecz się ma z rozdziałem 3, którego część eksperymentalna została przedstawiona w rozdziale monografii „Legal and technical aspects of artificial intelligence” (Wyd. UKSW, 2021, str. 311-322). Jednym z moich zadań, jako recenzenta jest wyrażenie opinii, czy „rozprawa doktorska prezentuje ogólną wiedzę teoretyczną kandydata w dyscyplinie albo dyscyplinach oraz umiejętność samodzielnego prowadzenia pracy naukowej” (Art. 187 ust. 1 ustawy Prawo o szkolnictwie wyższym i nauce). Z dostarczonej mi dokumentacji, zawierającej m.in. podpisane przez współautorów oświadczenia o ich udziale w pracy zbiorowej wynika, że Doktorant zaproponował koncepcję badań opisanych w wyżej cytowanych artykułach oraz był w te badania zaangażowany we wszystkich etapach – od opracowania metodyki, obejmującej rozwój narzędzi do analizy danych, opartych o uczenie maszynowe, do przygotowania publikacji. **W oparciu o przekazaną mi dokumentację wyrażam zatem przekonanie, że oceniana przeze mnie rozprawa doktorska prezentuje ogólną wiedzę teoretyczną Doktoranta w dyscyplinie informatyka techniczna i telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej.**

Chociaż, problematyka dotycząca ustrukturyzowanych raportów (również wykorzystywanych w diagnostyce raka prostaty) jest dobrze znana w środowisku lekarzy radiologów i doczekała się już opracowań książkowych np. Olga Brook;

Wieland H. Sommer: Radiology Structured Reporting Handbook: Disease-Specific Templates and Interpretation Pearls (Thieme, 2021), według mnie najbardziej istotną nowością wprowadzoną w rozprawie jest sformalizowanie translacji radiologicznych rekomendacji, dotyczących sposobu prowadzenia procesów diagnostycznych i mających formę narracyjną, do zestawu pojęć i łączących ich relacji, które z kolei stają się elementami systemu regułowego wspierającego proces diagnostyczny. W oparciu o system wspierający procesy diagnostyczne, implementujący zestaw reguł wyprowadzonych z rekomendacji diagnostycznych, Doktorant eksperymentalnie potwierdza korzyści stosowania proponowanej metodyki, polegające przede wszystkim na obiektywizacji procesu diagnostycznego. **Jest to - moim zdaniem - wymagane przez ustawę Prawo o szkolnictwie wyższym i nauce (Art. 187 ust. 2) dla pozytywnej oceny rozprawy doktorskiej oryginalne rozwiązanie podjętego przez Doktoranta problemu naukowego z obszaru dyscypliny informatyka techniczna i telekomunikacja, dotyczącego metodyki rozwoju systemów raportowania strukturalnego.**

IV. Uwagi do rozprawy

Rozdział 2 rozpoczyna opis bazy danych obrazowych użytej przez Doktoranta (ProstateX: <https://prostatex.grand-challenge.org/>). Baza zawiera zarówno treningowe obrazy zmian z udostępnionymi etykietami „zmiana istotna klinicznie”/”zmiana nieistotna” jak i obrazy testowe bez udostępnionych etykiet. W pracy Doktorant używa 330 obrazów treningowych (w tym 254 zmiany nieistotne).

Obrazy te są w pierwszej kolejności użyte do wytrenowania klasycznych klasyfikatorów, przyjmujących na wejściu cechy radiomiczne wyliczone dla wszystkich modalności, dla całych obrazów i dla wycinków obejmujących zmianę (ewentualnie z pewnym marginesem – informacja o obszarach obejmujące zmiany jest dostarczona razem z bazą danych). Doktorant trenuje dwa klasyczne klasyfikatory: sieć neuronową oraz kNN. Sieć neuronowa jest trenowana dla dwóch predefiniowanych zestawów cech wejściowych, a jej skuteczność jest oceniana w oparciu o schemat walidacji krzyżowej potrójnej. W tekście rozprawy jest podana informacja o architekturze użytej sieci razem z informacją, że inne architektury były również sprawdzone, ale wymieniona w tekście osiągnęła najlepsze wyniki (mierzone polem pod krzywą ROC) na zbiorach walidacyjnych. Dla klasyfikatora kNN wybierany był optymalny (w sensie pola pod krzywą ROC)

parametr N oraz zestaw zmiennych objaśniających – w tym celu zastosowano przeszukiwanie z wykorzystaniem algorytmu genetycznego.

Wyniki osiągnięte przez sieci neuronowe są słabe (AUC co najwyżej 0.73 na zbiorze testowym, wyniki przedstawiono w publikacji [91]), podczas gdy wyniki osiągnięte dla klasyfikatora kNN (AUC =0.92), opublikowane w 2017r. (artykuł [92]) są o 0.01 lepsze, niż najlepszy aktualny wynik na tablicy wyników <https://prostatex.grand-challenge.org>. Wynik 0.92 uzyskano jednak dla innego zbioru testowego, niż w publikacji [91].

