

RECENZJA

rozprawy doktorskiej mgr. inż. Norberta Kapińskiego
pt. *Proces gojenia ścięgna Achillesa oceniany przez fuzję danych z wykorzystaniem
głębokich sieci neuronowych.*

1. Podstawa formalna recenzji

Przedmiotem recenzji jest nowa wersja pracy doktorskiej mgr. inż. Norberta Kapińskiego, której pierwotna treść – na mocy §6 pkt. 6 Rozporządzenia MNiSzW z dnia 30.01.2018 (Dz.U. nr 261) – została zwrócona do niezbędnych uzupełnień i poprawy. Doktorant przedstawił nową wersję pracy w październiku 2020 r. (praca jest opatrzona nieprawidłową datą *grudzień 2019 r.*), którą skierowano do powtórnej recenzji.

2. Zakres i cel rozprawy

Zakres i cel rozprawy nie uległ zmianie. Poprawiona wersja pracy doktorskiej pozostaje rozprawą z zakresu komputerowego wspomaganie decyzji medycznych poprzez przetwarzanie, analizę i rozpoznawanie informacji obrazowej, ukierunkowanej na ocenę procesu gojenia ścięgna Achillesa po jego całkowitym zerwaniu i po zabiegu rekonstrukcyjnym w trakcie długoczasowego procesu rehabilitacji. Obejmuje złożoną problematykę analizy zadania oceny stanu ścięgna w celu zdefiniowania cech różnicujących stopień zaawansowania procesu gojenia, pozyskiwania obrazów MRI, ich przetwarzania w celu poprawy jakości, wydobycia obiektów i wyznaczenia cech morfometrycznych, a także metody i algorytmy regresji zmierzające do wyznaczenia stosownej liczbowej oceny rozpatrywanego procesu gojenia.

Z uwagi na poważne trudności, jakie przysparza lekarzom medyczna diagnostyka obrazowa w codziennej praktyce klinicznej (rozdzielenie obiektu zainteresowania od tła, wyodrębnienie klas obiektów istniejących w obrazie, wyznaczenie liczbowej oceny cech charakteryzujących analizowane obiekty), na co również zwraca uwagę Doktorant, zauważa się gwałtowny wzrost metod i narzędzi komputerowego przetwarzania i analizy obrazów medycznych, stanowiących dla lekarza istotne wsparcie jego wysiłków w stawianiu diagnozy i doborze terapii.

Niniejsza praca mieści się w tym właśnie bardzo aktualnym nurcie informatyki medycznej, dlatego tematykę podjętą w rozprawie uważam za oryginalną, uzasadnioną i ważną dla współczesnych prac w obszarze przetwarzania i analizy medycznej informacji obrazowej oraz wspomaganie procesów decyzyjnych w medycynie. Pozytywna ocena tematu rozprawy wynika również z faktu, iż każda praca z tego zakresu, której wynikiem są konstruktywne algorytmy przetwarzania, analizy i klasyfikacji informacji obrazowej ukierunkowane na konkretny medyczny problem decyzyjny, a do takich zaliczam recenzowaną rozprawę, jest celowa i ważna z praktycznego punktu widzenia.

3. Zawartość rozprawy

Układ rozprawy i zawartość poszczególnych rozdziałów uległy zmianie. Nowa rozprawa składa się z 8 rozdziałów poprzedzonych streszczeniami w języku polskim i angielskim, spisem rysunków i tabel, 2 dodatków, wykazu oznaczeń oraz bibliografii zawierającej 150 pozycji i obejmuje 155 stron maszynopisu.

Rozdział pierwszy (*Wstęp*) pozostał niezmieniony. Wprowadza on w „klimat” pracy, przedstawia informacje dotyczące szybkiego rozwoju technik obrazowania medycznego oraz zwraca uwagę na ich ścisły związek z nowatorskimi metodami komputerowego przetwarzania informacji, a wśród nich z głębokimi sieciami neuronowymi, mieszczącymi się w obszarze sztucznej inteligencji. Podkreślono tu również znaczenie architektur obliczeniowych pozwalających na efektywną komputerową implementację zaawansowanych algorytmów. Rozdział drugi o zmienionym na korzyść tytule (*Cel i struktura pracy*) zawiera sformułowanie tezy pracy oraz jej celów: głównego i pobocznych, a także medyczny kontekst problemu diagnostycznego rozpatrywanego w pracy. Przedstawiono tu również strukturę pracy oraz w ujęciu syntetycznym zmienioną w niektórych przypadkach zawartość poszczególnych rozdziałów.

