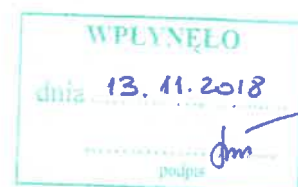


Prof. dr hab. Zbysław Tabor
Instytut Teleinformatyki
Wydział Fizyki, Matematyki i Informatyki
Politechnika Krakowska
ul. Warszawska 24, 31-155 Kraków
ztabor@pk.edu.pl

Kraków, 12 listopada 2018



RECENZJA

rozprawy doktorskiej Pani mgr inż. Darii Hammerling pt.:

**„WYKORZYSTANIE SYGNAŁU MOWY JAKO ŹRÓDŁA INFORMACJI DIAGNOSTYCZNEJ,
KONTROLNEJ I PROGNOZTYCZNEJ W WYBRANYCH PROBLEMACH MEDYCZNYCH ZWIĄZANYCH
Z OTOLARYNGOLOGIA”**

wykonanej pod kierunkiem promotora prof. dra hab. inż. Janusza Gajdy
i promotora pomocniczego dra inż. Andrzeja Skalskiego
na Wydziale Elektrotechniki, Automatyki, Informatyki i Inżynierii Biomedycznej
Akademii Górniczo-Hutniczej w Krakowie

I. Ogólna charakterystyka podjętego przez Doktorantkę problemu badawczego

W przedłożonej mi do oceny rozprawie Doktorantka przedstawia projekt systemu wspomagającego diagnostykę medyczną, wykorzystującego uczenie maszynowe i działającego w oparciu o analizę sygnałów akustycznych rejestrowanych w trakcie wypowiedzania przez osobę badaną zadanych samogłosek. Problem jest podjęty w związku z hipotezą, zgodnie z którą cechy sygnału są zaburzane przez zmiany patologiczne towarzyszące niektórym schorzeniom (na przykład laryngologicznym lub chorobie Parkinsona). Jeśli hipoteza jest prawdziwa, cechy sygnału akustycznego mogą być użyte do uzupełnienia obrazu diagnostycznego choroby, do oceny odpowiedzi na stosowaną terapię, a także do oceny przysłego stanu pacjenta.

Spśród tematów poruszonych przez Doktorantkę dla recenzenta szczególnie interesujące jest zastosowanie proponowanych przez Doktorantkę technik do oceny stanu pacjenta z chorobą Parkinsona oraz do predykcji zmian tego stanu.

Skuteczny system oceny stanu pacjenta z chorobą Parkinsona stanowiłby niezwykle przydatne narzędzie, wspierające podejmowanie decyzji terapeutycznych dotyczących w szczególności stosowania lewodopy.

Dlatego podjęty temat badawczy oceniam jako ważny, interesujący i wpisujący się w dyscyplinę biocybernetyki i inżynierii biomedycznej, w szczególności w obszar wykorzystania uczenia maszynowego w diagnostyce medycznej.

II. Ogólna charakterystyka rozprawy

W rozprawie poddane są analizie dane zgromadzone w dwóch bazach danych. Pierwsza z tych baz liczy 1410 nagrań samogłosek 'a', 'i' oraz 'u' wypowiedzianych przy różnych intonacjach. Analizowane nagrania pochodzą od osób zdrowych oraz cierpiących na trzy różne schorzenia laryngologiczne. Druga baza zawiera 27 nagrań samogłosek 'a', 'e', 'i', 'o' oraz 'u' wypowiedzianych przez osoby cierpiące na chorobę Parkinsona przez okres od 2 do 21 lat. Analiza danych przebiega według schematu: akwizycja, ekstrakcja cech, redukcja wymiarowości, klasyfikacja/regresja. Ekstrakcja cech jest oparta o analizę częstotliwościową i czasowo-częstotliwościową. Do redukcji wymiarowości używana są kPCA lub autoasocjacyjne sieci neuronowe. W problemie analizy stanu pacjentów z chorobą Parkinsona redukcja wymiarowości jest wykonywana „manualnie” poprzez wybór zmiennych objaśniających, które są najlepiej skorelowane ze zmienną objaśnianą – Doktorantka wyznacza w tym celu wartości współczynników korelacji Pearsona i Spearmana. Klasyfikacja jest wykonywana przy użyciu lasów losowych (klasyfikacja dwuklasowa: osoby zdrowe i osoby ze schorzeniami laryngologicznymi) i sieci neuronowych (klasyfikacja wieloklasowa osób zdrowych i osób cierpiących na trzy schorzenia laryngologiczne). Regresja jest wykonywana przy użyciu SVR (estymacja punktacji UPDRS-III w oparciu o parametry akustyczne) i sieci neuronowych (estymacja parametrów sygnału akustycznego pacjenta po 180 minutach od podania leku w oparciu o parametry akustyczne zarejestrowane czterech nagrań zarejestrowanych w momencie podania leku, 30, 60 i 120 minut po podaniu leku). Trening i optymalizacja klasyfikatorów/regresorów są oparte o metodę Out of Bag (lasy losowe) lub walidację krzyżową (SVR i sieci neuronowe). W przypadku sieci neuronowych zbiór danych jest dzielony na trzy części: dane treningowe, walidacyjne i testowe. Wyniki analizy są podsumowane przy użyciu standardowych w uczeniu maszynowym i właściwie dobranych miar (dokładność, precyzja, czułość).