W dalszej części rozprawy Doktorant przedstawia wyniki prac nad klasyfikatorami opartymi o głębokie sieci neuronowe, ale wyniki osiągane dla tych modeli (AUC=0.84, przedstawione również w publikacji [96]) są gorsze, niż wynik raportowany dla klasyfikatora kNN. Doktorant proponuje dwie architektury do segmentacji (z których jedna ma emulować proces diagnostyczny prowadzony wg schematu PI-RADS), przyjmujące na wejściu obrazy z trzech sekwencji RM, które osiągają porównywalne wyniki. Klasyfikatory oparte o sieci głębokie są ewaluowane na stronie ProstateX i plasują się aktualnie w okolicach trzechsetnej pozycji.

W dalszej części Doktorant porównuje krzywe ROC dla klasyfikatorów głębokich z krzywymi ROC uzyskanymi dla zadania klasyfikacji rozwiązywanego przez dwójkę lekarzy (jeden lekarz z mniejszym doświadczeniem w pracy z PI-RADS, drugi lekarz z większym doświadczeniem). Wyniki przedstawione na Rys. 2.3.4 pokazują, że pola pod krzywymi ROC lekarzy są mniejsze, niż dla klasyfikatora głębokiego.

W prezentacji wyników badań, opisanych w rozdziale 2, brakuje konsekwentnego stosowania identycznej metodyki, co utrudnia ocenę jakości tych wyników. Jest to po części efekt przyjętej przez Doktoranta koncepcji rozdziału 2, który jest skrótem z trzech publikacji opublikowanych w latach 2017-2021, a po części sposobu funkcjonowania witryny <https://prostatex.grand-challenge.org> w tym okresie. W rozprawie uderza przede wszystkim kontrast między wynikami raportowanymi dla klasyfikatora kNN, a wynikami uzyskanymi dla modeli głębokich, będący efektem sposobu ewaluacji obu modeli. Wydaje się, że rzetelna ewaluacja opisanych modeli mogłaby być wykonana na co najmniej dwa sposoby:

1. przy wysłaniu modelu z pracy [96] do <https://prostatex.grand-challenge.org> Doktorant mógł jednocześnie wysłać do ewaluacji model z pracy [92],

2. wszystkie modele powinny być przeliczone, dla celów ich prezentacji w rozprawie raz jeszcze na identycznych zbiorach testowych, treningowych i walidacyjnych.

W rozprawie Doktorant porównuje klasyfikatory, porównując pola pod krzywymi ROC. Nie zauważyłem w rozprawie wykorzystania w tym celu właściwych narzędzi statystycznych (np. NCSS Statistical Software: Comparing Two ROC Curves – Paired Design). Nie rozumiem idei wykorzystania w rozprawie testu Wilcoxona do porównania pól pod krzywymi ROC w sytuacji, gdy istnieją narzędzia statystyczne odpowiednie do tego typu porównań. Ponadto w rozprawie nie ma żadnej informacji o jakości klasyfikacji dla optymalnego (np. w sensie statystyki Youdena) progu decyzyjnego – z klinicznego punktu widzenia taka informacja jest równie istotna, jak informacja o AUC.

Nie jest jasno opisane, jak została wyznaczona krzywa ROC dla lekarzy – nawet, jeśli dyskretna, pięciostopniowa ocena PI-RADS została przetłumaczona na parametr decyzyjny, przyjmujący pięć wartości, porównywanie takiej pięciopunktowej krzywej z krzywą uzyskaną dla ciągłego parametru decyzyjnego klasyfikatorów nie jest uczciwe.

Na str. 76 rozprawy Doktorant pisze: „such suggestions play the role of a ‘second opinion’” w odniesieniu do oceny zmiany w skali PI-RADS, sugerowanej przez zaproponowany przez Doktoranta system regułowy. Jednak ocena PI-RADS jest proponowana w oparciu o wartości cech zmiany, wybrane manualnie przez lekarza. Cechy te z kolei determinują ocenę PI-RADS, zgodnie z rekomendacjami PI-RADS, wobec czego twierdzenie, że otrzymujemy w ten sposób drugą opinię jest nieco na wyrost. Druga ocena mogłaby być uzyskana, gdyby system sam proponował wartości cech, co być może da się osiągnąć integrując z rozwijanym przez Doktoranta systemem modele oparte o uczenie maszynowe, a taka integracja jest – zgodnie z sugestią Doktoranta – naturalnym kierunkiem rozwoju projektu systemu wspomagającego decyzje diagnostyczne.

V. Wnioski

W mojej opinii przedłożona mi do recenzji rozprawa spełnia wszystkie, określone ustawą Prawo o szkolnictwie wyższym kryteria, wymagane do jej pozytywnej oceny:

1. prezentuje ogólną wiedzę teoretyczną Doktoranta w dyscyplinie informatyka techniczna i telekomunikacja oraz umiejętność samodzielnego prowadzenia pracy naukowej,

2. zawiera oryginalne rozwiązanie podjętego przez Doktoranta problemu naukowego, dotyczącego metodyki rozwoju systemów raportowania strukturalnego.

Oceniając zatem rozprawę pozytywnie, wnioskuję o dopuszczenie Pana Piotra Sobeckiego do dalszych etapów postępowania w sprawie nadania stopnia doktora w dyscyplinie informatyka techniczna i telekomunikacja.

Zbysław Tabor

Zbysław Tabor