

Gliwice, 30.11.2022

RECENZJA ROZPRAWY DOKTORSKIEJ mgr inż. Alicji Płuciennik

Tytuł rozprawy: Algorytmy integracji danych i uczenia maszynowego w diagnostyce nowotworów

Przedłożona mi do recenzji rozprawa doktorska Pani mgr. inż. Alicji Płuciennik została wykonana w Katedrze Inżynierii i Biologii Systemów, Wydziału Automatyki, Elektroniki i Informatyki, Politechniki Śląskiej, pod kierunkiem prof. dr hab. inż. Krzysztofa Fujarewicza, we współpracy z firmą Wasco S.A., w ramach programu Doktorat wdrożeniowy.

Promotorem pomocniczym jest dr hab. Małgorzata Oczko-Wojciechowska z Narodowego Instytutu Onkologii im. Marii Skłodowskiej-Curie PIB, Oddział w Gliwicach, a opiekunem przemysłowym dr inż. Michał Bachorz.

Celem pracy było opracowanie strategii integracji danych molekularnych i klinicznych, dla utworzenia nowych narzędzi wspomagających diagnostykę nowotworów piersi i tarczycy. W szczególności celem Doktorantki było wykazanie, czy dodanie do klasyfikatora molekularnego danych klinicznych umożliwi redukcję liczby cech molekularnych koniecznych do oznaczenia, przy zachowaniu lub poprawie jakości klasyfikacji. Cechy molekularne wymagają oznaczenia/testowania metodami zaawansowanej diagnostyki molekularnej, co podnosi koszty, zaś dane kliniczne są zawsze dostępne w ramach rutynowej diagnostyki klinicznej.

W odniesieniu do nowotworów tarczycy, celem hybrydowego molekularno-klinicznego algorytmu diagnostycznego, nad którym pracowała Doktorantka było odróżnienie guzów łagodnych i złośliwych. Jest to ważny problem kliniczny. Standardowa diagnostyka nowotworów tarczycy obejmuje ocenę cytologiczną komórek guza pobranych metodą biopsji aspiracyjnej cienkoigłowej i klasyfikację wyniku zgodnie z systemem klasyfikacji Bethesda. Kategorie rozpoznania III, IV, a także V, w badaniu cytologicznym tarczycy wg klasyfikacji Bethesda stanowią istotny problem kliniczny, dla którego poszukiwane są molekularne klasyfikatory wspomagające diagnostykę, w celu lepszej oceny ryzyka złośliwości guza. Od tego zależą bowiem dalsze decyzje terapeutyczne, radykalność leczenia, i szanse na wyleczenie. Innym aspektem, który może ulec poprawie wskutek poprawy jakości diagnostyki dzięki zastosowaniu mieszanego klasyfikatora molekularno-klinicznego, jest uniknięcie niepotrzebnej terapii i nadmiarowych zabiegów diagnostycznych. Obecnie, około 30% przypadków zaliczanych do kategorii III-V wg. systemu Bethesda jest kierowanych do operacji diagnostycznej, która polega na usunięciu płata tarczycy z guzem, lub całej tarczycy. Dopiero po szczegółowej ocenie histopatologicznej materiału pooperacyjnego możliwe jest postawienie definitywnej diagnozy (rak lub zmiana łagodna). W krajach zachodnich, przede wszystkim w USA, z powodzeniem stosowane są testy molekularne, które poprawiają jakość diagnostyki przypadków wątpliwych (kat. III-V) i pozwalają zredukować

liczbę operacji diagnostycznych do ok. 10%. Testy te są kosztowne i praktycznie nie są w Polsce stosowane. A zatem cel, jakim było stworzenie testu molekularno-klinicznego, poprawiającego dokładność diagnozy w porównaniu do diagnostyki w oparciu o same parametry kliniczne, oraz tańszego niż opracowane na zachodzie testy molekularne, jest jak najbardziej racjonalny.