Wyniki uzyskane przez Doktorantkę zostały przedstawione również w szeregu publikacji, w tym w artykułach w międzynarodowych recenzowanych czasopismach oraz w artykułach prezentowanych na międzynarodowych konferencjach (pozycje bibliograficzne [4], [17], [113]-[115], [118]-[121]).

III. Uwagi ogólne do rozprawy

Być może najpoważniejszym obszarem zastosowań wyników uzyskanych przez Doktorantkę jest wsparcie w kontroli stanu pacjentów z chorobą Parkinsona. Indywidualny dobór planu terapii jest niezbędny, aby zredukować objawy choroby. Dobór ten jest często procesem długotrwałym między innymi (lub przede wszystkim) z powodu trudności organizacyjnych służby zdrowia. Wszelkie systemy teleinformatyczne wspomagające skutecznie ocenę aktualnego stanu pacjenta w powiązaniu z aktualnym planem terapii mogą poprawić jakość życia pacjentów z chorobą Parkinsona i ich rodzin. W tym kontekście należy zadać Doktorantce dwa pytania.

Po pierwsze, opisane badania wpasowują się w obszar „spersonalizowanej opieki medycznej”. Doktorantka na podstawie analizy sygnału akustycznego i metod uczenia maszynowego ocenia stan pacjenta w skali ilościowej używanej przez lekarzy neurologów. Ocena stanu pacjenta X w schemacie UPDRS-III jest wykonywana w oparciu o dane treningowe pochodzące od grupy pacjentów, do których X nie należy, co wynika z przyjętej przez Doktorantkę metodyki badawczej. Czy przyszłe badania, o ile zostaną podjęte, nie powinny skupić się na innym schemacie eksperymentu, w którym sygnały akustyczne od pacjenta X są nagrywane wielokrotnie w trakcie różnych sesji i na podstawie wielu nagrań, w oparciu o metody uczenia maszynowego estymowana jest punktacja UPDRS-III pacjenta X? Czy przy takim modelu eksperymentu dokładność estymacji punktacji UPDRS-III nie byłaby wyższa?

Po drugie, w spersonalizowanej medycynie bardzo aktualnym obszarem badań są tzw. „wearable sensor networks”. Tego typu sieci sensorów, oparte na przykład na akcelerometrach, dostarczyłyby łatwiejszych do uzyskania, niż sygnał akustyczny informacji o mimowolnych ruchach pacjenta, które są skojarzone z chorobą Parkinsona. Czy Doktorantka planuje rozszerzenie swoich przyszłych badań o rejestrację sygnałów z tego typu czujników?

