

Prof. dr hab. inż. Ewaryst Tkacz
Politechnika Śląska,
Wydział Inżynierii Biomedycznej,
Katedra Biosensorów i Przetwarzania
Sygnałów Biomedycznych

Zabrze, dn. 26.10.2018 r.

RECENZJA ROZPRAWY DOKTORSKIEJ

Tytuł rozprawy : „Zastosowania nowych metod bioinformatycznych do opisu sekwencji genomu wirusów”

Autor rozprawy : **Mgr Damian PANAS**

Promotor rozprawy : **Dr hab. Piotr Wąż**

1. Cel, zakres i charakter rozprawy

Recenzowana rozprawa doktorska dotyczy, jak wspomniano w jej tytule badania możliwości zastosowania metod bioinformatycznych do opisu sekwencji genomu wirusów. Tematykę tę postrzegam jako niezwykle ważną i aktualną biorąc np. pod uwagę zagadnienie tzw. ksenotransplantacji, czyli pozyskiwania organów odzwierzęcych dla zastosowania u ludzi. Zagadnienie na tym etapie wspomnianego zastosowania wymaga nieco szerszego komentarza, jednak nie jest bezpośrednio związane z tematyką rozprawy i dlatego nie będę tego komentarza czynił w chwili obecnej. Dodam tylko krótko, że aktualnie biorę udział w projekcie o akronimie PERV dotyczącym wbudowywania w ludzki genom endogennych retrowirusów świńskich. Ze względu na fakt, iż praca ma charakter wybitnie nowatorski, w określonym obszarze rozumianym jako pole aplikacyjne, Autor przedstawił również pewne subiektywne rozważania dotyczące możliwych aplikacji dzięki szeroko rozumianym badaniom symulacyjnym. Dzięki temu nabrał przekonania, iż przedstawiona koncepcja wykorzystania Dynamicznej Reprezentacji Sekwencji DNA/RNA jest co najmniej poprawna i gwarantuje ewentualny sukces aplikacyjny poprzedzony rozważeniem i właściwym przedstawieniem zagadnień klasyfikacyjnych, w tym wykorzystujących dopasowanie krzywych do wzorca i obserwacji oraz wykorzystaniem elementów uczenia maszynowego. Trzeba podkreślić, że na każdym z wspomnianych etapów daje się odczuć pewną dojrzałość Autora, śmiałość, czy wręcz pewność siebie w odniesieniu do prezentowanych zagadnień, tak teoretycznych jak i praktycznych. Uważam, że pewnym uzasadnieniem takiego mojego być może subiektywnego wrażenia mogą być sformułowane cele rozprawy, które brzmią:

1. Rozwinąć matematyczny formalizm Dynamicznej Reprezentacji Sekwencji DNA/RNA.
2. Połączyć Dynamiczną Reprezentację Sekwencji DNA/RNA z wybranymi algorytmami uczenia maszynowego.
3. Zbadać możliwości Dynamicznej Reprezentacji Sekwencji DNA/RNA pod kątem wykorzystania jej w filogenetyce molekularnej.
4. Określić jakość deskryptorów Dynamicznej Reprezentacji Sekwencji DNA/RNA poprzez ich korelacje, zużycie w drzewach losowych oraz algorytm Boruta.

Zapewnienie realizacji wspomnianych celów rozprawy Doktorant osiągnął poprzez skuteczną próbę udowodnienia następujących tez:

- Dynamiczna Reprezentacja Sekwencji DNA/RNA może stanowić alternatywę wobec powszechnie stosowanych metod dopasowania sekwencji.
- Dwuwymiarowa Dynamiczna Reprezentacja Sekwencji DNA/RNA może być wykorzystana w analizie filogenetycznej wirusów.
- Połączenie Dynamicznej Reprezentacji Sekwencji DNA/RNA z takimi algorytmami uczenia maszynowego jak C5.0 lub las losowy, umożliwia identyfikowanie nieznanymi szczepów wirusów, w szczególności podtypów wirusa grypy typu A.

