

Recenzja rozprawy doktorskiej
mgr. Oleksandra Myronova
pt. “In-silico Modeling of Antigen Recognition During Immune
Response by Analyzing the Sequential and Structural
Peptide-HLA-TCR Data Using Machine Learning”

1. Problematyka naukowa rozprawy

Procesy zachodzące w organizmie i związane z układem odpornościowym w dużej mierze bazują na rozpoznawalności przez organizm kształtów przestrzennych cząsteczek patogennych. Rozpoznawalności tej i odpowiedniej reakcji organizm uczy się na drodze kontaktu, gdzie dopasowanie powierzchniowe cząsteczki patogenu do cząsteczki białkowej układu odpornościowego jest kluczowe w walce z infekcją. Podobne procesy mają miejsce również w odniesieniu do zagrożenia, jakie stanowią komórki nowotworowe. Badanie możliwości i konsekwencji takiego kontaktu pozwala na lepsze zrozumienie odpowiedzi immunologicznej organizmu i wpływa na doskonalenie podejść terapeutycznych, w tym produktów farmaceutycznych.

Tematyka rozprawy wpisuje się w ten obszar badań i porusza bioinformatyczne aspekty związane z procesami zachodzącymi w trakcie zagrożenia immunologicznego u człowieka. Koncentruje się na wybranych rodzajach patogenów/komórek nowotworowych, dla których badane są szczegółowo powiązane procesy wiązania się cząsteczki patogennej (antygeny) z białkiem układu odpornościowego (przeciwciałem) i wywołania odpowiedzi odpornościowej organizmu. W tym celu mgr Oleksandr Myronov analizuje dane eksperymentalne pozyskane z użyciem nowoczesnych technologii, na drodze spektrometrii mas czy metod sekwencjonowania nowej generacji. Analizę swoją wspomaga podejściami z zakresu uczenia maszynowego, w szczególności sztucznych sieci neuronowych. Swoją pracą poszerza obecny stan wiedzy i zbiór dostępnych powiązanych narzędzi bioinformatycznych. Tematyka ta może być zatem przedmiotem rozprawy doktorskiej w dyscyplinie informatyka techniczna i telekomunikacja.

2. Opinia o rozprawie

Rozprawa składa się z czterech rozdziałów i bibliografii. Rozdział pierwszy zawiera przedstawienie celu i zakresu pracy oraz wprowadzenie do tematyki rozprawy od strony biologicznej i uczenia maszynowego. Rozdział drugi stanowi szczegółowe omówienie metod z zakresu uczenia maszynowego zastosowanych w ramach pracy doktorskiej, jak również zbiorów danych biologicznych wykorzystanych w testach. Rozdział trzeci, o rozbudowanej strukturze, poświęcony jest przedstawieniu wyników pracy własnej Doktoranta. Rozdział czwarty zawiera wnioski końcowe.

Modele i metody przedstawione w rozdziale trzecim służą, w ogólności, przewidywaniu reakcji układu immunologicznego człowieka na obecność wybranego antygeny. Rozwiązują trzy problemy z tym związane, do tego celu wykorzystując sztuczne sieci neuronowe, które uczone są na obszernych zbiorach danych eksperymentalnych. Predykcja taka ma duże znaczenie w praktyce, zwłaszcza w przemyśle farmaceutycznym, gdyż pozwala wstępnie zweryfikować wiele podejść i modeli bez kosztownych i czasochłonnych eksperymentów laboratoryjnych i testów klinicznych. Predykcja została oparta na klasyfikacji odwołującej się do sekwencji aminokwasowych peptydów i białek.

Pierwszy z rozwiązywanych problemów związany jest z wykrywaniem prezentacji patogenego peptydu na powierzchni komórki, co powinno prowadzić do uruchomienia reakcji obronnej organizmu. Doktorant współpracował model konwolucyjnej sieci neuronowej, która na podstawie sekwencji aminokwasowej peptydu i rodzaju cząsteczki HLA (reprezentowanej również przez sekwencję aminokwasów) klasyfikuje kompleks peptyd-HLA w kontekście odpowiedzi immunologicznej. Szczególny nacisk położył na przygotowanie reprezentatywnego zbioru danych do procesu uczenia modelu, w którym uwzględnił sztucznie wygenerowane przykłady z odpowiedzią negatywną. Inny model dwukierunkowej sieci neuronowej LSTM ukierunkowany został na następnym po prezentacji peptydu etapie tego procesu, wzbudzenie odpowiedzi immunologicznej organizmu. Na wejściu założono ten sam rodzaj danych jak poprzednio, czyli sekwencję aminokwasową peptydu i rodzaj cząsteczki HLA, natomiast inne były zbiory uczące, które obejmowały odpowiedzi pozytywne i negatywne i nie wymagały uzupełniania o sztucznie wygenerowane dane. W tym kontekście sprawdzony został również wyżej wspomniany model sieci konwolucyjnej. Modele te weszły w skład komercyjnego narzędzia ArdImmune Rank i zostały zweryfikowane w praktycznych badaniach dotyczących przewidywania reakcji układu odpornościowego na obecność patogenów związanych z chorobami COVID-19 i nowotworową.

