

Akceptuję
H. J. M.

RECENZJA ROZPRAWY DOKTORSKIEJ

**Rada Dyscypliny Nauk Medycznych
Warszawski Uniwersytet Medyczny**

Rozprawa na stopień naukowy doktora nauk medycznych i nauk o zdrowiu
w dyscyplinie nauki medyczne

Imiona i nazwisko autora rozprawy doktorskiej:
lek. Cezary Piotr Maciejewski

Tytuł rozprawy doktorskiej:
**"Wykorzystanie nowoczesnych technik analizy danych tekstowych
w elektronicznej dokumentacji medycznej,
w celu stworzenia narzędzi przyspieszających pozyskanie wartościowych naukowo
danych ustrukturyzowanych oraz zautomatyzowanych skal ryzyka w kardiologii"**

Promotor:
prof. dr hab. n. med. Paweł Balsam
Promotor pomocniczy:
dr hab. n. med. Krzysztof Ozierański

I Katedra i Klinika Kardiologii, Warszawski Uniwersytet Medyczny

Recenzent:
dr hab. n. med. Paweł T. Matusik, prof. UJ

1. **Wartość naukowa i merytoryczna oraz oryginalność rozprawy doktorskiej**

Elektroniczna dokumentacja medyczna stanowi źródło informacji zarówno w aspekcie klinicznym, jak również w kwestii prowadzonych badań naukowych. Elektroniczna dokumentacja medyczna zwiększa wiarygodność danych medycznych, ułatwia do nich dostęp, pozwala zaoszczędzić czas oraz wspiera podejmowanie decyzji klinicznych. Dokumentacja ta z biegiem czasu może zawierać bardzo dużą ilość danych medycznych, które mogą być w sposób niemal automatyczny umieszczane w bazach danych stanowiąc cenny materiał do analiz celem optymalizacji opieki, jak również prowadzenia badań naukowych. Tematyka przedstawionej rozprawy doktorskiej Pana lek. Cezarego Maciejewskiego dotycząca wykorzystania nowoczesnych technik analizy danych tekstowych w elektronicznej dokumentacji medycznej, w celu stworzenia narzędzi przyspieszających pozyskanie wartościowych naukowo danych ustrukturyzowanych oraz zautomatyzowanych skal ryzyka w kardiologii jest bardzo istotna i aktualna.

Lekarz Cezary Maciejewski w pracy doktorskiej w sposób umiętny wprowadza czytelnika w tematykę badawczą, podkreślając znaczenie elektronicznej dokumentacji medycznej (uwzględniając opinię międzynarodowej grupy eksperckiej) oraz omawiając dotychczasowy stan wiedzy w tym zakresie. Przedstawia ograniczenia danych gromadzonych głównie w celach administracyjnych / rozliczeniowych oraz wskazuje, iż około 80% danych w elektronicznej dokumentacji medycznej stanowią dane nieustrukturyzowane. W dalszej części rozprawy Pan lek. Cezary Maciejewski przedstawia hipotezy badawcze oraz cele ogólne prac w cyklu. Trzeba zauważyć, iż podjęta przez lek. Cezarego Maciejewskiego tematyka badawcza jest wymagająca oraz oryginalna. Dobrał on odpowiednie metody naukowe oraz przeprowadził stosowne analizy statystyczne zebranych danych.

Główne wyniki pracy doktorskiej przedstawione są w postaci cyklu 3 spójnych tematycznie oryginalnych artykułów naukowych:

1. Maciejewski C., Ozierański K., Basza M., Łodziński P., Śliwczyński A., Kraj L., Krajsman M., Prado Paulino J., Tymińska A., Opolski G., Cacko A., Grabowski M., Balsam P. Administrative Data in Cardiovascular Research—A Comparison of Polish National Health Fund and CRAFT Registry Data. *Int. J. Environ. Res. Public Health*. 2022; 19(19): 11964. Punktacja Ministerstwa Nauki i Szkolnictwa Wyższego (MNiSW): 140 ("article").

Wkład lek. Cezarego Maciejewskiego w powstanie publikacji: 86%.

