

Prof. dr hab. inż. Jacek Rumiński,
Katedra Inżynierii Biomedycznej
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

Gdańsk, 05.01.2024 r.

Recenzja

pracy doktorskiej mgr. inż. Jakuba Piskozuba

pt. „A data glove with a reduced number of sensors for the recognition of Polish Sign Language letters”.

Promotor: prof. dr hab. inż. Paweł Strumiłło

Recenzja została przygotowana w odpowiedzi na pismo Pani prof. dr. hab. inż. Ewy Piętki, Przewodniczącej Rady Dyscypliny Inżynieria Biomedyczna, Politechniki Śląskiej (RDIB.002.120.2023 z dnia 28 listopada 20213 r.) z prośbą o wykonanie recenzji rozprawy mgr. inż. Jakuba Piskozuba.

1. Tematyka, cele i tezy rozprawy

Tematyka pracy dotyczy problemu rozpoznawania gestów i liter języka migowego wykonywanych dłonią. Przedstawione aspekty pracy są bardzo aktualne. Dotyczą zarówno problemu projektowania interfejsów interakcji człowiek-komputer jak i nowoczesnej analizy danych, w szczególności z wykorzystaniem metod uczenia maszynowego.

Doktorant wskazał cel pracy, zarówno w streszczeniu jak i głównej części dokumentu. Pan mgr inż. Jakub Piskozub sformułował również tezę pracy zapisaną w rozdziale pierwszym (storna 5) jako:

„It is possible to effectively recognize the letters of the Polish Sign Language Alphabet using a sensor glove with a reduced number of sensors, which can simplify the construction of the glove and offer its improved ergonomics”.

Teza jest interesująca i wskazuje główne aspekty pracy, uważam jednak, że została zdefiniowana zbyt ogólnie i wymagałaby doprecyzowania. Rękawice sensoryczne do rozpoznawania znaków języka migowego były opracowane w wielu ośrodkach w Polsce (m.in. Politechnika Białostocka, Politechnika Gdańska, Politechnika Łódzka) demonstrując możliwość efektywnego rozpoznawania liter. Zatem określenie „effectively recognize” jest zbyt ogólne. Oczywiście ważnym aspektem tezy jest wskazana rola „zredukowanej liczby czujników”, ale Doktorant powinien wskazać albo w założeniach, albo w tezach szczegółowych, co stanowi

odniesienie do „zredukowanej liczby czujników”. Stosowne odniesienie pojawia się w dalszej części pracy.

2. Zawartość rozprawy

Rozprawa napisana jest w języku angielskim i składa się z sześciu głównych rozdziałów, spisu literatury (jako rozdział 7) oraz suplementu (jako rozdział 8), wskazującego link do repozytorium kodu.

W rozdziale pierwszym Autor właściwie przedstawia motywację podjęcia tematu oraz przedstawia tezę jak i przegląd pracy.

Rozdział drugi prezentuje wybrane aspekty teoretyczne i techniczne w zakresie tematyki rozprawy, w szczególności informacje dotyczące Polskiego Języka Migowego oraz przegląd prac badawczych i aspektów związanych z projektowaniem i wykorzystaniem rękawic sensorycznych w kontekście rozpoznawania gestów. Przedstawia również krótką charakterystykę wybranych metod związanych z uczeniem maszynowym, wykorzystywanym w rozpoznawaniu gestów. Krótko wspomina sieci neuronowe, w szczególności modele rekurencyjne. Nie wiem jednak, dlaczego nie opisuje szczegółowo sieci splotowych, skoro z nich korzysta w ramach dalszych prac. W tym samym rozdziale Doktorant opisuje również problem redukcji wymiarowości danych (głównie klasyczne metody jak PCA i LDA) oraz przedstawia wybrane techniki selekcji cech. Właściwa analiza problemu i aktualnych prac innych zespołów wskazuje ogólną wiedzę teoretyczną Doktoranta w obszarze tematycznym związanym z inżynierią biomedyczną.

W kolejnym, trzecim rozdziale pracy, Pan mgr inż. Jakub Piskozub przedstawia projekty rękawic oraz systemu akwizycji danych. Praca doktorska ma charakter eksperymentalny i zaprojektowanie rękawicy stanowi ważną część pracy Doktoranta. Oczywiście projekty podobnych rękawic znane są od wielu lat (np. projekt studencki z 2014 roku z Cornell University wykorzystujący Spectra Symbol Flex-Sensors: https://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/f2014/rdv28_mjl256/webpape/) i szereg z nich było podstawą prac badawczych, o których wspomina również Pan mgr inż. Jakub Piskozub. Niemniej, opracowanie rękawicy sensorycznej, umożliwiło Doktorantowi przeprowadzenie szeregu eksperymentów związanych z podejmowaną tematyką.