Mam również ogólną uwagę do układu pracy. Układ przyjęty przez Doktorantkę przypomina układ artykułu skierowanego do odbiorcy, który jest

specjalistą w szczegółowym obszarze, którego dotyczy rozprawa. Często jednak Doktoranci konstruują plan swoich rozpraw tak, aby były one zrozumiałe dla szerszego grona odbiorców. W takim przypadku plan rozprawy Doktorantki powinien wyglądać następująco:

1. Wprowadzenie i motywacja
2. Anatomiczne podstawy procesu mowy
3. Akwizycja i preprocessing sygnału akustycznego
4. Cechy ilościowe opisujące sygnał akustyczny
5. Metody klasyfikacji i regresji w analizie sygnałów
6. Analiza aktualnego stanu wiedzy
7. Zastosowanie metod uczenia maszynowego w diagnostyce otolaryngologicznej (opis całego eksperymentu, tzn, opis bazy danych, analizowanych sygnałów akustycznych itd.)
8. Zastosowanie metod uczenia maszynowego w diagnostyce choroby Parkinsona (opis całego eksperymentu, tzn, opis bazy danych, analizowanych sygnałów akustycznych itd.)
9. Podsumowanie

Przy obecnym kształcie analiza stanu wiedzy pojawia się na początku i są w niej cytowane parametry, wprowadzane przez Doktorantkę wiele stron dalej, w związku z czym czytającemu trudno jest ocenić znaczenie wyników już opublikowanych. Podobnie, elementy preprocessingu (autoasocjacyjne sieci, używane do redukcji wymiarowości) są opisane razem z metodami regresji.

IV. Uwagi szczegółowe do rozprawy

W pracy Doktorantka nie ustrzegła się błędów językowych i niejasności, które wymieniam poniżej. Podkreśleniem oznaczam te moje zastrzeżenia, które mogą wymagać więcej uwagi Doktorantki.

1. Str. 11, linia 8: „Zazwyczaj takie systemy wykorzystują odległość [...]” - jaką odległość i w jakiej przestrzeni?
2. Str. 12, linia 6 od dołu: „estymacja punktacji w skali UPDRS [...]” - celem pracy jest estymacja punktacji tylko w obrębie części III UPDRS.
3. Str. 13, Analiza stanu wiedzy: pojawiają się nazwy cech sygnałów, cechy te nie zostały jeszcze zdefiniowane.

4. Str. 13, linia 9 od dołu: odniesienie do cech „parametrycznych” i „nieparametrycznych”. Czy Doktorantka ma na myśli cechy ilościowe i jakościowe?
5. Str. 14: liczne błędy językowe np. „Parametrami opisującymi sygnał [...] były parametry [...]”, „Nagrania wykorzystane na poczet tej pracy [...]”, „baza została zaimplementowana [...]”
6. Str. 15: dokładność klasyfikacji jest definiowana w 10 linii od dołu, a użyta po raz pierwszy w 6 linii od góry.
7. Str. 17: pojawia się niezdefiniowany (choć oczywisty) skrót PD. Skrót ten jest definiowany na str. 26. W pracy używane jest zamiennie „choroba Parkinson'a” lub „choroba Parkinsona”.
8. Str. 18: „W wyniku 10-krotnego sprawdzianu krzyżowego otrzymano [...] różnicę 2 punktów [...]”. Nie jest jasne, czy te dwa punkty, to odchylenie standardowe, średni błąd, błąd systematyczny...
9. Str. 20, nad rysunkiem: wprowadzone pojęcie fonacji, zdefiniowane wiele stron później.
10. Str. 21, 4 linia od dołu: błąd językowy „Głównym jego parametrem jest częstotliwością”.
11. Str. 21, 1 i 2 linia od dołu: „zależą od [...] ich masy [...]. Parametry te mogą być modulowane [...]”. Chodzi o modyfikację, a nie modulację? Czy masa na pewno może być modyfikowana?
12. Str. 22, 2 linia w rozdziale 2.1: błąd językowy „do całkowitego jego zaniku i możliwości porozumiewania się”. Na marginesie: możliwości porozumiewania się nie obejmują tylko mowy.
13. Str. 22, linie 3 i 5 w rozdziale 2.1; najpierw jest mowa o czynnikach funkcjonalnych, a następnie o funkcyjnych.
14. Str. 23, linia 9: błąd literowy „Dysfonie czynnościową”.
15. Str. 26, linia 10 od dołu: niejasne „Większość obszarów mózgu [...] można badać poprzez analizę części korowej mózgu [...]”. Na czym polega analiza części korowej mózgu?
16. Str. 28, linia 6 od dołu: niejasne „Przedstawiona dysfunkcja pokazuje potencjał metod”.
17. Str. 28, linia 4 od dołu: niejasne „Analizę należy wzbogacić [...] w celu kontroli wszystkich zaburzeń [...]”. Co doktorantka ma na myśli, pisząc o kontroli i czy na pewno istnieje sposób kontroli wszystkich możliwych zaburzeń?