2. Maciejewski C., Ozierański K., Barwiołek A., Basza M., Bożym A., Ciurla M., Krajsman M., Maciejewska M., Łodziński P., Opolski G., Grabowski M., Cacko A., Balsam P. AssistMED project: transforming cardiology cohort characterisation from electronic health records through natural language processing – algorithm design, preliminary results, and field prospects. *Int J Med Inform*. 2024; 185: 105380. IF = 4,9, punktacja MNiSW: 140 ("research article").

Wkład lek. Cezarego Maciejewskiego w powstanie publikacji: 81%.

3. Maciejewski C., Ozierański K., Basza M., Barwiołek A., Ciurla M., Bozym A., Krajsman M., Łodziński P., Opolski G., Grabowski M., Cacko A., Balsam P. Practical use case of natural language processing for observational clinical research data retrieval from electronic health records: AssistMED project. *Pol Arch Intern Med.* 2024; Mar 19: 16704.

IF = 4,8, punkty MNiSW: 200 ("original article").

Wkład lek. Cezarego Maciejewskiego w powstanie publikacji: 82%.

Opublikowane prace poddane zostały recenzji, w tym 2 w czasopismach naukowych posiadających współczynnik Impact Factor (na czas złożenia pracy doktorskiej sumaryczny Impact Factor prac wynosił 9,7, natomiast suma punktów MNiSW wynosiła 480).

Pan lek. Cezary Maciejewski we wszystkich artykułach jest pierwszym autorem oraz posiada bardzo znaczący wkład w ich powstanie (wynoszący we wszystkich pracach ponad 80%).

Lek. Cezary Maciejewski dokonał krytycznej analizy oraz interpretacji uzyskanych wyników i dokonał szczegółowego przeglądu piśmiennictwa z zakresu tematyki rozprawy doktorskiej. Lek. Cezary Maciejewski w pracy doktorskiej przedstawia możliwości wykorzystania uzyskanych wyników (w tym oszczędności czasowe i finansowe pozyskiwania dużych ilości danych), proponuje przyszłe perspektywy wykorzystania technik procesowania języka naturalnego oraz omawia ich ograniczenia. Praca doktorska kończy się podsumowaniem i wnioskami, oświadczeniami współautorów publikacji określającymi indywidualny wkład każdego z nich w powstanie poszczególnych publikacji (z jednoczesnym wyrażeniem przez współautorów publikacji zgody na wykorzystanie publikacji jako części rozprawy doktorskiej przez lek. Cezarego Maciejewskiego) oraz bibliografią.

Celem pierwszej publikacji wchodzącej w skład cyklu była ocena dokładności danych zbieranych przez Narodowy Fundusz Zdrowia (NFZ, z perspektywy badacza) w odniesieniu do pacjentów z migotaniem przedsionków. Uzyskano dane NFZ dotyczące migotania przedsionków i częstych współistniejących chorób sercowo-naczyniowych, które porównano z danymi zebranymi manualnie z indywidualnej dokumentacji medycznej pacjentów w retrospektywnym rejestrze CRAFT. Dane z NFZ nie doszacowały odsetka pacjentów z migotaniem przedsionków (NFZ=83% vs. 100%), jednocześnie przeszacowały odsetek pacjentów z innymi chorobami układu krążenia, bądź też czynnikami ryzyka (wyższe ryzyko w skalach CHA₂DS₂-VASc oraz HAS-BLED), podkreślając różnice pomiędzy manualną analizą dokumentacji medycznej oraz danymi rozliczeniowymi.

W drugiej publikacji Pan lek. Cezary Maciejewski w ramach projektu AssistMED wraz ze współautorami opracował praktyczne narzędzie procesowania języka naturalnego, które automatycznie zapewnia kompleksową charakterystykę kliniczną pacjentów na podstawie elektronicznej dokumentacji medycznej. Narzędzie to wspiera pobieranie danych dotyczących stanu klinicznego, stosowanych leków wraz z ich dawkowaniem i parametrów echokardiograficznych. Sprawdzono działanie algorytmu w porównaniu z manualnym zbieraniem danych. W programie AssistMED zbadano obecność 56 stanów klinicznych, leków