Pierwsza część rozdziału 3. zatytułowanego *Monitorowanie procesu gojenia ścięgna Achillesa*, poświęcona medycznemu aspektowi rozpatrywanego problemu diagnostycznego, uległa drobnym redakcyjnym zmianom nie ingerującym w jej istotę treściową. W pierwszej kolejności Autor informuje Czytelnika bardzo szczegółowo o anatomii ścięgna Achillesa, jego biomechanice i roli jaką pełni w organizmie, urazach ścięgna i ich kategoriach, przyczynach urazów i częstości ich występowania, czynnikach (wewnętrznych i zewnętrznych) sprzyjających urazom ścięgna oraz o całościowym przebiegu procesu leczenia (metody leczenia, rehabilitacja, fazy procesu gojenia). W drugiej części rozdziału znajdujemy dość szczegółową wiedzę na temat wykorzystywanej dalej techniki obrazowania medycznego zwanej rezonansem magnetycznym (MRI), poczynwszy od podstaw techniki mającej swe źródła w mechanice kwantowej, poprzez różne fizyczne zjawiska i prawa matematyczne nimi rządzące (np. równanie Larmora), aż do sposobów ustawień urządzeń do MRI i parametrów obrazowania prowadzących do różnych wynikowych sekwencji obrazów i ich modalności, dających różne możliwości diagnostyczne. Łącznie przedstawiono tu 10 sekwencji (wraz z podgrupami). Autor wzbogacił prezentację dodając graficzne zestawienie z odpowiednim tekstowym komentarzem wykorzystywanych sekwencji rezonansu magnetycznego z widoczną strukturą ścięgna Achillesa. W opisie badania ultrasonograficznego dodano ilustrację przedstawiającą obszar obrazowania ścięgna Achillesa w badaniu USG. Rozdział kończy część (pozostawiona bez zmian) poświęcona badaniom biomechanicznym, których celem jest określenie własności mechanicznych różnych struktur organizmów żywych. Opisane tu zostały między innymi pomiary biomechaniczne wykorzystywane w pracy.

W rozdziale 4. zatytułowanym *Konwolucyjne sieci neuronowe* nastąpiły dość znaczące zmiany organizacyjne: zrezygnowano z punktu 4.4 (*Zastosowania w medycynie*), a pewne istotne syntetyczne i poglądowe informacje przeniesiono do punktu 4.1, który zamiast tytułu *Rys historyczny* przyjął trochę mylącą nazwę *Ewolucja i jej implikacje w radiologii* (nie wiadomo, o jaką ewolucję chodzi). Punkt 4.3 pozostał praktycznie niezmieniony i – jak poprzednio – obejmuje prezentację algorytmów uczenia głębokich sieci neuronowych oraz

architektury wybranych sieci konwolucyjnych (AlexNet, GoogLeNet, ResNet), które były stosowane w zaproponowanej koncepcji wnioskowania o stanie gojącego się ścięgna Achillesa i w prowadzonych badaniach eksperymentalnych. Usunięto jedynie klasyfikator Bayesa jako metodę kombinowania wyników modeli bazowych (klasyfikacji i regresji). W punkcie 4.2 o zmienionym tytule (*Szkolenie*) w dalszym ciągu przedstawiono szereg różnych zagadnień koncepcyjnych związanych z problemem nadmiernego dopasowania i zmniejszeniem wymiarowości wektora cech opisującego obiekt (selekcja i redukcja). W punkcie dodano wskaźniki stosowane w ocenie jakości modeli regresyjnych.

Dwa kolejne rozdziały zawierają oryginalne wyniki Doktoranta. W rozdziale 5 (*Nowa metoda oceny procesu gojenia ścięgna Achillesa*) przedstawiono opracowaną metodę i jej składowe, zbiory danych rzeczywistych wykorzystywane w badaniach eksperymentalnych, sposób wyznaczenia wzorcowej (referencyjnej) oceny procesu gojenia ścięgna Achillesa przez zespół ekspertów (radiolodzy, ortopedzi). Rozdział zawiera wyniki poszczególnych eksperymentów przeprowadzonych na różnych zbiorach danych i ukierunkowanych na konkretne zagadnienia poznawcze (rozdzielenie ścięgna zdrowego i po zerwaniu, redukcja wymiaru wektora cech, wybór sekwencji MRI, ocena procesu gojenia z wykorzystaniem macierzy błędu i krzywej gojenia). Pod względem strukturalnym rozdział wzbogacono o punkt 5.3.3 (*Ocena zbiorcza*), w którym przedstawiono jeden syntetyczny wskaźnik oceny stanu ścięgna Achillesa. Dokonano znaczących zmian treści i ilustracji zdecydowanie poprawiających przekaz informacyjny.

