

Warszawa, 10.02.2025

Prof. Dr hab. Henryk Rybinski
Instytut Informatyki Politechniki Warszawskiej
Henryk.rybinski@pw.edu.pl

Recenzja rozprawy doktorskiej mgra Artura Chudzika
p/t. „Machine Learning Methods for Parkinson’s Disease Datasets”

Dane ogólne

Przedstawiona do recenzji rozprawa jest napisana w języku angielskim. Składa się

1. ze streszczenia w języku polskim i angielskim,
2. z autoreferatu poświęconego omówieniu cyklu ośmiu publikacji,
3. ze streszczeń tych publikacji, oraz
4. z pełnych tekstów publikacji.

Publikacje stanowią zasadniczą część rozprawy, jednak obszerny autoreferat opracowany przez doktoranta (liczący 49 stron) jest również ważną częścią rozprawy.

W skład Autoreferatu wchodzi sześć rozdziałów oraz sekcja Referencji (60 prac, 6 stron): Rozdział 1 jest poświęcony wprowadzeniu w problematykę oraz prezentuje tezę rozprawy; Rozdział 2 omawia problematykę zbiorów danych wykorzystywanych w badaniach nad metodami sztucznej inteligencji w rozpoznawaniu chorób neurodegeneracyjnych, Rozdział 3 jest poświęcony modelom klasyfikacji choroby Parkinsona. W Rozdziale 4 doktorant omawia swój wkład w badania nad uczeniem maszynowym w zakresie choroby Parkinsona, natomiast w Rozdziałach 5 i 6 przedstawia odpowiednio dyskusję dotyczącą uzyskanych wyników oraz podsumowanie badań. Uzupełnieniem Autoreferatu jest sekcja abstraktów trzech dodatkowych współautorskich prac poświęconych problematyce chorób neurodegeneracyjnych.

Załączone publikacje, to:

- A1. Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, “DTI Helps to Predict Parkinson’s Patient’s Symptoms Using Data Mining Techniques,” in Intelligent Information and Database Systems, F. L. and H. T.-P. and T. B. Nguyen Ngoc Thanh and Gaol, Ed., Cham: Springer International Publishing, pp. 615–623, 2019.
- A2. W. Przybyszewski, A. Śledzianowski, A. Chudzik, S. Szlufik, and D. Koziorowski, “Machine Learning and Eye Movements Give Insights into Neurodegenerative

- Disease Mechanisms,” *Sensors*, vol. 23, no. 4, p. 2145, Feb. 2023, doi: 10.3390/s23042145.
- A3.W. Przybyszewski, A. Chudzik, S. Szlufik, P. Habela, and D. M. Koziarowski, “Comparison of Different Data Mining Methods to Determine Disease Progression in Dissimilar Groups of Parkinson’s Patients,” *Fundamenta Informaticae*, vol. 176, no. 2, pp. 167–181, Dec. 2020.
- A4.Chudzik, A. Szymański, J. P. Nowacki, and A. W. Przybyszewski, “Eye-Tracking and Machine Learning Significance in Parkinson’s Disease Symptoms Prediction,” in *Intelligent Information and Database Systems*, K. and S. A. and T. B. and C. S. Nguyen Ngoc Thanh and Jearanaitanakij, Ed., Cham: Springer International Publishing, pp. 537–547, 2020.
- A5.Chudzik, A. Śledzianowski, and A. W. Przybyszewski, “Machine Learning and Digital Biomarkers Can Detect Early Stages of Neurodegenerative Diseases,” *Sensors*, vol. 24, no. 5, p. 1572, Feb. 2024.
- A6.Chudzik, A. Drabik, and A. W. Przybyszewski, “Investigating the Impact of Parkinson’s Disease on Brain Computations: An Online Study of Healthy Controls and PD Patients,” in *Intelligent Information and Database Systems: 15th Asian Conference, ACIIDS 2023, Phuket, Thailand, July 24–26, 2023, Proceedings, Part II*, Berlin, Heidelberg: Springer-Verlag, pp. 235–246, 2023.
- A7.Chudzik and A. W. Przybyszewski, “Classification of Parkinson’s Disease Using Machine Learning with MoCA Response Dynamics,” *Applied Sciences*, vol. 14, no. 7, p. 2979, Apr. 2024.
- A8.Chudzik, J. P. Nowacki, and A. W. Przybyszewski, “Recognizing Patterns of Parkinson’s Disease using Online Trail Making Test and Response Dynamics – Preliminary Study,” *ICPR 2024. PREPRINT*.

Łączna objętość rozprawy (wraz z załączonymi publikacjami) wynosi ok. 203 stron. W sześciu pracach doktorant jest pierwszym autorem, prace [A2, A3, A5, A7] są opublikowane w czasopismach, z tego [A2, A3, A5] w czasopismach z listy JCR, pozostałe publikacje są rozdziałami z materiałów konferencji międzynarodowych, będących na liście MEiN (70 pkt). Do rozprawy dołączono oświadczenia autora o jego udziale (poświadczone przez współautorów).