z 16 grup wraz z dawkowaniem oraz 15 numerycznych parametrów echokardiograficznych w grupie 400 pacjentów hospitalizowanych na oddziale kardiologicznym. Nie wykazano istotnych różnic pomiędzy wyszukiwaniem z wykorzystaniem algorytmu a wyszukiwaniem manualnym przez człowieka. Na podstawie analizy jakościowej stwierdzono, że brak zgodności z manualnie zbieranymi danymi wynikał głównie z przypadkowych błędów algorytmu, błędów literowych w elektronicznej dokumentacji medycznej, skomplikowanego opisu dawkowania i braku zaawansowanej analizy kontekstu zastosowanego narzędzia.

Trzecia publikacja miała na celu przedstawienie praktycznego wykorzystania procesowania języka naturalnego w dużym badaniu retrospektywnym w porównaniu do manualnego pozyskiwania danych. W badaniu uwzględniono anonimową dokumentację wypisową pacjentów z oddziału kardiologicznego przeanalizowaną pod kątem włączenia do rejestru CRAFT. Zidentyfikowano odpowiednio 3030 i 3029 pacjentów za pomocą podejścia manualnego i z wykorzystaniem procesowania języka naturalnego, odzwierciedlając dokładność tej drugiej metody w wykrywaniu migotania przedsionków na poziomie 99,93%. Kompleksowa charakterystyka pacjentów metodą procesowania języka naturalnego była szybsza niż analiza manualna (3 godziny i 15 minut w porównaniu do 71 godzin i 12 minut). Skalkulowane wyniki w skalach CHA₂DS₂-VASc oraz HAS-BLED nie różniły się istotnie pomiędzy sobą w zależności od metody pozyskania danych. W przypadku większości danych stwierdzono niemal idealną zgodność pomiędzy danymi uzyskanymi za pomocą procesowania języka naturalnego i metodą manualną. Identyfikacja dawki dziennej była najmniej dokładną cechą pozyskiwaną z wykorzystaniem procesowania języka naturalnego.

2. **Poprawność redakcyjna rozprawy doktorskiej**

Przedstawiona praca doktorska ma zwięzłą strukturę zawartą na 65 stronach. W części opisowej rozprawy znajduje się 18 pozycji piśmiennictwa, w tym publikacje opublikowane w wiodących czasopismach naukowych, jak *Circulation*, *BMJ* oraz *Lancet*. To skłania do stwierdzenia, iż lek. Cezary Maciejewski posiada umiejętność doboru odpowiedniego piśmiennictwa. Praca doktorska posiada typowy dla tego rodzaju opracowań układ.

3. **Sugestie / uwagi**

Pomimo dużej wartości rozprawy doktorskiej z obowiązku recenzenta trzeba odnieść się do kilku drobnych aspektów dotyczących rozprawy doktorskiej:

Publikacja 1 (Maciejewski C., Ozierański K., Basza M., Łodziński P., Śliwczyński A., Kraj L., Krajsman M., Prado Paulino J., Tymińska A., Opolski G., Cacko A., Grabowski M., Balsam P. Administrative Data in Cardiovascular Research—A Comparison of Polish National Health Fund and CRAFT Registry Data. *Int. J. Environ. Res. Public Health*. 2022, 19(19), 11964).

1. W tekście pracy, podobnie jak w rozprawie skale CHA₂DS₂-VASc oraz HAS-BLED mogłyby być określane zgodnie z przyjętymi w literaturze (w tym w wytycznych

Europejskiego Towarzystwa Kardiologicznego) określeniami (zamiast „CHA2DS2VASc” i „HASBLED”). Ponadto w związku z relatywnie dużymi modyfikacjami skali HAS-BLED wynikającymi z przyjętej metodologii mogłaby być ona określona jako „modified HAS-BLED score”.

2. Zarówno na rycinie 2, jak również na rycinie 3 mogłyby być podane wartości P dla określenia istotności statystycznej uzyskanych wyników.
3. Wartościowe byłoby zbadanie różnic pomiędzy danymi zbieranymi prospektywnie, specyficznie dla chorych z migotaniem przedsionków na podstawie wywiadu oraz dostępnej dokumentacji medycznej, a szczegółowo raportowanymi danymi do Narodowego Funduszu Zdrowia. Czy Autor rozprawy planuje wykonanie takiego badania w przyszłości?