Rozdział 6 (*Testy nowej metody*) obejmuje kolejne wyniki badań eksperymentalnych, ale związane z analizami porównawczymi z innymi metodami automatycznej oceny procesu gojenia ścięgna Achillesa. Opracowaną metodę porównano z oceną generowaną przez wybrane sieci konwolucyjne przetwarzające obrazy sekwencji MRI oraz sieci konwolucyjne, ale działające na obrazach USG. Dokonano również porównania metody z wynikami badań biomechanicznych oraz testu ATRS. Treść rozdziału w porównaniu z wersją pierwotną pracy została znacząco zmieniona i uzupełniona.

W nowej wersji rozprawy dodano rozdział 7 (*Wnioski*) zawierający wnioski praktyczne wynikające z przeprowadzonych badań eksperymentalnych i analiz porównawczych. Pracę kończy rozdział 8 (*Podsumowanie*) zawierający syntetyczne podsumowanie uzyskanych wyników oraz propozycje dalszych prac w rozpatrywanym obszarze. W dodatku A przedstawiono krótką informację o aplikacji webowej AchillesDL powstałej w ramach projektu START i służącej do przeglądania i etykietowania danych radiologicznych. W nowym dodatku B szczegółowo zaprezentowane zostały cechy Haralicka opisujące teksturę ROI.

W podsumowaniu tego fragmentu opinii należy z uznaniem podkreślić dbałość Autora o wysoki poziom edytorski rozprawy, ze szczególnym uwzględnieniem bogatej i trafnie skomponowanej sfery ilustracyjnej. Na pozytywną ocenę zasługuje także logiczny układ pracy i przyjęty podział na rozdziały, który prawidłowo oddaje koncepcyjno-eksperymentalno-praktyczny charakter rozprawy.

W ocenie pierwszej wersji pracy sformułowano zarzut niewłaściwego rozróżnienia treści ważnych dla pełnego zrozumienia osiągnięć Doktoranta i treści nieistotnych, do których w dalszej części pracy się nie nawiązuje. Choć pod tym względem wersja poprawiona i uzupełniona prezentuje się znacznie lepiej, to wciąż z jednej strony, Autor skąpi czytelnikom

ważnych szczegółów nt. stosowanych metod i procedur, a z drugiej zaś chętnie kreśli ich obszerne tło historyczne, co – jak wynika z przekazanego opisu wprowadzonych zmian – jest pochodną zainteresowań Doktoranta. W takim wzbogaceniu treści rozprawy o dodatkowe elementy, niezwiązane z oryginalnymi wynikami pracy doktorskiej, nie ma niczego złego (może poza utrudnieniem pracy recenzentom), pod warunkiem wszak, że w pierwszej kolejności należycie zadba się o kompletność i wystarczającą szczegółowość w prezentacji tej rozprawy. Ten warunek w ocenianym tekście wciąż jeszcze nie do końca został spełniony, o czym będzie mowa w dalszej części recenzji.

4. Uzyskane wyniki

Celem pracy było wykazanie następującej tezy (str.5):

Cechy ekstrahowane przez głębokie sieci neuronowe będą skuteczne do oceny procesów patofizjologicznych widocznych w obrazowaniu medycznym, takich jak gojenie tkanki miękkiej. Aby zweryfikować sformułowaną hipotezę, Autor przedstawia następujące cele pracy:

Cel główny (cyt.): Opracowanie automatycznej metody oceny gojenia się ścięgna Achillesa, bazującej na fuzji danych z wykorzystaniem głębokich sieci neuronowych.