W przypadku nowotworów piersi, problemem diagnostycznym nad którym pracowała Doktorantka było odróżnienie/klasyfikacja dwóch podtypów molekularnych raka piersi: podtypu luminalnego A (Lum A) i podtypu luminalnego B (Lum B). Podtypy te różnią się rokowaniem, guzy Lum A wykazują najlepszą odpowiedź na leczenie i najlepsze rokowanie spośród wszystkich podtypów raka piersi; guzy Lum B charakteryzują się gorszym rokowaniem. W niektórych krajach, molekularna klasyfikacja podtypów raka piersi zaczyna być uwzględniana w podejmowaniu decyzji terapeutycznych. Spośród licznych testów wielogenowych opracowanych do klasyfikacji molekularnej raków piersi (np. Mammaprint, Endopredict, PAM50, Breast Cancer Index czy Oncotype DX), wiele jest wciąż jeszcze w trakcie badań klinicznych, niektóre mają już akceptację FDA. Dotychczas tylko jeden z testów, Oncotype DX, został uwzględniony w klasyfikacji AJCC (American Joint Committee on Cancer) w systemie klasyfikacji nowotworów TNM (VIII edycja systemu klasyfikacji z 2016 r.). Wynik tego testu molekularnego, polegającego na oznaczeniu ekspresji 21 genów za pomocą ilościowej reakcji PCR, określa ryzyko nawrotu raka u chorych z guzami ER+ i bez przerzutów do regionalnych węzłów chłonnych, u których zastosowano co najmniej leczenie adiuwantowe tamoksifenem. Z pewnością kolejne wielogenowe klasyfikatory molekularne będą w przyszłości pełnić coraz ważniejszą rolę w diagnostyce raka piersi i kwalifikacji chorych do indywidualnego schematu leczenia. Nowy, molekularno-kliniczny klasyfikator, czulszy i bardziej swoisty oraz tańszy od testów molekularnych jest z pewnością oczekiwany.

Poszukiwanie klasyfikatora molekularno-klinicznego dla raka tarczycy prowadzono w oparciu o serię 200 próbek tkankowych (77 raków i 123 zmiany łagodne), pochodzących od pacjentów Instytutu Onkologii w Gliwicach. Doktorantka dysponowała danymi molekularnymi tych próbek obejmującymi profil ekspresji genów oznaczony za pomocą mikromacierzy oligonukleotydowych Affymetrix Human Transcriptome Array 2.0 oraz dane kliniczne (wiek, płeć, wielkość guza, wynik oceny cytologicznej wg klasyfikacji Bethesda). Opracowano również za pomocą algorytmów wykorzystujących sieci Bayesowskie, syntetyczną zmienną kliniczną nazwaną „ryzykiem klinicznym” (w odniesieniu do raka tarczycy „ryzyko kliniczne” oznacza wystąpienie nowotworu złośliwego, zaś w przypadku raka piersi – gorzej rokującego podtypu Lum B; zmienna ciągła). Jak rozumiem z opisu zawartego w rozdziale 3.1.3., opracowanie algorytmu pozwalającego wyliczyć „ryzyko kliniczne” nie wchodzi w zakres niniejszej pracy i zostało wykonane wcześniej; interesuje mnie, jaki był udział Doktorantki? Jako klasyfikator molekularny stosowano albo autorską (Fujarewicz et al., 2017) 40-genową sygnaturę opracowaną wcześniej w celu diagnostyki raka brodawkowatego albo zaczerpniętą z literatury sygnaturę 163 genów różnicujących złośliwe guzy tarczycy od łagodnych (Alexander et al., 2012); sygnatury zawierały 7 genów wspólnych.

Opracowanie klasyfikatora molekularno-klinicznego dla raka piersi prowadzono w oparciu o dane zaczerpnięte z bazy The Cancer Genome Atlas (TCGA). Dane dotyczyły 264 próbek raka piersi (166 raków LUM A i 98 raków Lum B). Dla próbek tych dostępne były dane molekularne w postaci danych mikromacierzowych i proteomicznych, a także dane kliniczne. W klasyfikatorze wykorzystywano dane o ekspresji 50 genów wchodzących w skład sygnatury PAM50 (Parker et al., 2009). Spośród danych klinicznych wykorzystano dane dotyczące poziomu ekspresji markerów oznaczanych immunohistochemicznie: ER, PR, HER2 oraz stopień zaawansowania wg. skali TMN.