Rozdział czwarty przedstawia opis wybranych aspektów metody badawczej. Autor opisuje cztery fazy badań (począwszy od fazy pilotowej) oraz poszczególne elementy metody badawczej począwszy od doboru uczestników badań, poprzez opis procedury pomiarowej (dla poszczególnych faz badań), prezentację metod przetwarzania wstępnego (w tym technik poszerzania zbioru danych), a skończywszy na aspektach uczenia maszynowego wykorzystywanego w pracy.

Wyniki badań zaprezentowano w rozdziale piątym. W kolejnych podrozdziałach Doktorant przedstawia rezultaty badań wstępnych, klasyfikacji gestów statycznych i klasyfikacji znaków Polskiego Języka Migowego z pełnym i zredukowanym zbiorem cech. Na końcu rozdziału właściwie przedstawia wyniki własnych prac w odniesieniu do rezultatów podobnych badań innych zespołów.

Ostatni, szósty rozdział pracy, przedstawia krótką dyskusję uzyskanych wyników i konkluzje pracy.

Autor cytuje 59 pozycji literatury. Część cytowanych prac, to recenzowane publikacje w czasopismach lub materiałach konferencyjnych. Około 20 prac recenzowanych, to pozycje opublikowane w ciągu ostatnich pięciu lat. Uważam, że zakres literatury można było uzupełnić o szereg nowych prac opublikowanych w źródłach recenzowanych. Jedynie w bazie IEEE Xplore można znaleźć co najmniej kilkadziesiąt powiązanych, nowych pozycji, które podkreślają stan wiedzy w zakresie nowoczesnych metod zarówno konstrukcji rękawic sensorycznych (między innymi użycie nowoczesnych materiałów i tekstyliów) jak i propozycji szeregu nowoczesnych metod rozpoznawania gestów. Przykładowym artykułem tego rodzaju jest np. M. Pan, Y. Tang and H. Li, "State-of-the-Art in Data Gloves: A Review of Hardware, Algorithms, and Applications," in *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-15, 2023, Art no. 4002515, doi: 10.1109/TIM.2023.3243614.

Na końcu rozprawy Doktorant załączył suplement wskazujący link do repozytorium GitHub, w którym dostępnych jest kilka projektów związanych z eksperymentalną częścią pracy.

3. Oryginalne osiągnięcia i ocena merytoryczna rozprawy

Oryginalne rozwiązanie problemu naukowego w rozprawie dotyczy oceny możliwości rozpoznawania gestów wykonywanych rękawicą sensoryczną przy zastosowaniu mniejszej liczby czujników (cech). Do oryginalnych osiągnięć zaliczam:

- O1: Opracowanie systemu złożonego z rękawicy sensorycznej oraz algorytmów uczenia maszynowego i weryfikację za pomocą tego systemu hipotezy dotyczącej możliwości uzyskania dokładności rozpoznawania znaków Polskiego Języka Migowego z dokładnością 95 procent dla gestów statycznych i około 90 procent dla gestów statycznych i dynamicznych, uzyskanych w przypadku zredukowania liczby zestawów danych do pomiarów z trzech czujników.
- O2: Wykazanie, z wykorzystaniem eksperymentów zrealizowanych za pomocą wyżej opisanego systemu i dla wyżej przedstawionego kontekstu, że opracowanie spersonalizowanych modeli, czyli takich, które przygotowano ze szczególnym wykorzystaniem danych od danej, pojedynczej osoby, umożliwia uzyskanie dużej dokładności rozpoznawania gestów dla tej osoby.

Doktorant rozwiązał problem naukowy postawiony w tezie i udokumentował rozwiązanie poprzez prezentację i analizę wyników szeregu eksperymentów pokazując wartości miar różnych konfiguracji i liczby czujników.

Za cenne osiągnięcia w obszarze nauk inżynierijno-technicznych uważam również rozwiązania techniczne dotyczące: a) zaprojektowania, realizacji i przetestowania rękawicy sensorycznej, b) zaimplementowania i przetestowania oprogramowania uczenia maszynowego dla rozpatrywanego problemu i c) zebranie i udostępnienie innym naukowcom bazy danych zarejestrowanych sygnałów.

Powyższe osiągnięcia, zarówno naukowe jak i techniczne, wskazują, że rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej przez Doktoranta. Ponadto, opisane osiągnięcia na tle stanu wiedzy wskazują, że rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego.