1. Problem badawczy i jego znaczenie

1.1 Problem rozważany w rozprawie

Przytoczone publikacje dotyczą badań mających charakter interdyscyplinarny. Moja opinia ogranicza się jedynie do problemów dotyczących badań informatycznych w przedstawionych pracach.

Tematyka rozprawy ściśle wiąże się z metodami odkrywania wiedzy ze zbiorów danych oraz uczenia maszynowego. Dziedzina ta od pewnego czasu leży w obszarze

zainteresowań naukowców zajmujących się metodami sztucznej inteligencji, w szczególności metodami uczenia maszynowego. Badania w tej dziedzinie mają ogromne znaczenie, zarówno praktyczne jak też teoretyczne, przede wszystkim w kontekście systemów akwizycji wiedzy, uczenia maszynowego, a także metod wnioskowania indukcyjnego przez analizę przykładów.

Szczególnie w ostatnich kilku latach dziedzina ta nabiera ogromnego znaczenia, przede wszystkim w kontekście pojawiających się nowych możliwości zastosowań. Jednym z bardziej spektakularnych zastosowań sztucznej inteligencji jest medycyna, bowiem analiza danych biomedycznych stwarza istotne narzędzia pracy zarówno w naukach medycznych, jak też w praktycznych zastosowaniach.

Opiniowana praca jest kontynuacją badań zespołu profesora Przybyszewskiego, mieszczących się w tym właśnie nurcie i dotyczy problematyki odkrywania wiedzy na potrzeby diagnozowania stopnia zaawansowania choroby Parkinsona. Doktorant koncentruje się nad możliwościami zbudowania nowych modeli klasyfikacji i predykcji, których celem jest dokładniejsze diagnozowanie stanu choroby Parkinsona. Proponowane modele są oparte na technikach uczenia maszynowego, eksploracji danych oraz metodach zbiorów przybliżonych.

Podsumowując tę część rozprawy, stwierdzam, że tematyka rozważana przez doktoranta jest niezwykle ważna i niewątpliwie może być przedmiotem rozprawy doktorskiej.

1.2 Cel rozprawy

W autoreferacie autor zwraca uwagę na istotne problemy związane z różnorodnością objawów choroby Parkinsona, szczególnie w różnych fazach zaawansowania choroby oraz na nakładanie się tych samych objawów w różnych chorobach neurologicznych. W związku z tym jako zasadniczy cel pracy doktorant definiuje opracowanie nowych metod analizy danych, prostych w zastosowaniu, pozwalających wcześniej diagnozować chorobę Parkinsona oraz różnicować ją z innymi chorobami neurodegeneracyjnymi. Obszarem badań dla znalezienia takich metod są dla doktoranta metody eksploracji danych oraz uczenia maszynowego. W szczególności autor koncentruje się na trzech podejściach do zbierania i analizy danych:

1. obrazowanie tensora dyfuzji (DTI),

2. śledzenie ruchu gałek ocznych,
3. testy kognitywne online, powiązane z wybranymi algorytmami ML zaprojektowanymi do modelowania wzorców choroby Parkinsona w tych danych.

1.3 Charakter rozprawy i znaczenie praktyczne badań

Praca ma przede wszystkim charakter eksperymentalny, jest jednak mocno poparta technicznym warsztatem autora, pozwalającym konstruować nowe rozwiązania techniczne. Pozytywnie oceniam też warsztat teoretyczny doktoranta. Podstawowe narzędzia stosowane w rozprawie wiążą się z metodami statystycznymi oraz metodami uczenia maszynowego, (w tym m.in. metodami konwolucyjnych sieci neuronowych (CNN), losowych lasów (*random forest*), czy metodami zbiorów przybliżonych Pawłaka (RST).

Propozycje autora są wsparte zaawansowanymi implementacjami oraz eksperymentami. Wskazują one na

1. solidny warsztat doktoranta w zakresie narzędzi uczenia maszynowego;
2. duży potencjał praktycznego wdrażania opracowanych algorytmów;

2. Wkład autora

Wkład autora obejmuje szereg interesujących elementów związanych z metodami eksploracji danych i uczeniem maszynowym na potrzeby analizy danych otrzymywanych w procesie diagnozowania pacjentów:

1. W pracy A1 wkładem autora jest zweryfikowanie przydatności metod zbiorów przybliżonych do analizy danych otrzymywanych z obrazowania tensora dyfuzji. Autor utworzył model, który wykrywa zaburzenia w odręcznym piśmie. Zastosowane podejście pozwoliło odkrywać zmiany mikrostrukturalne w mózgu u osób z chorobą Parkinsona.
2. W pracach A2, A3 i A4 doktorant pokazuje możliwości wykorzystania metod uczenia maszynowego z użyciem danych uzyskiwanych z obserwacji gałek ocznych. Opracowane metody pozwalają identyfikować cechy specyficzne u osób z chorobą Parkinsona. Doktorant bada możliwości wykorzystania m.in. takich technik ML, jak naiwna metoda Bayesa, drzewa decyzyjne, klasyfikatory LSVC (Linear Support Vector Classifiers) i KNN (K-Nearest Neighbors) i RF (Random

Forest). Metody te są porównane z metodami bazującymi na metodach zbiorów przybliżonych. Autorowi udało się uzyskać odpowiedź na pytanie, w jakich sytuacjach poszczególne podejścia są lepsze. Jednocześnie doktorant potwierdził fakt, że dane uzyskane z badania ruchu gałek ocznych są niezwykle ważne w diagnozowaniu zaawansowania choroby Parkinsona.