W tym miejscu recenzji rozprawy doktorskiej trzeba odnieść się do stanu podobnych badań na świecie. Tematyka bowiem nie jest zupełnie nowa i nikt dotąd się nią nie zajmował. Świadczy o tym dość obszerny, liczący 116 pozycji spis literatury. Jednak sens zajmowania się tematem jest dość prosty do uzasadnienia, i to właśnie w poszczególnych rozdziałach pracy zaprezentował Doktorant, gdyż wyimaginowana krzywa opisująca potrzeby w tym zakresie w moim odczuciu ma tendencję wyraźnie rosnącą. Zatem zajmowanie się tematyką opisaną w rozprawie jest od strony teoretycznej wybitnie uzasadnione, co potwierdza praktyka w tym zakresie. Chcemy żyć coraz bardziej komfortowo, zapewniając przy tym niezbędne podstawowe potrzeby a wśród nich możliwość projektowania elementów budowy i cech funkcjonalnych tkanek w odniesieniu lub może przy użyciu narzędzi do biologii molekularnej. W takim właśnie kontekście lokują charakter rozprawy.

2. Zawartość rozprawy

Rozprawa doktorska mgr Damiana Panasa jest stosunkowo niezbyt obszernym 84 stronicowym opracowaniem precyzyjnie określonego tematu i składa się z 6 rozdziałów obejmujących po kolei:

- **wprowadzenie**, (w którym przedstawiono kilka uwag o charakterze bardziej filozoficznym wraz z naświetleniem historycznym, które za zadanie mają wprowadzić ewentualnego czytelnika w celowość realizacji badań sformułowanych w tytule rozprawy);
- **cele i tezy pracy**, (który przedstawia precyzyjnie sformułowane 4 cele rozprawy oraz wyraża przekonanie, iż dla ich realizacji potrzebne będzie udowodnienie 3 tez);
- **podstawa biologiczna** (zawiera opis podstaw biologicznych czyli genomikę jak również mechanizmy ewolucji genomów czy filogenetykę molekularną; na zakończenie rozdział ten koncentruje uwagę recenzenta na zasadniczym przedmiocie badań eksperymentalnych tj. na wirusach: Zika i grypy);
- **wykorzystane metody** (gdzie Autor pokusił się o zaprezentowanie metod wykorzystywanych w dalszej części rozprawy; przedstawił ogólnie w jakim celu próbujemy dopasowywać sekwencje, jak możemy je ze sobą nawzajem porównywać, wreszcie po co jest nam potrzebna dynamiczna reprezentacja sekwencji DNA/RNA w szczególności w odniesieniu do pozostałych istniejących metod służących między innymi do podobnych celów);
- **badania eksperymentalne** (to zdecydowanie najciekawszy rozdział rozprawy zawierający wyniki badań oraz właściwą ich prezentację: analityczną, graficzną i tabelaryczną; należy jednak krytycznie stwierdzić, że Autor być może ze względu na swoją skromność jest dość oszczędny w słowach, gdyż w tym właśnie rozdziale, po wykonaniu sporej liczby badań można by najlepiej przedstawić wartość naukową rozprawy, nowatorstwo podejścia oraz pewną przydatność praktyczną).

- **Podsumowanie oraz wnioski z badań** (ten rozdział to zaledwie trzy strony niezwykle syntetycznego podsumowania odnośnie osiągnięcia celów i udowodnienia tez wraz ze stosownym komentarzem).
- **Dwa dodatki:**
 - A: Grafy dynamiczne szczepów ZIKV
 - B: Wykorzystane szczepy IAV

Rozprawa doktorska zakończona jest, wspomnianym już wcześniej spisem bibliograficznym zawierającym 116 pozycji, w tym dwoma pracami współautorstwa Doktoranta. Od strony edytorskiej rozprawa doktorska mgr Damiana Panasa zaopatrzona jest w spis rysunków i spis tabel, których w tekście pracy jest sporo. Znajduje się w niej również spis używanych skrótów i oznaczeń, których jest nie mało. Te cechy świadczą niewątpliwie o pewnej dojrzałości Autora rozprawy. Nie odmówię sobie przyjemności docenienia Autora za przygotowanie tekstu w środowisku LaTeX, czyli profesjonalnego edytora tekstów z kompilatorem. Ponadto Autor zadbał o pewien komfort pracy recenzenta polegający na tym, iż wszystkie pozycje w spisie literatury zaopatrzone są w informację odnośnie cytowania w pracy. Wiadomo bowiem, że spis literatury nie może mieć charakteru redundancyjnego (nadmiarowego). Nie umieszcza się w nim pozycje, które nie są później cytowane w tekście. We wspomnianym środowisku istnieje pewien mechanizm ułatwiający ten aspekt pracy recenzenta.