W ramach pracy doktorskiej, mgr inż. Płuciennik testowała dwie strategie fuzji danych, określane jako wczesna i późna fuzja danych. Badano także korelacje pomiędzy cechami klinicznymi („ryzykiem klinicznym”) a cechami molekularnymi, w poszukiwaniu pojedynczych cech, które wykazywałyby silną zależność. Znalaziono poszczególne takie cechy, jednak nie zostały one wymienione, ani omówione. W kolejnym kroku analizowano obecność informacji wzajemnej (mutual information, MI) pomiędzy cechami molekularnymi. Ponownie, nie zostały omówione cechy o wysokiej MI, co mogłoby być interesujące z biologicznego punktu widzenia, pozwalając wskazać geny/białka, które są sprzężone funkcjonalnie lub wykazują inny rodzaj zależności. Jak rozumiem, cechy o wysokiej wartości MI zostały następnie usunięte z klasyfikatora, jako nie wnoszące żadnej nowej informacji, a podrażające koszty testowania. Analizowano także MI dla cech klinicznych i molekularnych, tu nie znalaziono jednak wysokich wartości, które, jak rozumiem, mogłyby wskazywać na możliwość zastąpienia w klasyfikatorze cechy z jednej kategorii cechą z drugiej kategorii.

Analiza wpływu strategii fuzji na jakość klasyfikacji wykazała, że do późnej fuzji danych można z powodzeniem wykorzystać cechy, o których wiadomo, że same wykazują wysoką jakość klasyfikacji. Natomiast jeżeli nie ma pewności, czy dodawane cechy wykazują dobrą jakość klasyfikacji, korzystniejsze jest zastosowanie wczesnej fuzji; wówczas algorytm selekcji cech wybierze najlepsze zmienne.

W odniesieniu do nowotworów tarczycy uzyskano poprawę klasyfikacji po dodaniu do modelu zmiennej klinicznej w postaci przetworzonej zmiennej syntetycznej (tzw. „ryzyko kliniczne”). Dodanie tej cechy klinicznej do modelu pozwoliło zredukować liczbę cech molekularnych, co było zgodne z założeniami projektu. Spośród wszystkich cech klinicznych jedynie stopień klasyfikacji wg systemu Bethesda był cechą wykazującą istotne zależności z cechami molekularnymi i dawał jakość klasyfikacji zbliżoną do najlepszych klasyfikatorów molekularnych. Jednak dodanie jej w procesie fuzji nie poprawiało wyniku klasyfikacji.

W przypadku nowotworów piersi, dodanie do modelu cech klinicznych nie poprawiło jakości klasyfikacji. Dodanie cech klinicznych metodą wczesnej fuzji nie pogarszało znacznie jakości klasyfikacji, natomiast przy zastosowaniu późnej fuzji, klasyfikacja ulegała widocznemu pogorszeniu. Powstaje pytanie, czy zjawisko to nie ma związku z faktem, że użyty klasyfikator molekularny PAM50 został stworzony dokładnie w celu klasyfikacji podtypów molekularnych raka piersi i dlatego w odniesieniu do rozróżnienia między typami Lum A vs. Lum B sprawuje się najlepiej?

W dyskusji Doktorantka wskazuje, że przedstawione metody mogą znaleźć zastosowanie do tworzenia innych hybrydowych klasyfikatorów diagnostycznych, np. wykorzystujących fuzję z danymi z badań obrazowych, czy wykorzystujących inne cechy molekularne (np. mutacje czy zmienność liczby kopii DNA).