Niemniej, oceniając pracę należy zwrócić uwagę na kilka ważnych aspektów:

- 1.) Zebrany zbiór danych, który jest podstawą głównych eksperymentów liczy 5760 sekwencji (36 znaków x 16 osób x 10 powtórzeń) i nie jest dla mnie jasne jak ten zbiór został podzielony na podzbiory treningowy, testowy (walidacyjny). Problem dotyczy tego, że w przypadku klasyfikacji referencyjnej (tj. z użyciem wszystkich danych) nie jest określone, czy Doktorant dzielił dane na podzbiory ze względu na zróżnicowanie osób (czyli dane od danej osoby są wyłącznie w zbiorze treningowym lub testowym), czy mieszał dane od tych samych osób zarówno w zbiorze treningowym jak i testowym (ze względu na powtórzenia). Jeśli bowiem zastosował drugie z przedstawionych wyżej podejść, to z natury metod uczenia maszynowego wynika, że uzyskane modele będą prawdopodobnie nadmiernie dopasowane do danych. Dotyczy to również deklarowanej wielokrotnie w rozprawie walidacji krzyżowej (z pięciokrotnym podziałem na zbiór treningowy i testowy).
- 2.) Nie jest jasne jak zróżnicowano warunki wykonywania gestów w czasie eksperymentów. Przykładowo, czy dla danego uczestnika robiono przerwy, jak długie, co uczestnicy robili w czasie przerw (tj. w celu zmniejszenia efektu „pamięci mięśni” przy powtarzaniu gestu), itp. Czy wykonywano powtórzenie jednej litery dziesięć razy i następnie kolejna litera, czy dziesięć razy powtarzano cały alfabet. Są to bardzo ważne różnice istotne zarówno dla znaczenia uczenia modeli z użyciem wszystkich pomiarów, jak i dla podejścia spersonalizowanego. Jeśli bowiem powtórzenia liter były wykonywane w serii, wówczas wartość uzyskanego zbioru danych dla problemu uczenia uniwersalnego modelu jest znacznie mniejsza.
- 3.) Nie jest jasne, jak przygotowano i wykorzystano zbiór poszerzający dane pomiarowe. Doktorant w rozdziale 4.4 wskazuje zastosowanie typowych metod, np. rozciągania sygnałów, podczas gdy w rozdziale 5.3 jest wskazane, że „(...) tested by multiplying 200 times the amount of data from each participant and re-training the model only on data from one person”. Nie jest zatem jasne, jak zrealizowano augmentację danych dla poszczególnych eksperymentów (tj. jakie konkretnie operacje, ile razy zastosowano dane operacje, czy stosowano je tylko do danych treningowych, itd.). Bez szczegółowych informacji można przypuszczać, że model wyuczony na tak przygotowanych danych może być bezużyteczny (zbyt mocno dopasowany do danych) dla przyszłych zastosowań.
- 4.) W rozdziale 4.1 Doktorant charakteryzuje grupy osób, biorące udział w eksperymentach związanych z wykonywaniem gestów. Brak szczegółowych informacji czy osoby znały język migowy, w jakim zakresie, od ilu lat się nim posługują, czy przeszły wstępne szkolenie, jeśli tak, to jakie, itd. Dla czwartej fazy badań (16 osób) podano, że jedna osoba posiadała stosowny certyfikat w zakresie komunikacji z wykorzystaniem języka migowego. Dużą wartością pracy byłaby korelacja i porównanie możliwości rozpoznawania gestów z zastosowaniem proponowanego systemu dla osób o różnym doświadczeniu w zakresie posługiwania się językiem migowym.
- 5.) W rozdziale 5.3, na stronie 50, Doktorant wskazuje, że „The result was confirmed by 5-fold cross validation” oraz na stronie 51: „(...) with the data divided into training, test, and validation sets, including 5-fold cross validation”. Nie znalazłem ani w pracy,

ani we wskazanym repozytorium, szczegółów dotyczących walidacji krzyżowej i jej wyników (dla poszczególnych, pięciu podziałów). Tym bardziej, że w kodzie źródłowym (projekt HandPi-Classifer, plik handclas.py) wykorzystana jest tylko funkcja „train_test_split()” do podziału danych na zbiór treningowy i testowy w proporcji 0,8 i 0,2 (oraz zbiór testowy na zbiór walidacyjny i testowy w proporcji 0,2 i 0,8).

Wszystkie powyższe uwagi uważam, za kluczowe ze względu na moc argumentacji w rozprawie, bazującej na użytych zbiorach danych do przygotowania modeli i wnioskowania na podstawie wyników tych modeli. Dlatego, w dalszej części rozprawy sformułowałem pytania do Doktoranta, które powinny być przedstawione w czasie dyskusji podczas obrony pracy doktorskiej.

4. Uwagi szczegółowe i pytania

Praca w większości napisana jest poprawnie. Niemniej chciałbym zwrócić uwagę na szereg drobnych, dodatkowych elementów dotyczących głównie redakcji pracy.