Wrażenie ogólne z czytania rozprawy jest bardzo dobre, ponieważ została ona przygotowana w sposób profesjonalny, a może bardziej precyzyjnie, przy użyciu profesjonalnych narzędzi. Zaznaczam, że to pozytywne wrażenie nie wyklucza poprzednich uwag (oszczędność w formułowaniu myśli), natomiast chcę wyraźnie podkreślić, że w moim odczuciu Doktorant osiągnął cel polegający na pokazaniu recenzentowi elementów swojego profesjonalizmu.

3. Poprawność i oryginalność tezy rozprawy

Obserwując niezwykle dynamiczny rozwój poszczególnych obszarów inżynierii biomedycznej, w szczególności zaś bioinformatyki dokonujący się w ostatnich kilkunastu latach, z pełną odpowiedzialnością uznaję tematykę rozprawy, określone w niej cele jak też ich realizację za nadzwyczajnie aktualną. Na tak zdefiniowanym tle, praca doktorska mgr Damiana Panasa wnosi istotny wkład w rozwój badań w obszarze metod analitycznych dotyczących genomu mówiąc ogólnie zaś w szczególności dopasowania sekwencji wirusów.

W związku z powyższym uznaję postawione w rozprawie tezy dotyczące potrzeby opracowania nowej filozofii metrologicznej łączącej zalety innych metod pomiarów sekwencji genetycznych za poprawne, oryginalne oraz przede wszystkim za udowodnione w stopniu wystarczającym do potwierdzenia kwalifikacji Autora aplikującego o stopień naukowy doktora nauk o zdrowiu.

4. Analiza źródeł

Wspomniano już wcześniej, że w rozprawie włączono spis bibliograficzny liczący 116 pozycji. Trzeba jednak zaznaczyć, że poruszana w rozprawie tematyka, czerpiąca inspirację z nauk biomedycznych, obfituje w niezwykle bogactwo literatury w zakresie zagadnień dotyczących analizy genomu ogólnie, w szczególności zaś dopasowania sekwencji wirusów. Przełożenie tych inspiracji na obszar zagadnień z zakresu „sekwenomiki” (takiego terminu użył przedstawiciel firmy Illumina oferującej Sekwenatory Nowej Generacji) mówiąc bardzo ogólnie, jest jednak zagadnieniem nowym, zatem w moim odczuciu Autor wykazał znacznie

większą niż dostateczna wiedzę, odnoszącą się do wyboru źródeł oraz sposobu ich wykorzystania w tekście rozprawy.

Wydaje się, że pozycja rozprawy w stosunku do stanu wiedzy i aktualnych możliwości technicznych i technologicznych reprezentowanych przez literaturę światową w obszarze metod sekwencjonowania genomu przy pomocy metodologii uczenia maszynowego jest znacznie więcej niż przyzwoita. Niektóre, drobne uchybienia wskazane poniżej mogą być usprawiedliwione sporym wskaźnikiem elementów nowatorskich zaczerpniętych z nauk doświadczalnych.