W rzeczywistości, metody fuzji danych będące przedmiotem pracy doktorskiej zostały wykorzystane w opracowaniu wdrożenia, które bazuje na jeszcze innych cechach molekularnych, tj. ekspresji zestawu genów oznaczanej metodą ilościowej reakcji RT-PCR. To nie jest jednak jasno w pracy wyrażone. Jak czytamy w rozdziale 6 – fuzja umożliwiła poprawę czułości i specyficzności modelu.

Brakuje mi również w pracy schematu, który pokazywałby zależności pomiędzy zbiorami próbek klinicznych użytych w różnych etapach prac. O ile zestawy próbek użyte do analizy różnych metod fuzji danych są dokładnie opisane, to w opisie wdrożenia (Rozdział 6) pojawiają się enigmatyczne informacje o zestawie 275 próbek raka tarczycy, oraz nieznaną liczbę próbek raka piersi

pochodzących z kolekcji Instytutu Onkologii, które były wykorzystane do opracowania klasyfikatora w oparciu o ilościową reakcję RT-PCR.

Ocena rozprawy doktorskiej

W mojej subiektywnej ocenie rozprawa doktorska jest napisana niestarannie, zarówno pod względem językowym jak i edytorskim. Czyta się ją ciężko; Autorka posługuje się często skrótami myślowymi i żargonem technicznym, zwłaszcza w obszarze omawianych zagadnień obliczeniowych. Sporo jest literówek i błędów edytorskich, a także błędów gramatycznych i stylistycznych, utrudniających rozumienie treści. Przykładem jest zdanie ze str. 9: „Ponieważ diagnostyka przy pomocy BAC oraz klasyfikacji Bethesda nie daje w wielu przypadkach jednoznacznej odpowiedzi co do złośliwości guza.” Co Autorka miała na myśli?

Str. 10: „W wyniku różnic między populacjami oraz zastosowaniem narzędzi statystycznych w przypadku danych genomowych [...] powstało wiele zestawów genów umożliwiających klasyfikację guzów o bardzo dobrej jakości, przeznaczonych do określania złośliwości określonych podtypów guzów, np. raków brodawkowatych [...] – prosiłabym Doktorantkę o sformułowanie tego zdania w taki sposób, aby podmiot był jednoznacznie określony, bez błędów gramatycznych i w sposób jasno zrozumiały.

Z kolei lektura rozdziałów 1.1 i 1.2 przynosi wrażenie, że Doktorantka dość pobieżnie poznała zagadnienia onkologiczne, nad którymi pracowała. Opis zawiera błędy merytoryczne i jest kiepsko napisany. Odnosi się wrażenie, że Doktorantka nie otrzymała odpowiedniego wsparcia w tym obszarze ze strony swoich promotorów, gdyż ta część rozprawy mogłaby wyglądać znacznie lepiej.

Uwagi i pytania do tej części pracy:

- Rutynowa diagnostyka nowotworów tarczycy prowadzona jest w oparciu o ocenę cytologiczną materiału komórkowego z biopsji aspiracyjnej cienkoigłowej – nie tylko w Polsce! Taki jest standard międzynarodowych towarzystw tyreologicznych na całym świecie.
- str. 5 – nie wiem, na jakich danych opiera się Autorka pisząc, że w raku piersi wyróżniamy kilkadziesiąt podtypów? Można prosić o wyliczenie?
- BAC wykonuje się zawsze pod kontrolą USG, a nie „czasami” (str. 10)
- Proszę, żeby Doktorantka wyjaśniła, co oznacza pojęcie „leczenie radykalne” w onkologii, jakie mogą być metody leczenia radykalnego i co ma ono na celu? (str. 9)
- Str. 11 – Ki67 to nie jest receptor. Proszę Doktorantkę o wyjaśnienie, jakie funkcje pełni w komórce białko Ki67
- Str. 13 – proszę o wyjaśnienie, co Autorka rozumie pod określeniem, że sygnatura PAM50 „uznawana jest za złoty standard w diagnostyce molekularnej raka piersi”? Czy na pewno?
- Czy na pewno test RT-qPCR na preparatach parafinowych „może być przeprowadzony w niemal dowolnym laboratorium na świecie”?