W szczególności:

- W opisie dotyczącym doboru grupy badawczej nie jest dla mnie jasne jak określano rozmiary dłoni dla kolejnych grup, ponieważ raz wymiary są podane w mm (np. „hand sizes between 16 and 21.5mm”), w innych miejscach w cm (np. „They had similar hand size ranging from 18 do 21 cm [...]”).
- Spis skrótowców i streszczenia nie są ujęte żadną numeracją stron.
- Spis skrótowców nie jest uporządkowany ani alfabetycznie, ani według kolejności pojawiania się w tekście. Nie zawiera również szereg skrótowców stosowanych w tekście rozprawy.
- Dokument byłby o wiele bardziej czytelny, gdyby każdy główny rozdział rozpoczynał się od nowej strony.
- Numeracja rozdziałów w języku angielskim nie kończy się zwyczajowo kropką (w odróżnieniu od publikacji w języku polskim). Przykładowo zamiast „2. Theoretical [...]”, powinno być „2 Theoretical [...]”.
- Umieszczanie skrótowców i różnych wtrąceń w tekście zapisano z zastosowaniem znaków „- -”, co prowadzi do pogorszenia czytelności pracy, np. „[...] - BSL -, etc. -.”, „- Fig. 2.3.1 -: „”.
- Zapis numeracji wzorów i zapis ich cytowania jest również mało czytelny i niestandardowy, np. „ - 1 - ” i dalej (strona 46) „The classification accuracy was expressed using - 1 -.”
- Nieuporządkowane stosowanie zapisu liczb zmiennoprzecinkowych. W niektórych miejscach zapisane z przecinkiem (np. tabele 5.1.1, 5.1.2) w innych miejscach właściwie z kropką (np. 5.2.1).
- Niektóre zdania są trudne do zinterpretowania, szczególnie w kontekście merytorycznym. Przykładowo (strona 43): „Optimization of the model is implemented using the Adam algorithm with an initial learning rate of LR [...]”. Jaka jest zatem wartość początkowa lub jak należy interpretować LR (brak opisu, brak w wykazie)?
- Podpis rysunku 4.4.1 nie jest na tej samej stronie co rysunek (zaczyna nową stronę).

- Niektóre rysunki są identyczne jak w publikacjach autora (np. 2.3.2, 4.2.1). Należałoby wspomnieć o tym w podpisie lub odniesieniu do rysunku.

Pytania/prośby do autora rozprawy

- P1: Proszę o wyjaśnienie sposobów podziału danych na zbiory treningowe, testowe i walidacyjne (oraz dla pięciu podziałów walidacji krzyżowej) zastosowanych w pracy zarówno dla danych pomiarowych, jak i dla danych wykorzystujących augmentację.
- P2: Proszę o uszczegółowienie warunków wykonywania gestów z powtórzeniami w poszczególnych eksperymentach.
- P3: Proszę o wyjaśnienie, jak zrealizowano augmentację danych dla poszczególnych eksperymentów (tj. jakie konkretnie operacje, ile razy te operacje, czy tylko do danych treningowych, itd.).
- P4: Proszę o przedstawienie szczegółów dotyczących architektury sieci neuronowej (rysunek 4.5.1) z podaniem liczby parametrów modelu podlegających uczeniu i omówienie tych parametrów w funkcji liczby przykładów zbioru treningowego bez i z augmentacją.

5. Konkluzja recenzji

Podsumowując, stwierdzam, że przedstawione w rozprawie cele badawcze zostały osiągnięte, metody badawcze w większości właściwie dobrane i zastosowane, a eksperymenty w większości poprawnie przeprowadzone. Teza została potwierdzona.

Biorąc pod uwagę przedstawione wyżej analizy i uzasadnienia stwierdzam, że rozprawa doktorska prezentuje ogólną wiedzę teoretyczną Doktoranta w dyscyplinie inżynieria biomedyczna. Doktorant wykazuje umiejętność samodzielnego prowadzenia pracy naukowej a rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego.

Dlatego stwierdzam, że rozprawa mgr. inż. Jakuba Piskozuba spełnia wymagania sformułowane w Ustawie z dnia 20 lipca 2018 roku – Prawo o szkolnictwie wyższym i nauce, Dz. U. 2018, poz. 1668, z późniejszymi zmianami. Wnioskuje do Rady Dyscypliny Inżynieria Biomedyczna Politechniki Śląskiej o dopuszczenie mgr. inż. Jakuba Piskozuba do dalszych etapów postępowania w sprawie nadania stopnia naukowego doktora.

prof. dr hab. inż. Jacek Rumiński, prof. PG
Katedra Inżynierii Biomedycznej, ETI, PG