Inny problem związany z reakcją organizmu na patogenne peptydy dotyczy wykrywania możliwości utworzenia kompleksu peptyd-TCR, które to zdarzenie może zajść po prezentacji peptydu na powierzchni komórki. Doktorant opracował dla tego problemu model sieci typu transformer, przyjmującej na wejściu sekwencje aminokwasowe peptydu i cząsteczki TCR. Tutaj również dużą uwagę poświęcił przygotowaniu sztucznie wygenerowanych negatywnych przykładów, o które uzupełniony został zbiór użyty w procesie uczenia modelu. Celem poprawienia jakości zbioru uczącego odfiltrowuje niektóre cząsteczki TCR, m.in. na podstawie wyników klastrowania. Weryfikacja modelu na danych obejmujących patogeny

wirusowe i nowotworowe wykazała w jego przypadku skuteczniejszą predykcję niż w przypadku czterech innych klasyfikatorów dedykowanych temu problemowi.

Formułując swoją ocenę rozprawy, pragnę na wstępie podkreślić dużą wiedzę mgr. Oleksandra Myronova nt. biologicznej strony rozwiązywanych problemów, dostępnych wyników badań eksperymentalnych z nimi związanych i aktualnego stanu prac z zakresu immunologii obliczeniowej prezentowanych w literaturze. Dzięki temu modele sieci neuronowych, które opisuje w rozprawie, mogły zostać w odpowiedni sposób wyuczone i zweryfikowane, a to z kolei prowadzi do wiarygodnych wniosków dotyczących przewidywanych reakcji układu odpornościowego człowieka.

Do najważniejszych osiągnięć badawczych przedstawionych w rozprawie zaliczyłabym:

- Wkład w rozwój komercyjnego narzędzia bioinformatycznego, które zostało użyte w praktyce celem przyczynienia się do postępów w aktualnych i ważnych społecznie badaniach biotechnologicznych prowadzonych w kierunku opracowania szczepionek na COVID-19 i chorobę nowotworową. Szczególny wkład w obszerne analizy opisane w podrozdziale 3.8, prowadzone w celu zbadania odpowiedzi immunologicznej organizmu człowieka na antygeny nowotworowe różnego rodzaju.
- Opracowanie metody BERTrand przewyższającej inne znane z literatury klasyfikatory oparte na sieciach neuronowych, na której skuteczność miało wpływ nowe podejście do przygotowania danych uczących.
- Uwzględnienie w fazach uczenia modeli i ich testowania obszernych i uznanych zbiorów danych biomedycznych, pochodzących z eksperymentów z zastosowaniem technik laboratoryjnych o różnorodnym charakterze: spektrometrii mas, sekwencjonowania nowej generacji, testów ELISA i ELISPOT.

W trakcie czytania rozprawy nasunęły mi się następujące uwagi krytyczne:

- Doktorant nie precyzuje (z wyjątkiem podrozdziału 3.9), jaki był jego udział w tworzeniu architektury proponowanych sieci neuronowych i implementacji oprogramowania. Większą uwagę poświęca swojemu udziałowi w trenowaniu stworzonych modeli i ich eksperymentalnej weryfikacji, chociaż i tu trudno niekiedy określić precyzyjnie jego wkład.
- Wspomina, że był jednym z twórców narzędzia ArdImmune Vax i że składało się na nie m.in. narzędzie ArdImmune Rank, ale nie podaje, czy i w jakim zakresie jego wkład w ArdImmune Vax wykraczał poza ramy ArdImmune Rank.
- W odniesieniu do przedstawionych w podrozdziale 3.7 wyników badań zabrakło wzmianki, jaki był w nich udział Doktoranta.
- W podrozdziale 3.9 Doktorant pisze o grupowaniu sekwencji peptydowych na podstawie odległości Levenshteina. Czy w tym grupowaniu została uwzględniona macierz substytucji aminokwasów?

Powyższe uwagi nie podważają mojej pozytywnej oceny pracy naukowej Doktoranta. Układ rozprawy i wykorzystanie źródeł literaturowych nie budzą zastrzeżeń.

3. Podsumowanie

Pan mgr Oleksandr Myronov swoją pracą doktorską przyczynił się do uzyskania wyników istotnych w obszarze immunologii obliczeniowej. Do rozwiązania powiązanych z tym obszarem problemów bioinformatycznych zastosował zaawansowane podejścia z zakresu uczenia maszynowego, współtworzył modele sieci neuronowych, a do ich uczenia zastosował zróżnicowane i najnowsze dane eksperymentalne, które opracował w autorski sposób, uwzględniając potrzebę odpowiedniej ich reprezentacji. Wykazał się zatem wiedzą i umiejętnościami z zakresu informatyki, ale także dodatkową głęboką znajomością problemów i literatury biologicznej. Wyniki przedstawione w rozprawie zostały opublikowane w dwóch bardzo dobrych czasopismach, *Frontiers in Genetics* (140 pkt. MEiN) i *Bioinformatics* (200 pkt.); w tym drugim, które jest uznawane za najlepsze czasopismo bioinformatyczne, Doktorant jest pierwszym autorem.

Wartość badań Doktoranta dodatkowo potwierdzona została praktycznym wykorzystaniem jego pracy w firmie bioinformatycznej działającej w obszarze zastosowań medycznych, w której jest zatrudniony. To wdrożenie uważam za dodatkowy aspekt przemawiający na korzyść tej pracy.

Na podstawie wyrażonych powyżej opinii stwierdzam, że rozprawa pt. “In-silico Modeling of Antigen Recognition During Immune Response by Analyzing the Sequential and Structural Peptide-HLA-TCR Data Using Machine Learning” autorstwa mgr. Oleksandra Myronova spełnia warunki stawiane rozprawom doktorskim przez obowiązującą ustawę o stopniach naukowych i tytule naukowym. Wnoszę o dopuszczenie tej rozprawy do publicznej obrony.