Cele poboczne (cyt.):

- 1. Wybór efektywnego kosztowo i czasowo protokołu badania bazującego na technikach obrazowania medycznego, a dokładniej rezonansu magnetycznego.*
- 2. Przetestowanie różnego rodzaju podejść związanych ze szkoleniem głębokich sieci neuronowych.*
- 3. Porównanie wyników oceny nowej metody z wynikami klasyfikacji bazującej na danych z ultrasonografii.*
- 4. Porównanie wyników oceny nowej metody z oceną funkcjonalną, rutynowo stosowaną do wspomagania rehabilitacji po urazie ścięgna.*

Realizacja sformułowanych celów doprowadziła do uzyskania następujących wyników szczegółowych:

1. Zaproponowanie systemu regresyjnego, który wyznacza ocenę procesu gojenia uszkodzonego ścięgna Achillesa (w skali 0 – 7) dla 6 różnych wskaźników na podstawie obrazów przekrojowych uzyskanych z badań MRI oraz 46 cech obrazowych uzyskanych z wysegmentowanego obszaru ROI. W pierwszej kolejności następuje ekstrakcja 4096 cech z obrazu z wykorzystaniem wytrenowanej (dla zadania dychotomii: ścięgno zdrowe/chore) głębokiej sieci konwolucyjnej, a następnie odbywa się redukcja cech z zastosowaniem metody PCA. Uzyskane w ten sposób 200 cech wraz z cechami obrazowymi są kolejno przetwarzane przez algorytm regresji liniowej LASSO, gdzie następuje dodatkowa selekcja cech, a następnie przez tzw. meta-regresor wyniki dla różnych obrazów przekrojowych są uśrednione z odrzuceniem wartości skrajnych (tzw. średnia trymowana). Do budowy zaproponowanego systemu zastosowano procedury uczenia nadzorowanego, których podstawą był zgromadzony zbiór danych klinicznych. Na zbiór ten składały się wyniki badań MRI dla 29 zdrowych osób oraz badania 60 pacjentów z zerwanym ścięgnem Achillesa przeprowadzone dla 10 różnych odstępów czasowych od urazu (przed operacją oraz 9 badań w okresach od 1 tygodnia do 1 roku po operacji). Każde badanie zostało opisane przez radiologa-eksperta pod względem 6

parametrów oceniających w skali 0-7 stan ścięzna (stopień zaawansowania procesu gojenia) pod różnym kątem. Zgromadzone dane posłużyły także w badaniach eksperymentalnych jako dane testowe – zastosowano tutaj metodę wielokrotnej walidacji krzyżowej.

2. Przeprowadzenie bogatego zestawu badań eksperymentalnych oraz analiz porównawczych:

2.1. Eksperymentalny wybór rodzaju głębokiej sieci neuronowej (spośród 3 porównywanych sieci: AlexNet, ResNet-18 i GoogleNet), która w zaproponowanym systemie wnioskowania będzie pełniła rolę ekstraktora cech opisujących obrazy przekrojowe MRI. Jako kryterium przyjęto tutaj jakość klasyfikatora binarnego (zdrowy/chory) wykorzystującego wyekstrahowane cechy oraz czas uczenia sieci. Wybór padł na sieć AlexNet – była ona zdecydowanie najszybsza i dawała jakość klasyfikacji powyżej 99% poprawnych decyzji.

2.2. Badanie procedury redukcji cech PCA w celu między innymi określenia liczby składowych głównych. Wybrano 200 składowych, co daje poziom bliski 99% skumulowanej wariancji.

2.3. Wybór sekwencji rezonansu magnetycznego, które będą stosowane w zaproponowanej metodzie automatycznej oceny. W pierwszej kolejności przeanalizowano wizualnie histogramy dla 10 różnych sekwencji i modalności zwracając uwagę na ich lokalne maksima. Przeprowadzony ogląd wraz z wyznaczonymi współczynnikami korelacji dla niektórych sekwencji doprowadził do wyboru sekwencji T2*GRE TE_MIN.

2.4. Kolejny eksperyment miał na celu wybór parametru α w kryterium metody regresji liniowej LASSO. Najlepszy rezultat uzyskano dla $\alpha=0.1$, dla którego przedstawiono również predykaty (dla 6 różnych wielkości opisujących stan ścięzna) pozostawionych po selekcji metodą LASSO. Liczba wyselekcjonowanych predykatów waha się od 12 do 16.