-
18. Str. 31, linia 9 od dołu: pojawia się odwołanie do stanu off, jako momentu zupełnego wysycenia leków. Ta definicja jest niejasna. W jaki sposób jest sprawdzane wystąpienie stanu off? Czy badana jest krew pacjenta?
 19. Str. 32, linia 5 od dołu: wydaje się, że zdanie „Celem analizy [...] tonu krtaniowego” jest wyrwane z kontekstu.
 20. Str. 34, linia 1: „konieczne było wyselekcjonowanie cech”. Przy zastosowaniu głębokiego uczenia selekcja cech być może nie byłaby konieczna.
 21. Str. 34, 18 linia od dołu: jest mowa o „produkcji” samogłosek.
 22. Str. 34, 3 linia od dołu: „Cepstrum definiowane jest jako logarytm dziesiętny widma [...] poddanej odwrotnej transformacie [...]”. Błąd językowy i formalny – logarytm jest definiowany dla liczb, a nie dla funkcji.
 23. Str. 35, 1 linia: „Algorytm oparty na analizie cepstralnej [...]” - o jaki algorytm chodzi?
 24. Wzór (3.1) – nadmiarowy nawias okrągły, poza tym logarytmu nie stosuje się do funkcji.
 25. Str. 35, linia 2 pod wzorem (3.1): „poddany procesowi segmentacji i okienkowania sygnał”. Ani segmentacja, ani okienkowanie nie zostały w pracy w sposób jawny zdefiniowane.
 26. Wzór (3.2) – pojawia się niezdefiniowana liczba ramek R .
 27. Str. 35, 4 linia pod wzorem (3.2) - „z wygładzeniem dla zdefiniowanej liczby okresów”. Nieprecyzyjna definicja parametrów.
 28. Wzór (3.3) – nadmiarowy nawias okrągły.
 29. Str. 35, linia 1 pod wzorem (3.3) - „z wygładzaniem dla kolejnych okresów [...]”. Ponownie nieprecyzyjna definicja parametrów.
 30. Wzór (3.4) – przy definicji energii podanej przez Doktorantkę, energia będzie rosła przy zwiększaniu częstotliwości próbkowania lub przy zwiększaniu liczby okien. Czy na pewno tak powinno być?
 31. Str. 35, 3 linia od dołu - pojawia się odwołanie do widma czasowo-częstotliwościowego $G(t,f)$. W jaki sposób widmo to zostało uzyskane? Czy stosowano transformate falkowa? Jeśli tak, to jaka?
 32. Str. 38, 3 akapit: opis metody LPC nie jest wystarczający do zrozumienia zasady jej działania.
 33. Str. 41, filtracja melowa: podane informacje nie definiują jednoznacznie procesu filtracji: filtry zachodzą na siebie, ale nie wiadomo jak bardzo, filtry są trójkątne, ale w skali logarytmicznej czy pierwotnej?

