

Dr hab. inż. Paweł Piotr Łabaj  
Małopolskie Centrum Biotechnologii  
Uniwersytet Jagielloński  
Gronostajowa 7a, 30-387 Kraków  
[pawel.labaj@uj.edu.pl](mailto:pawel.labaj@uj.edu.pl)

Kraków, 29.05.2023

## Recenzja rozprawy doktorskiej

Tytuł rozprawy: Machine learning methods in support of multiomics signature identification for breast cancer patient subpopulations

Autor rozprawy: mgr inż. Joanna Tobiasz

Promotor rozprawy: Prof. dr hab. inż. Joanna Polańska

Dziedzina: nauki inżynieryjno-techniczne

Dyscyplina: Inżynieria Biomedyczna

Rozwój wysokoprzepustowych biomedycznych technik pomiarowych pozwolił w ostatnich latach na znaczny postęp w naukach biologicznych i medycznych. Umożliwił również na dokładniejsze przyjrzenie się stosowanym od lat klasyfikacjom klinicznym wielu chorób ale z punktu widzenia szerokiej/głębokiej charakteryzacji molekularnej. Jednym z przykładów jest nowotwór piersi gdzie stosowana powszechnie klasyfikacja kliniczna oparta jest na poziomie kilku białek markerowych i pozostaje niezmienna od lat. Natomiast istniejąca, również już jakiś czas, klasyfikacja molekularna, bazująca na ekspresji genów, nie oddaje w pełni dużego zróżnicowania obrazu molekularnego i związanego z tym zróżnicowania zachowań klinicznych. Jednak właśnie dzięki dostępności wysokoprzepustowych technik pomiarowych otwiera się możliwość nowego spojrzenia na kwestię klasyfikacji nowotworu piersi. Dokładniejsze scharakteryzowanie jego podtypów może pomóc w lepszej stratyfikacji pacjentów oraz w optymalnym dobraniu procedury leczniczej. Jednocześnie zrozumienie określonej charakterystyki może wesprzeć poszukiwanie nowych, bardziej efektywnych procedur terapeutycznych.

Taka dokładniejsza charakteryzacja wymaga analizy znacznej ilości heterogennych danych generowanych przez wysokoprzepustowe techniki pomiarowe. Konieczne jest zastosowanie właściwych technik „eksploracji danych” a następnie „uczenia maszynowego”. W przedstawionej rozprawie doktorskiej Autorka adresuje problem zastosowania metod uczenia maszynowego do identyfikacji i klinicznego oraz molekularnego scharakteryzowania podpopulacji pacjentek dotkniętych nowotworem piersi. Analizując stosowane obecnie podejścia Autorka przetestowała kombinacje metod inżynierii cech i klastrowania a następnie porównała je w oparciu o zaproponowane

metryki. W rezultacie Autorka zidentyfikowała w oparciu o profil białkowy sześć podpopulacji pacjentek gdzie trzy wykazały dużą zgodność ze stosowanymi podtypami opartymi o poziom mRNA, a kolejne trzy okazały się być podkategoriami kolejnego podtypu opartego o poziom mRNA. Zidentyfikowane przez Autorkę podpopulacje różnią się przeżywalnością oraz profilami białkowymi i transkryptomycznymi. W toku analizy zidentyfikowano zestawy markerów specyficznych dla podpopulacji oraz skonstruowano sygnatury umożliwiające rozróżnienie wszystkich podtypów. Zauważono jednak, że dla nowo zidentyfikowanych podtypów zróżnicowanie w przestrzeni transkryptomicznej jest mniejsze niż w białkowej (do tej obserwacji odniosę się poniżej). Wykonana przez Autorkę analiza jasno wykazuje, że dotychczas stosowane klasyfikacje były niekompletne i że wykryte podpopulacje, po dodatkowej walidacji, powinny być włączone do praktyki. Natomiast uzyskane sygnatury powinny stać się przyczynkiem do dalszych badań w celu lepszego zrozumienia podtypów nowotworu piersi i opracowania nowych terapii celowanych w rozwoju medycyny spersonalizowanej.

Układ rozprawy jest typowy dla tego typu opracowań. W kolejnych rozdziałach Autorka:

- opisuje „biologiczne tło” dla prowadzonych analiz przybliżające czytelnikowi zarówno sam nowotwór piersi jak i stosowane podejścia do jego klasyfikacji (Rozdział 2),
- przedstawia zestaw danych użytych w analizie jak i przybliża rozpatrywane podejścia analizy, w rozdziale tym zdefiniowane są również metryki wykorzystane następnie do porównania (Rozdział 3),
- przedstawia wyniki stosowania różnych podejść klasteryzacji, omówienie ich porównania jak i argumentację za ostatecznym wyborem (Rozdział 4),
- przedstawia charakterystykę kliniczną dla wybranego podziału na subpopulacje (Rozdział 5)
- przedstawia charakterystykę molekularną dla wybranego podziału na subpopulacje wraz z identyfikacją potencjalnych markerów różnicujących (Rozdział 6)
- prezentuje podsumowanie pracy z uwypukleniem głównych osiągnięć (Rozdział 7)
- przedstawia materiały dodatkowe (Rozdział 8)

Do pracy został również dołączony imponujący wykaz dorobku naukowego Doktorantki.