2.5. Eksperymentalne porównanie wyników oceny procesu gojenia przez zaproponowaną metodę z oceną radiologa-eksperta. Przeprowadzone badania składały się z 3 eksperymentów. W pierwszym eksperymencie wyznaczono tzw. macierz błędów, której komórki zawierają błędy MAE dla badań, dla których moment wykonania (tydzień, jaki upłynął od zabiegu) był wcześniej rozpoznany przez radiologa i metodę automatyczną. Najmniejsze błędy, świadczące – zdaniem Doktoranta – o dobrym modelowaniu oceny eksperta przez metodę automatyczną powinny układać się na przekątnej macierzy błędów. Odległość rzeczywistego przebiegu najmniejszych błędów od przekątnej oceniono przy pomocy zaproponowanego wskaźnika będącego polem powierzchni pomiędzy krzywą określającą wartości minimalne błędów MAE a wartościami na przekątnej. W drugim eksperymencie problem oceny ścięzna sformułowano jako zadanie klasyfikacji z 8 klasami oznaczającymi ocenę sprowadzoną do liczb całkowitych od 0 do 7. Przedstawiona tutaj macierz pomyłek precyzuje charakter poprawnych i błędnych ocen. Trzeci eksperyment pokazuje na szeregu wykresach przebieg tzw. krzywej gojenia według oceny radiologa-eksperta oraz według oceny uzyskanej przy pomocy opracowanej metody (uzyskanej przez zaokrąglenie wyniku regresji do najbliższej wartości całkowitoliczbowej). Wyniki zostały przedstawione dla wszystkich 6 sformułowanych wskaźników oceny stanu ścięzna oraz dla 4 różnych pacjentów.

2.6. W kolejnym eksperymencie przedstawionym w podrozdziale 6.1 dokonano porównania opracowanej metody z systemami regresyjnymi bazującymi na sieciach konwolucyjnych AlexNet, Inception-v3 (wersja GoogleNet) i ResNet. W pierwszej kolejności sieci wytrenowano w procedurze minimalizującej średni błąd kwadratowy (MSE), a następnie porównano pod względem błędów MAE, MAX_AE i Corr na zbiorze trenującym. Pozwoliło to

zidentyfikować najlepszy regresor, którym okazała się sieć Inception-v3. Zidentyfikowana najlepsza sieć (choć, jak wynika z poprzednich badań, ma ona najdłuższy czas treningu – patrz rys. 5.3) została porównana z opracowaną metodą oceny stopnia gojenia ścięgna Achillesa (z meta-regresem w postaci algorytmu SVR) pod względem tych samych kryteriów dla 6 wskaźników oceny, ale tym razem na zbiorze testowym. Uzyskane wyniki porównawcze pokazują, iż opracowana metoda charakteryzuje się lepszą jakością predykcji.

2.7. Eksperymentalne porównanie oceny automatycznej na podstawie badań MRI z oceną wykorzystującą obrazy USG. Wpierw w przeprowadzonych badaniach zbudowano klasyfikatory binarne (zdrowy/chory) wykorzystujące konwolucyjne sieci neuronowe (AlexNet, Inception v3, ResNet) i przetwarzające obrazy USG ścięgna Achillesa w projekcji osiowej i strzałkowej. Uzyskane klasyfikatory zostały porównane z klasyfikatorem zbudowanym w oparciu o zaproponowaną metodę wykorzystującym dane z obrazów MRI. Uzyskane wyniki pokazują wyższość opracowanej metody dokonującej fuzji informacji z cechami wyznaczonymi na bazie danych radiologicznych MRI. Podobne wyniki uzyskano dla porównywanych metod, które pracowały jako regresory i wyznaczały oceny procesu gojenia ścięgna Achillesa po uszkodzeniu i zabiegu rekonstrukcyjnym.

2.8. W eksperymentach porównujących wyniki opracowanej metody z badaniami biomechanicznymi, wpierw posługując się metodami analizy czynnikowej, dokonano redukcji wielkości opisujących biomechanikę z 16 do 4 zmiennych. Następnie określono współczynniki korelacji pomiędzy automatyczną oceną uzyskaną z opracowanej metody oraz oceny podanej przez radiologa z wyznaczonymi 4 zmiennymi oraz dodatkowo z badaniem biomechanicznym ATRS. Uzyskane wyniki pokazują słabe skorelowanie ocen uzyskanych na podstawie badań MRI z wynikami pochodzącymi z testów biomechanicznych.

5. Uwagi krytyczne

Mimo, iż wiele uwag zawartych w recenzji pierwszej wersji pracy zostało przez Autora uwzględnionych we wprowadzonych poprawkach i uzupełnieniach, to wciąż w rozprawie zauważa się niepoprawione błędy merytoryczne i redakcyjne, a także niejasności lub błędne sformułowania, które mogą powodować niezrozumienie istoty stosowanych przez Autora rozwiązań i koncepcji lub ich niewłaściwą interpretację.