3. W pracach A5-A8 doktorant zajmuje się najnowszym podejściem, polegającym na wykorzystaniu danych z internetowych testów kognitywnych do identyfikacji motorycznych i poznawczych objawów choroby Parkinsona. W celu przeprowadzenia badań doktorant zaprojektował i zrealizował specjalizowany test zbierający dane związane z badaniami motoryki i możliwości poznawczych. W pracy A5 prezentowane jest nisko kosztowe efektywne narzędzie do wykrywania wczesnych symptomów choroby neurodegeneracyjnej. W A6 prezentowany jest model doktoranta wykorzystujący regresję logistyczną, charakteryzujący się dużą dokładnością. Dodatkowe badania dotyczące możliwości wykorzystania innych narzędzi ML w oparciu o dane MoCA (*Montreal Cognitive Assessment*) zostały zaprezentowane w A7. Pokazały one dużą przydatność metody zbiorów przybliżonych. Natomiast praca A8 zaprezentowała wyniki badań mające na celu znalezienia sposobów rozróżnienia symptomów różnych faz zaawansowania choroby Parkinsona – średniej i zaawansowanej. W efekcie uzyskane wyniki pokazały, że stworzone przez doktoranta nieinwazyjne metody badania stanowią narzędzia do wczesnego wykrywania i diagnozowania symptomów choroby.

Zaproponowane rozwiązania mają charakter techniczny, wymagały one jednak od autora dobrego warsztatu naukowego i znajomości narzędzi sztucznej inteligencji i eksploracji danych. Wartość opracowanych metod analizy danych na potrzeby diagnostyki w chorobie Parkinsona została potwierdzona publikacjami zamieszczonymi w pracy. Warto też zwrócić uwagę na to, że uzyskano interesujące wyniki za pomocą prostych metod zbierania danych z wykorzystaniem rzeczywistości wirtualnej.

3. Wiedza kandydata

W rozprawie doktorant wykazuje się dużą wiedzą, przede wszystkim w zakresie uczenia maszynowego, w metodach analizy danych, w tym tych bazujących na zbiorach

przybliżonych oraz rozmytych-przybliżonych. Niezależnie od rozprawy na uwagę zasługuje także całokształt dorobku doktoranta (14 publikacji w GS, cytawalność 73 pozycji, i indeks H=4).

4. Inne uwagi

Autoreferat jest napisany dobrym angielskim. Edycja rozprawy nie budzi zastrzeżeń. Zamieszczenie w autoreferacie Słownik jest pomocny w trakcie zapoznawania się z rozprawą. Cytowana literatura jest aktualna.

Podstawą rozprawy jest 8 publikacji, ponadto autor ponadto wyróżnia 3 publikacje „pomocnicze”, w których jest współautorem. Moim zdaniem cel tego zabiegu jest niejasny, wprowadza nieco zamieszania. Wystarczyło dodać te publikacje do sekcji References. Tym bardziej, że uważam, iż przedstawione artykuły A1-A8 stanowią dorobek, który jest dobrą podstawą do rozprawy doktorskiej.

W trakcie czytania rozprawy nasunęły mi się uwagi:

Jeden z nurtów prac doktoranta jest związany z analizą danych otrzymywanych w badaniach ruchu gałek ocznych. Nurt ten był podstawowym przedmiotem rozprawy doktorskiej Alfreda Śledzianowskiego, w związku z tym zabrakło mi w autoreferacie odniesienia do doktoratu Śledzianowskiego. Nasuwa mi się w związku z tym kilka pytań:

1. jakie są różnice pomiędzy modelami zastosowanym w rozprawie dra Śledzianowskiego oraz w pracy doktoranta bazującymi na RST i FRST;
2. Czy modele te są budowane w oparciu o te same dane i/lub metody;
3. Czy opracowane modele są konkurencyjne w stosunku do wcześniejszych, czy się uzupełniają.

Mam też pytanie, jak doktorant widzi możliwości wykorzystania modeli LLM w badaniach związanych z chorobami neurodegeneracyjnymi.

5. Podsumowanie

Pozytywnie oceniam przedstawione w doktoracie artykuły i wkład doktoranta. Rozprawa spełnia wymagania zawarte w obowiązujących przepisach dotyczących rozpraw doktorskich, wnoszę zatem o dopuszczenie mgra Artura Chudzika do publicznej obrony.