Rozdział 3.3.1. - USG, MRI, RTG – to nie są metody „badawcze” tylko diagnostyczne

- „zebrane obserwacje o pacjencie (np. wielkość guza czy ilość zajętych węzłów chłonnych)” – takie dane określa się nazwą „dane kliniczne”, a nie „obserwacje o pacjencie”.

Praca jest też napisana nierówno. Dużo bardziej przejrzyste i dobrym językiem napisane są rozdziały Dyskusja (5) oraz Wdrożenie (6).

Niektóre przykłady błędów językowych, żargonu i braku staranności w redakcji pracy:

Str. 5. – „Jednym z problemów przy leczeniu wspomnianych nowotworów są trudności wynikające z nieprawidłowej diagnostyki, ...” – nie diagnostyki, tylko diagnozy

Str. 8. - „... diagnostyki, która umożliwi na rozpoznanie...”

Str. 17 – „fuzja zbiera dane z różnych pod procesów i następuje jedna fuzja ...”

Str. 18 – „Przykładem pracy wykorzystującej fuzję danych obrazowych do diagnostyki tarczycy jest framework opracowany przez zespół...”

Str. 19 – „... wiele publikacji daje perspektywę na znalezienie klasyfikatora...”

Str. 22 – „Analizy dotyczące zbioru próbek guzów tarczycy w niniejszej pracy prowadzona była...”

Str. 26 – Na guzkach wykonano biopsję ...”

Str. 30 – „... wyróżnia się trzy grupy wiekowe ... mniej niż 20 lat, 20-45 lat, 46-60 lat, więcej niż 60 lat.”

Str. 43. „Histogramy współczynników korelacji pomiędzy cechami molekularnymi a obliczoną zmienną ryzyko w analizowanych zbiorach danych.”

Str. 31 „Sieć dla zbioru cech klinicznych zawiera tylko dwa połączenia – pomiędzy Bethesda a ryzykiem złośliwości oraz [...], przy czym silniejsze jest połączenie z Bethesda.”

Str. 59 – Jakości klasyfikacji dla modeli zbudowanych wyłącznie o cechy kliniczne ... była dość niska ...”

Str. 79 „Wśród zmiennych klinicznych jedyną wykazującą istotne zależności z cechami molekularnymi była Bethesda ...”

Żargon techniczny używany jest na zmianę z poprawnymi określeniami, np.

- na str. 44 znajdujemy eleganckie sformułowanie: „kategoria cytologiczna badania BAC zgodnie z klasyfikacją Bethesda”

- na str. 22, 26 i w wielu innych miejscach znajdujemy żargonowe określenia „procedura preprocessingu”, „wstępny preprocessing danych...” a na schemacie 1.3, czy na str. 16 – „wstępne przetwarzanie (danych)”

- na str. 30 „... z wykorzystaniem k-krotnej krosvalidacji”, a na str 34 „uczenie maszynowe z walidacją krzyżową z wielokrotnym losowaniem”

W spisie skrótów brak wszystkich stosowanych skrótów (np. str. 9 – AUS, FLUS; str. 11 – ASCO); w tekście, w niektórych przypadkach również brak rozwinięcia skrótów przy pierwszym zastosowaniu

Cytowania - Na str. 8 – brak cytowania: „... ze względu na niską efektywność leczenia [?...]”, na str 11 - brak cytowania „... za pomocą symboli w skali ASCO [?]”; cytowanie nr 27 – praca nie jest autorstwa J. Jassema i M Krzakowskiego, tylko pod redakcją tych osób

Autorka stosuje dziwny styl cytowania – w tekście cytowanie numerami, zaś w spis referencji ułożony alfabetycznie, wg. nazwiska pierwszego autora (przy czym nazwiska są wymieniane z podaniem inicjałów imion na początku), i numerowany wg kolejności. Przyznam, że nie spotkałam się dotąd z takim sposobem cytowania referencji. Jest to również o tyle nietypowe, że w pracach doktorskich zazwyczaj cytuje się w treści nazwiska autorów, co zwiększa czytelność pracy, pozwala się łatwo zorientować, o jaką pracę, czy zespół badawczy chodzi.