1. W podrozdziale 4.2.2 Autor przedstawia algorytm regresji LASSO oraz metodę redukcji cech PCA, które są wykorzystywane w zaproponowanej koncepcji komputerowej oceny procesu gojenia ścięgna Achillesa. Oprócz tych ważnych informacji punkt obejmuje także „małe co nieco” o procedurach selekcji cech, choć nie są one stosowane w autorskiej metodzie i w badaniach eksperymentalnych (oprócz techniki LASSO, która posiada naturalne własności selekcji). Taki rozszerzony kontekst informacji o metodach zmniejszania wymiarowości wektora cech ma sens, ale prezentacja taka powinna być starannie przemyślana. Przekaz informacyjny przyjęty przez Autora jest daleki od doskonałości: przedstawiono przypadkowo wybrane 2 kryteria oceny cech stosowane w metodach typu *filtr*, nie wspominając nawet o metodach obudowanych (*wrappery*) i wbudowanych, a w samych formułach matematycznych wciąż niewyjaśnione (mimo prośby recenzenta w poprzedniej opinii) pozostają znaczenia zmiennych losowych i rozkładów prawdopodobieństwa, bez czego prezentacja obu formuł mija

się z celem. Dodatkowo wspomniano o możliwości wykorzystania do selekcji cech teorii zbiorów przybliżonych powołując się na literaturę, która nie dotyczy selekcji cech.

2. W podrozdziale 4.1 Autor przedstawił krótki rys historyczny systemów neuronowych poczynając od koncepcji sztucznego neuronu McCullocha-Pittsa z 1943 r., niestety, nie unikając przy tym błędów. Przykładowo: rys. 4.2. nie przedstawia sieci (jak napisano w tekście), ale pojedynczy neuron; wzór 4.3 jest nieprawidłowy – powinno być nie n , ale potencjał aktywacyjny Z ; funkcja aktywacji neuronu ujmuje zależność jego wyjścia (y) od potencjału aktywacyjnego (Z); stwierdzenie, iż neuron z funkcją progową wyznacza hiperpłaszczyznę jest nieprecyzyjne – taką płaszczyznę wyznacza potencjał aktywacyjny, a neuron zwraca informację (np. 0, 1, lub dla bipolarnej funkcji progowej -1, +1), po której stronie hiperpłaszczyzny jest wejściowy wektor x (hiperpłaszczyznę modeluje neuron z tożsamościową funkcją aktywacji – *adaline*); we wzorze 4.5 zamiast x (czyli – według oznaczeń pierwszej współrzędnej prostokątnej(?)) powinien być potencjał aktywacyjny Z .

3. Ważnym krokiem w zaproponowanej przez Doktoranta metodzie jest fuzja wyników modeli bazowych, co znalazło odbicie w tytule rozprawy. Przegląd metod zespołowych (dla klasyfikacji i regresji) został przedstawiony w podrozdziale 4.3.4 *Złożenia* i ograniczony do 3 procedur: uśrednianie z funkcją *softmax*, głosowanie i *stacking*. Przegląd ten zajmuje 1,5 strony (więcej miejsca poświęcono podstawom fizycznym badań MRI i USG, które nie są istotne dla prezentacji wyników pracy) i w najmniejszym stopniu nie stanowi wyboru reprezentatywnego, ani też nie uświadamia czytelnikom bogactwa i różnorodności metod kombinowania modeli bazowych. Można odnieść tutaj wrażenia (zapewne mylne), iż również Autor nie jest świadom tej różnorodności, tym bardziej, że przedstawiony przegląd został bezrefleksyjnie powtórzony za [62] (w tej pozycji literaturowej jest jeszcze bayesowska metoda fuzji, którą Autor ujął w pierwszej wersji pracy, a w drugiej wersji – nie wiadomo dlaczego – z niej zrezygnował). A przecież w tym punkcie powinno pojawić się także merytoryczne uzasadnienie wybranej w dalszych badaniach do fuzji modeli regresyjnych metody generalizacji stosów. Również z tego względu *stacking* powinien być przedstawiony w sposób bardziej szczegółowy, aby wszystkie pytania związane z jego realizacją w zadaniu modelowania procesu gojenia ścięgni Achillesa były zaspokojone. Można było tu skorzystać z wymienionej już poz. [62], albo z pracy źródłowej L. Breiman, Stacked Regressions, Machine Learning 24, 49-64 (1996). Recenzent sugeruje Doktorantowi, aby – w przypadku kontynuacji prowadzonych badań z wykorzystaniem metod zespołowych sięgnął do literatury źródłowej i przeglądowej (np. dla klasyfikacji: L. Kuncheva, Combining Pattern Classifiers, John Wiley, 2014, dla regresji: Mendes-Moreira i inni, Ensemble approaches for regression: A survey, ACM Comput. Surv. 45, 10:1 – 10:40), ze szczególnym uwzględnieniem bardzo skutecznego i popularnego schematu DES (*Dynamic Ensemble Selection*) z miarami kompetencji modeli bazowych i dywersyfikacji zespołów modeli (np. R. M. Cruz, R. Sabourin, G. D. Cavalcanti, Dynamic classifier selection: Recent advances and perspectives, Information Fusion 41 (2018) 195–216).