-
34. Str. 42, logarytmowanie sygnału pozwala na „rozdzielenie poszczególnych składowych”. Niejasne w jaki sposób logarytmowanie pozwala rozdzielać składowe.
35. Str. 42, rys. 3.5 i 3.6 – we wzorze (3.14) współczynniki MFCC są indeksowane liczbami naturalnymi, a na rysunkach czasem, który jest liczbą rzeczywistą.
36. Str. 43, drugi akapit: wprowadzane są kolejne cechy sygnału, definicje mogłyby być sformalizowane.
37. Str. 44, 5 linia w rozdziale 3.3.1: „z punktu widzenia największej wariancji” - nieprecyzyjne.
38. Wzór (3.23) – wprowadzona macierz K i wektor α , na kolejnej stronie błąd w definicji elementów tej macierzy. Nie ma definicji wektora α , jest definicja wektora a .
39. Str. 46, 1 linia pod wzorem (3.25) - „wykorzystano funkcje Gaussa”. Rozumiem, że ponieważ wektory x_p są wielowymiarowe, to funkcja Gaussa operuje na takich właśnie wektorach? W takim przypadku wykorzystuje się raczej macierz kowariancji, a nie szerokość jadra σ .
40. Wzór (3.26) – dlaczego po prawej stronie równania jest zależność od indeksu i , a po lewej już nie?
41. Str. 47, linia 6 pod rysunkiem: w jakim sensie „Lasy losowe nie wymagają wiedzy eksperckiej”?
42. Str. 48, drugi akapit: „zbiór wejściowy jest dzielony na dwa podzbiory: zbiór uczący oraz zbiór OOB”. Rozumiem, że Doktorantka wykorzystwała w budowie lasu losowego techniki baggingu wobec czego każdy element zbioru danych należy do zbioru treningowego ok. 2/3 drzew losowych, a więc może być użyty do testowania pozostałych 1/3? Wobec tego sformułowanie użyte przez Doktorantkę nie jest precyzyjne.
43. Str. 50, 1 linia pod rysunkiem: „parametry względem ich hierarchii” - o jaką hierarchię chodzi?
44. Str. 50, 2 i 3 linia pod rysunkiem: „uszeregowane są zgodnie z malejącą wariancją parametrów akustycznych” - chyba chodzi o część wariancji w zbiorze danych, wyjaśnianą przez kolejną składową główną.
45. Str. 50, ostatnie dwie linie: „Próg akceptacji [...] przyjęto równy 0.73”. Skąd wzięła się liczba 0.73? W zagadnieniach klasyfikacji wybór progu odciecia jest wynikiem dedykowanych analiz, na przykład analizy ROC. Tego typu analizy

służą wybraniu optymalnych (w zadanym przez użytkownika sensie) wartości parametru odcięcia. Czy taka analiza została przeprowadzona?

46. Str. 52 i następne – opis SVR nie jest niepotrzebnie szczegółowy i powtarza podręcznikowe informacje.
47. Str. 55 – opis sposobu wykorzystania sieci neuronowej w regresji jest z kolei zbyt ubogi.
48. Str. 59, linia 7 pod rysunkiem: dokładność klasyfikacji jest definiowana po raz kolejny.
49. Str. 59, linia 8 i następne pod rysunkiem: „Zwiększanie wymiaru wektora składowych głównych [...]. Wpływ σ na dokładność klasyfikacji [...]”. Czy chodzi o wymiar składowych głównych, które są wektorami? W jaki sposób zwiększano ten wymiar. Czy zwiększano również liczbę składowych głównych użytych w klasyfikacji? Jeśli badano wpływ wymiaru składowych głównych i σ na dokładność klasyfikacji, to na rys. 4.1 brakuje pętli sprzężenia od klasyfikacji do kPCA.
50. Str. 60, drugi akapit – mowa jest o wartości predykcyjnej składowych głównych, które są nieliniowymi funkcjami cech sygnału akustycznego oraz o indeksach Giniego tych składowych. Skąd wobec tego Doktorantka wzięła indeksy Giniego dla np. energii sygnału lub współczynnika jitter, skoro po wykonaniu kPCA one nie występują bezpośrednio w analizie.
51. Str. 62, 5 linia pod Tabelą 4.2 - „Precyzja określa stosunek [...]”. Zwykle używa się innej definicji precyzji: $\text{true positive} / (\text{true positive} + \text{false positive})$, a więc w mianowniku używa się innej wielkości, niż wskazana przez Doktorantkę.
52. Str. 63, linia 10 pod tabelą: „badanie dokładności klasyfikacji, było wyższe [...]” - powinno być „dokładność klasyfikacji była wyższa”.
53. Str. 64, przedostatni akapit – jest mowa o treningu sieci neuronowej i o jej architekturze. Trening powinien być opisany w części dotyczącej metodyki. Nie jest wiadomo jakie przesłanki wpływały na wybór architektury.
54. Str. 66. drugi akapit – czy w celu przeprowadzenia walidacji krzyżowej stosowano randomizację stratyfikowaną?
55. Tabela 4.6 i następne – wyniki prezentowane niekonsystentnie albo jako procenty, raz jako ułamki dziesiętne.
56. Str. 70, rozdział 4.2 – nie jest jasne, jakiej sytuacji dotyczyła kontrola i regresja SVR. Czy pacjentów w stanie off. czy w jakimś czasie po podaniu leków?