W mojej ocenie, mocną stroną pracy są schematy (rys 1.3, 2.1, 3.2, 3.3 i tabele – tab. 3.1), które bardzo czytelnie obrazują strukturę procesu obliczeniowego. Niemniej, mam wrażenie, że Rysunek 3.3 mógłby zawierać część D (model bez fuzji danych; dane kliniczne), gdyż w tabeli 4.1 na str. 53 taka analiza jest pokazana (klasyfikacja w oparciu o cechy kliniczne). Jest to zresztą logiczne, bo klasyfikacja w oparciu o cechy kliniczne jest przecież punktem wyjścia do wszelkich dalszych rozważań i analiz. Dobrze są napisane i wiele wyjaśniają także rozdziały 5 i 6.

Ogromnie żałuję, że dysertacja nie została należycie dopracowana, bo to z pewnością podniosłoby jej wartość. Dodatkowo, dla Doktoranta pisanie rozprawy pod opieką promotorów powinno być cennym czasem pozwalającym szkolić się w zakresie pisania tekstów naukowych i technicznych, czego tutaj najwyraźniej zabrakło. Szkoda, bo umiejętność ta jest niezbędna zarówno w dalszej karierze naukowej, jak i zawodowej. Bycie inżynierem nie zwalnia z powinności znajomości języka ojczystego, szczególnie, jeżeli się aspiruje do stopnia doktora, czyli elity intelektualnej.

Nie znam osobiście Doktorantki, ale zasięgnęłam informacji u osób, zarówno z Politechniki jak i z Instytutu Onkologii, z którymi miała okazję współpracować i dowiedziałam się, że cieszy się ona opinią osoby zdolnej i bardzo pracowitej. Ponieważ umiejętności językowe rozwijają się wskutek ćwiczenia, jestem przekonana, że w przyszłości Doktorantka z sukcesem zadba o ten aspekt swojego rozwoju.

Podsumowanie

Abstrahując od moich uwag krytycznych, uważam, że przedstawiona mi do oceny praca podejmuje ważne zagadnienie, istotne z klinicznego punktu widzenia i oryginalne ze strony obliczeniowej, spełniając wymogi stawiane przed doktoratem wdrożeniowym.

Praca ma charakter interdyscyplinarny, plasując się na pograniczu biologii medycznej i technik obliczeniowych, co podnosi jej złożoność i tłumaczy trudności, z którymi musiała się uporać Doktorantka, z wykształcenia inżynier. Praca ma również ważny aspekt wdrożeniowy - jej elementy znalazły zastosowanie w opracowaniu zgłoszenia patentowego dotyczącego testu molekularnego ułatwiającego odróżnianie łagodnych i złośliwych guzów tarczycy. Wyniki wchodzące w skład rozprawy zostały niedawno opublikowane w czasopiśmie IJMS (IF=6,2), co oznacza, że otrzymały pozytywne opinie od co najmniej jeszcze dwóch dalszych recenzentów. Doktorantka jest pierwszym autorem tej pracy.

Praca spełnia warunki określone w art. 13 Ustawy z dnia 14 marca 2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz.U. nr 65, poz. 595 z późn. zm.) w związku z art. 179 ust. 1 ustawy z dn. 3 lipca 2018. Przepisy wprowadzające ustawę –Prawo o szkolnictwie wyższym i nauce (DZ.U. z 2018 r. poz. 1669 z późn.zm.), dlatego wnioskuję do Rady Dyscypliny Inżynieria Biomedyczna Politechniki Śląskiej o dopuszczenie rozprawy mgr inż. Alicji Płuciennik do dalszych etapów przewodu doktorskiego.

Prof. dr hab. Katarzyna Lisowska

Narodowy Instytut Onkologii im. M. Skłodowskiej-Curie PIB, Oddział w Gliwicach