5. Znaczenie uzyskanych wyników dla dyscypliny naukowej

Rozprawa doktorska mgr Damiana Panasa dokonuje wnikliwej analizy możliwych do zastosowania metod dotyczących wykorzystania uczenia maszynowego w analizie sekwencji genetycznych i niestety tę część pracy, ustosunkowując się do niej krytycznie należy uznać za nieco odtwórczą, chociaż, co warto podkreślić, Autor umiejętnie porusza się we wszystkich nachodzących na siebie obszarach problemowych proponując w ten sposób nowe podejście lub nawet kreując pewien nowy standard postępowania w odniesieniu do istniejących metod sekwencjonowania. Chcę więc podkreślić, że niezwykle trudno jednoznacznie ocenić znaczenie zaprezentowanych wyników dla dyscypliny naukowej, ponieważ jeśli udałoby się je jeszcze w inny sposób zweryfikować, nie tylko poprzez badania symulacyjne, to może ono być ogromne. Zaprezentowane w pracy przykłady zawierają wprawdzie wyniki analiz sekwencjonowania szczepów wirusów pochodzących z różnych obszarów Ziemi, ale wrażenie moje, jako recenzenta jest, że większy nacisk położono na sprawy biologiczne, niż na elementy interpretacyjne wynikające z przeprowadzonych badań, których w pracy jest co najmniej wystarczająca ilość. Stąd, poproszę Doktoranta o odpowiedź na pytanie teoretyczne dotyczące metody Dynamic Time-Warping Method odnoszącej się do sekwencjonowania ciągów nukleotydów o nierównej długości; poproszę o wskazanie wad i zalet takiej metody alternatywnej w stosunku do zawartości rozprawy. Z wielką chęcią ponadto poznałbym ponadto stosunek Doktoranta do zaproponowanej przez L.A. Zadeha w 1965 roku logiki rozmytej dla klasyfikacji analizowanych sekwencji. Czy zdaniem pana magistra pewna subtelna probabilistyka, w postaci funkcji przynależności może w pana analizach znaleźć punkt zaczepienia.

6. Uwagi krytyczne

W świetle przedstawionych dotąd elementów rozprawy doktorskiej mgr Damiana Panasa z radością muszę podkreślić, że uwag krytycznych nie mam zbyt dużo. Jednak, aby recenzja nie była kolorową laurką muszę, wspomnieć, o co najmniej kilku. Dotyczą one głównie niektórych nadużywanych w pracy sformułowań, jak np. „...zidentyfikowanie podtypów...”, co jest wysoce nieprecyzyjne biorąc pod uwagę definicję pojęcia **identyfikacja**. Na stronie 18 Autor piszę o „...rozprzesztrzeniu się wirusa...”. Na tej samej stronie dalej napisano „...NS5 kodujące polimerazę RNA zależną od RNA...”. Bez utraty ogólności sformułowania ominąłbym to „...zależną od RNA...”. W pracy znalazłem też trochę literówek. Cześć z nich, na szczęście, jest jednak niewykrywalna dla większości tzw. „check-spellerów”. Te, wymienione przeze mnie, momentami niezręczne sformułowania są jednak zrekompensowane w znakomity sposób doskonałą jakością rysunków, jak się domyślam autorstwa w większości lub może w całości Doktoranta. Sprawiają one, swoją starannością, świetne wrażenie podwyższające jakość pracy jako całości i to oczywiście nie jest uwaga krytyczna.

7. Wniosek końcowy

Biorąc pod uwagę wskazane powyżej elementy rozprawy doktorskiej mgr Damiana Panasa stwierdzam, że przedstawiona mi do recenzji rozprawa doktorska spełnia z nadmiarem wymagania Ustawy o stopniach i tytule naukowym z dnia 14 marca 2003 (art.13 ust.1 i ust.2). W związku z powyższym, w oparciu o tę recenzję, wnoszę do Wysokiej Rady Wydziału Nauk o Zdrowi Gdańskiego Uniwersytetu Medycznego o dopuszczenie Autora do dalszych etapów przewodu doktorskiego. W szczególności do publicznej obrony rozprawy.

Dodatkowo, analizując sylwetkę Doktoranta, a w szczególności Jego dotychczasowe osiągnięcia naukowe, w tym 2 opublikowane artykuły w czasopismach z tzw. listy filadelfijskiej wnoszę do Wysokiej Rady postulat dotyczący rozważenia możliwości wyróżnienia recenzowanej rozprawy doktorskiej. Wniosek ten poparty jest przekonaniem, iż zarówno stopień naukowy jak i ewentualne wyróżnienie dostaną się w ręce bardzo dobrze rokującego Kandydata.