Publikacja 2 (Maciejewski C., Ozierański K., Barwiolek A., Basza M., Bożym A., Ciurla M., Krajsman M., Maciejewska M., Łodziński P., Opolski G., Grabowski M., Cacko A., Balsam P. AssistMED project: transforming cardiology cohort characterisation from electronic health records through natural language processing – algorithm design, preliminary results, and field prospects. Int J Med Inform. 2024 May;185:105380).

4. Tabela 1 oraz niektóre dane zamieszczone w suplemencie. Cyfry rozdzielono przecinkami zamiast kropkami (właściwymi dla określenia liczb dziesiętnych podawanych w języku angielskim).
5. Użyteczna byłaby walidacja uzyskanych wyników w innym ośrodku klinicznym. Jakie potencjalne trudności widzi Autor rozprawy w tym zakresie?

Publikacja 3 (Maciejewski C., Ozierański K., Basza M., Barwiolek A., Ciurla M., Bożym A., Krajsman M., Łodziński P., Opolski G., Grabowski M., Cacko A., Balsam P. Practical use case of natural language processing for observational clinical research data retrieval from electronic health records: AssistMED project. Pol Arch Intern Med . 2024 Mar 19:16704).


6. Przydatne byłoby uwzględnienie materiałów umieszczonych w suplemencie tego artykułu w tekście rozprawy (również w przypadku pozostałych prac, obecnie są dostępne online).

4. Ocena końcowa przedstawionej rozprawy doktorskiej

Drobne mankamenty rozprawy nie umniejszają jej istotnej wartości naukowej. Badania przeprowadzone przez Pana lek. Cezarego Maciejewskiego wskazują na ograniczenia danych, którymi dysponuje Narodowy Fundusz Zdrowia oraz mogą przyczynić się do optymalizacji działalności naukowej. Kolejną korzyścią wynikającą z przedstawionego cyklu prac jest możliwy ich wpływ na zalecenia dotyczące prowadzenia dokumentacji medycznej, aby zwiększać wiarygodność gromadzonych danych medycznych oraz wspierać i ułatwiać prace naukowe.

Wyrażam pozytywną ocenę przedstawionej mi rozprawy doktorskiej lek. Cezarego Maciejewskiego oraz stwierdzam, że spełnia ona warunki określone w *art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. 2018 poz. 1668)* i jednocześnie wnioskuję do Wysokiej Rady Dyscypliny Nauk Medycznych WUM o dopuszczenie Pana lek. Cezarego Maciejewskiego do dalszych etapów przewodu doktorskiego.

Ponadto chciałbym wystąpić o wyróżnienie przedstawionej rozprawy doktorskiej. Wniosek argumentuję podjęciem się przez doktoranta nowatorskich badań opartych na analizie elektronicznej dokumentacji medycznej i zastosowaniu m.in. procesowania języka naturalnego. Praca ta ma interdyscyplinarny charakter. Przedstawione badania wymagały ekspertyzy z zakresu różnych dziedzin oraz wiązały się z opracowaniem narzędzia do pozyskiwania danych klinicznych z elektronicznej dokumentacji medycznej. Wyróżnienie pracy uzasadniają również parametry bibliometryczne cyklu artykułów (łączy współczynnik Impact Factor czasopism, w których opublikowano 2 prace na czas złożenia pracy doktorskiej wynoszący 9,7 oraz sumaryczna punktacja MNiSW, w tym samym czasie, artykułów opublikowanych w tych czasopismach równa 480 punktów).

dr hab. n. med. Paweł Matusik
prof. UJ, specjalista kardiolog

2 6 4 4 6 6 9

dr hab. n. med. Paweł T. Matusik, prof. UJ
Klinika Elektrokardiologii, Instytut Kardiologii, Wydział Lekarski,
Uniwersytet Jagielloński Collegium Medicum
ul. Prądnicka 80, 31-202 Kraków
podpis i dane adresowe recenzenta

1 lipiec 2024 r.
data sporządzenia recenzji