Autorka bardzo dokładnie omawia obecny stan wiedzy co świadczy o dużym odczytaniu i dobrym przygotowaniu do podjęcia zagadnień poruszanych w dalszych częściach pracy. Na wyróżnienie zasługuje fakt, że pomimo ogromu zaprezentowanego materiału czytelnik jest przeprowadzany przez kolejne rozdziały w bardzo umiejętny sposób. Cały układ pracy jest bardzo dobrze przemyślany przez co czytelnik w żadnym momencie nie traci z oczu celu pracy. Ogrom prezentowanych wyników nie wywołuje uczucia przytłoczenia czy zagubienia, a wręcz dzięki umiejętnemu podaniu wzbudza podziw.

Tezy rozprawy są sformułowane jasno i przystępnie oraz są w pełni poparte danymi zawartymi w poszczególnych rozdziałach rozprawy. Podsumowanie rozprawy jest



syntetycznym wykazaniem, że założone w pracy cele zostały osiągnięte. Doktorantka dla każdej z tez wskazuje, która część rozprawy się do niej odnosi.

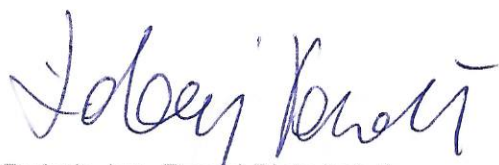
Tak naprawdę „przyczepić” się mogę tylko do dwóch kwestii: i) W 3.2.2 jednym z omawianych narzędzi jest `combat` z pakietu `sva`. Jeśli Autorka sięgnęła do tego pakietu to należało również omówić pozostałe funkcje, a te umożliwiają nie tylko usunięcie znanego „efektu paczki” ale również identyfikację nieznanymi czynników zakłócających i ich późniejsze wykorzystanie w modelu; ii) Autorka zauważa, że dla nowo zidentyfikowanych podtypów zróżnicowanie w przestrzeni transkryptomowej jest mniejsze niż w białkowej. Brakuje mi tutaj zastanowienia się nad możliwym powodem. W mojej opinii, ponieważ analiza transkryptomowa była wykonywana w oparciu o mikromacierze powodem mogą być ograniczenia tej technologii, które da się częściowo przezwyciężyć jak pokazała praca konsorcjum SEQC [<https://doi.org/10.1038%2Fnb.2957>]. Jeśli Autorka będzie kontynuowała ten temat sugeruję również przyjrzenie się danym RNA-Seq. Być może wtedy obraz transkryptomowy będzie pełniejszy.

Rozprawa zawiera 51 szczegółowych rycin oraz 30 tabel, które co do zasady są jasne i czytelne. Piśmiennictwo obejmuje 168 dobrze dobranych i aktualnych pozycji.

Podsumowując, przedstawiona do oceny praca doktorska stanowi bardzo wartościowe uzupełnienie obecnego stanu wiedzy odnośnie klasyfikacji podpopulacji pacjentek dotkniętych nowotworem piersi. Praca ta jednocześnie wyznacza kierunek dalszych badań w celu lepszego zrozumienia podtypów nowotworu piersi i opracowania nowych terapii celowanych w rozwoju medycyny spersonalizowanej. Praca ta w pełni odpowiada warunkom stawianym rozprawom doktorskim oraz wypełnia istotną lukę w obecnym stanie wiedzy. Należy też podkreślić ogrom pracy wykonanej przez Doktorantkę jak również imponujący dorobek naukowy.

Na podstawie powyższej oceny stwierdzam, że wymieniona rozprawa doktorska w pełni odpowiada warunkom stawianym w ustawie Prawo o szkolnictwie wyższym i nauce / Dz. U. z 2022 r. poz. 574, w zakresie nadawania stopni naukowych i na tej podstawie wnoszę do Wysokiej Rady Dyscypliny Inżynieria Biomedyczna Politechniki Śląskiej o dopuszczenie mgr inż. Joanny Tobiasz do dalszych etapów przewodu doktorskiego. Równocześnie chciałbym zgłosić pod dyskusję przyznanie wyróżnienia. Motywacją jest zarówno ogrom jak i wysoka jakość wykonanej pracy ale także bardzo wysoka jakość przygotowanej rozprawy.

Nie mam wątpliwości, że doświadczenie zgromadzone przez Autorkę stawia cały zespół badawczy w doskonałej pozycji wśród międzynarodowych grup zajmujących się tą tematyką.



Dr hab. inż. Paweł Piotr Łabaj