-
57. Str. 71, linia 8: jest mowa o „skuteczności klasyfikacji pacjentów”, podczas gdy problem opisany w rozdziale 4.2 dotyczy oceny stanu pacjenta, a nie klasyfikacji.
58. Tabela 4.13 – błąd w formatowaniu.
59. Str.73, pierwszy akapit: wyniki procentowe powinny być przedstawione w odniesieniu do wartości w Tabeli 3.2 na stronie 33, a nie do maksymalnej wartości punktowej w UPDRS-III. O ile więc błąd 9.5 to tylko 8.8% maksymalnej wartości 108 w UPDRS-III, to jest to jednak ok. 25% średniej UPDRS-III w stanie off dla danych w Tabeli 3.2. Może należałoby zaprezentować wykres wartości przewidywanych względem rzeczywistych, np. wykres Blanda-Altmana? Należałoby również wykonać test istotności różnic między obiema metodami p. test Wilcoxon. Bez tego typu informacji trudno jest ocenić jakość estymacji stanu pacjenta w oparciu o analizę sygnału akustycznego.
60. Str. 75, linia 10: „Dane ze zbioru uczącego były wykorzystane [...]”. Informacja podręcznikowa – powinna znaleźć się w ogólnym opisie metod, a nie w opisie wyników eksperymentu.
61. Rysunek 4.9. Czy można podać jeden model regresji uzyskany w wyniku walidacji krzyżowej, który mógłby być zastosowany dla całego zbioru danych? Innymi słowy, jak duża jest wariancja dla współczynników regresji liniowej uzyskanych przy użyciu sieci neuronowej.
62. Tabele 4.14 i 4.15 – wyniki procentowe powinny być wyznaczone w odniesieniu do wyników indywidualnego pacjenta, a nie maksymalnej wartości dla UPDRS-III. Cztery punkty różnicy w badaniu 180 min po podaniu lekařstwa to ok. 15% średniej wartości UPDRS-III. Jak w punkcie 59 - może należałoby zaprezentować wykres wartości przewidywanych względem rzeczywistych, np. wykres Blanda-Altmana? Należałoby również wykonać test istotności różnic między obiema metodami p. test Wilcoxon.
63. Str. 80, 5 linia pod rysunkiem: „odchylenie standardowe wskazujące na stabilność otrzymanej wartości” - w jakim sensie odchylenie standardowe wskazuje na stabilność wartości? Odchylenie jest wyznaczane dla grupy pacjentów, a nie dla zmieniającego się sygnału od jednego pacjenta wobec czego wydaje się, że nie może wskazywać na stabilność.
64. W bibliografii Doktorantka nie stosuje jednolitych konwencji. Pozycja [65] zastanawiająca.

V. Wnioski

Oceniając wstępną część rozprawy, na którą składają się rozdziały od 1 do 3 stwierdzam, że Doktorantka przeprowadziła analizę źródłową podjętego tematu badawczego w zakresie świadczącym o ogólnej wiedzy teoretycznej Doktorantki w dyscyplinie biocybernetyki i inżynierii biomedycznej.

Oceniając część rozprawy zawierającą prezentację i dyskusję wyników badań (rozdział 4) stwierdzam, że Doktorantka przedstawiła oryginalne rozwiązanie problemu badawczego, obejmującego przede wszystkim zastosowanie metod uczenia maszynowego do predykcji stanu i zmian stanu pacjenta z chorobą Parkinsona na podstawie analizy sygnałów akustycznych rejestrowanych w trakcie wypowiedzania przez pacjenta wybranych samogłosek. Proponując oryginalne rozwiązanie Doktorantka zademonstrowała umiejętność samodzielnego prowadzenia pracy naukowej. Konkluzje te potwierdzają dodatkowo publikacje doktorantki, tj. pozycje bibliograficzne [4], [17], [113]-[115], [118]-[121].

Podsumowując, stwierdzam, że rozprawa doktorska Pani mgr. inż. Darii Hammerling pt.: „Wykorzystanie sygnału mowy jako źródła informacji diagnostycznej, kontrolnej i prognostycznej w wybranych problemach medycznych związanych z otolaryngologią” spełnia warunki określone w art. 13 ust. 1 Ustawy o stopniach i tytule naukowym i wnioskuję o dopuszczenie jej do publicznej obrony.

Zbisto Tle