4. Rozdział 5.1. przedstawia kluczową dla pracy oryginalną metodykę Doktoranta pozwalającą poprzez:

- fuzję cech obrazowych (46 cech) pozyskanych z ROI (na marginesie – wyznaczenie ROI wymaga ręcznej segmentacji, a więc proponowana metoda nie jest w pełni automatyczna) i

sieci konwolucyjnej wyszkolonej dla dychotomicznego zadania klasyfikacji (tkanka zdrowa/chora) z redukcją PCA (z 4096 do 200 cech),

- procedurę selekcji z zastosowaniem metody LASSO (z 246 cech do kilkunastu) oraz
- metodę integracji bazowych wyników regresji (*stacking*) dla poszczególnych przekrojów poprzecznych obrazowania RMI ścięgna Achillesa,

na wyznaczenie wartości 6 parametrów opisujących proces gojenia się ścięgna Achillesa.

Ponieważ prezentacja ma wiele luk, Recenzent prosi o wyjaśnienie następujących kwestii:

4.1. Jak zastosowane metody przetwarzania obrazów ogólnego przeznaczenia (cechy statystyczne, cechy Haralicka) pozwalają uwzględnić szczegółową wiedzę o morfologii i strukturze tkanek wewnątrz ościęgna (str. 70)?

4.2. Co to znaczy, że (cyt.) *otrzymano optymalne podzbiory o liczebności poniżej 20 cech* (str. 71)?

4.4. Ile było przekrojów n ?

4.4. Jak wyglądały bazowe regresory R ?

4.5. Jak wyglądało uczenie meta-regresora w *stackingu* (w szczególności dla konkretnych realizacji (SVR, MPR, RF i LR) stosowanych w eksperymentach)?

4.6. Co oznacza TM w (5.1)? Jeśli jest to średnia trymowana wyników regresji R , to gdzie jest miejsce na *stacking*?

4.7. Czy pozyskiwanie cech wykorzystywanych w zadaniu oceny stopnia gojenia się ścięgna Achillesa (w skali 0 – 7) z głębokiej sieci konwolucyjnej szkolonej w binarnym zadaniu klasyfikacji (zdrowy/chory) nie jest procedura błędną metodologicznie? Czy nie byłoby lepiej wytrenować sieć dla zadania klasyfikacji z 8 klasami i wykorzystać tak wyekstrahowane cechy?

4.8. Wartości parametrów oceniających stopień gojenia się ścięgna Achillesa, podawane przez ekspertów były całkowitoliczbowe (chyba?) z przedziału 0 – 7. Czy zatem nie lepiej byłoby potraktowanie problemu automatycznego wyznaczania oceny jako zadania klasyfikacji, a nie regresji? Jak się wydaje, zrobiono tak w drugim eksperymencie z rozdziału 5.3 (rys. 5.12), ale nie stosując klasyfikacji, lecz zaokrąglając wynik regresji.

4.9. Czy sieć (żółty fragment na rys. 5.1) była trenowana dla każdego przekroju (x_i) oddzielnie, co skutkowałoby innymi wyekstrahowanymi cechami dla każdego przekroju, czy też trening był wspólny, czyli wyekstrahowane cechy nie zależały od przekroju?

4.10. Gdzie jest miejsce w strukturze z rys. 5.1. na różne modalności i sekwencje?

5. Wątpliwości związane z badaniami eksperymentalnymi:

5.1. Wciąż niejasna formuła (5.2) i rys. 5.4. Jeżeli H_{PC} jest miarą oceniającą stopień gojenia ścięgna uzyskaną na podstawie jednej cechy z PCA, to dlaczego jej wartości są ujemne?

5.2. Recenzent wciąż nie jest przekonany do sensowności eksperymentu przedstawionego w punkcie 5.3 (*Ocena procesu gojenia*) związanego z rozpoznawaniem (przez eksperta i przez automat) tygodnia badania. Ponieważ dana ta jest oczywiście dostępna, dlatego jej rozpoznanie nie posiada praktycznego znaczenia. Nie jest też prawdą, co napisał Doktorant (cyt.) *Z wykorzystaniem macierzy można określić w jakim stopniu wyniki ocen w poszczególnych tygodniach są do siebie zbliżone* (str. 94). Taki rezultat otrzymamy porównując wyniki automatu i eksperta w rozbiciu na badania z poszczególnych tygodni. Komórki macierzy z rys. 5.10 takich wyników nie zawierają. Komórka (i,j) zawiera

syntetyczną różnicę wyników oceny eksperta i automatu dla badań, które ekspert uznał, że pochodzą z i -tego tygodnia, a automat, że z j -tego. Recenzent nie potrafi wskazać praktycznego znaczenia takiego wyniku. Kolejne pytania związane z eksperymentem: jaka była stosowana metoda klasyfikacji (automatycznej) tygodnia badania i jaka była jej jakość? Jaka była jakość rozpoznawania tygodnia badania przez eksperta? Czy wypełniając ankietę (punkt 5.1.2.) ekspert znał tydzień badania, czy nie? A może należało wykonać eksperyment porównujący oceny eksperta (a więc oceny przyjęte za referencyjne) dla dwóch przypadków (znany/nieznany okres badania) – mielibyśmy wtedy sposobność ocenić, czy ta informacja wpływa na wydany werdykt.

5.3. Metoda autorska SVR jest metodą zespołową, gdyż meta-regresor SVR jest wykorzystany do integracji. Metoda porównywana (*end-to-end*) bazująca na sieci głębokiej to jeden system regresorowy. Czy w tym kontekście nie należy uznać, że wyniki na korzyść SVR (tab. 6.2.) są oczywiste?

6. Spis oznaczeń dodany przez Autora ma sens jedynie wtedy, gdy wymienione w nim symbole są konsekwentnie stosowane w całej pracy. Niestety, w opiniowanej rozprawie tak nie jest. Przykładowo: X (str. 53) – brak w spisie, Y (str. 53) wg. spisu oznacza macierz zmiennych PCA, Z (str. 38) – wg. spisu jest impedancja akustyczna, Z_l z rys. 4.3 nie ma w spisie oznaczeń, x_i (wzór 5.1) – wg. spisu oznacza sygnał wejściowy neuronu.

7. Recenzent prosi o wyjaśnienie, jak – zdaniem Doktoranta – winien brzmieć zwrot związany z uczeniem nadzorowanym: uczenie (szkolenie), czy uczenie (szkolenie) się. Autor w zasadzie używa drugiego terminu, ale czasami pojawia się sformułowanie pierwsze (np. w sformułowaniu celu pobocznego).

5. Podsumowanie recenzji - ocena rozprawy

W pracy niewątpliwie został sformułowany bardzo interesujący (zarówno z poznawczego (teoretycznego), jak i praktycznego punktu widzenia) problem badawczy automatycznej oceny stanu ścięgna Achillesa z wykorzystaniem nowoczesnych metod uczenia maszynowego oraz przetwarzania i analizy informacji obrazowej. Zawarte w niej rezultaty obejmujące metody ekstrakcji cech z przekrojów MRI, ich fuzję z cechami obrazowymi, redukcję i selekcję cech oraz budowę zespołowego systemu regresji wyliczającego wartości liczbowych wskaźników oceny procesu gojenia ścięgna Achillesa, a także różnorodne badania eksperymentalne i analizy porównawcze przeprowadzone na danych rzeczywistych, są oryginalne i stanowią znaczący wkład w dziedzinę automatycznego wspomagania medycznej diagnostyki obrazowej. Autor rozprawy wykazał się dużą wiedzą z medycznego obszaru dziedzinowego i znajomością metod obrazowania medycznego (w szczególności badań MRI), a także opanowaniem sztuki trenowania głębokich sieci konwolucyjnych i sprawnością w posługiwaniu się warsztatem badawczym.

Choć sposób prezentacji wyników pracy i poziom matematycznych formalizmów wciąż jeszcze wywołują niedosyt, to uważam, iż druga wersja pracy doktorskiej Pana mgr. inż. Norberta Kapińskiego lokująca się w dyscyplinie Inżynieria Biomedyczna spełnia wymogi stawiane pracom doktorskim przez odpowiednią Ustawę i wnioskuję o dopuszczenie jej do publicznej obrony.