

dr hab. n. med. Magdalena Spólnicka
Centrum Nauk Sądowych
Uniwersytet Warszawski

Warszawa, dn. 08.07.2024 r.

Recenzja rozprawy doktorskiej

POMORSKI UNIWERSYTET MEDYCZNY W SZCZECINIE

Tytuł: ANALIZA METYLACJI SEKWENCJI CpG W OPARCIU O UCZENIE
MASZYNOWE I SIECI NEURONOWE

Autor: lek. Tomasz Falgowski

Dyscyplina: nauki medyczne

Promotor: prof. dr hab. n. med. Tadeusz Sulikowski

W ostatniej dekadzie rozwój nowoczesnych metod analizy genetycznej, zwłaszcza technologii mikromacierzy i masowego sekwencjonowania równoległego (MPS/NGS), znacząco zwiększył naszą wiedzę na temat mechanizmów genetycznych i epigenetycznych leżących u podstaw funkcji żywych organizmów.

Do analizy metylacji szeroko wykorzystuje się platformę Illumina Infinium BeadChips. Pomiar stopnia metylacji oparty na wartości β jest trudny do analizy przy użyciu standardowych testów statystycznych. W związku z tym poszukuje się nowych metod analizy. Uczenie maszynowe oferuje szerokie możliwości w badaniu metylacji DNA, wprowadzając zaawansowane narzędzia analityczne, które mogą przełamać ograniczenia tradycyjnych metod. Przede wszystkim, algorytmy uczenia maszynowego mogą przetwarzać ogromne ilości danych metylacyjnych, identyfikując wzorce, które mogą być niewidoczne dla ludzkiego oka lub trudne do wykrycia za pomocą konwencjonalnych technik statystycznych. Dzięki zdolności do analizy złożonych, wielowymiarowych zbiorów danych, te algorytmy mogą wykrywać subtelne różnice w metylacji DNA, które mogą wskazywać na obecność chorób lub inne biologiczne procesy.

Sieci neuronowe, w przeciwieństwie do standardowych testów statystycznych, charakteryzują się większą odpornością na problemy takie jak wysoka wariancja, szum, niekompletność danych czy heteroskedastyczność. Dodatkowym atutem tej technologii jest zdolność do adaptacji, personalizacji i ulepszania w miarę uczenia się z danych. Niezależnie od wybranej metody badawczej, dane uzyskane w analizie metylacji powinny być reprezentatywne i wskazywać na rzeczywiste znaczenie biologiczne.

Rozprawa doktorska lek. Tomasza Falgowskiego pt. "Analiza metylacji sekwencji CpG w oparciu o uczenie maszynowe i sieci neuronowe" porusza tematykę kluczową dla współczesnej medycyny molekularnej. Celem pracy było zastosowanie zaawansowanych technik bioinformatycznych w analizie wzorców metylacji DNA, co ma istotne znaczenie w kontekście badań nad nowotworami i innymi chorobami genetycznymi.

Praca doktorska stanowi istotny wkład w rozwój badań nad metylacją DNA, ze szczególnym uwzględnieniem zastosowania zaawansowanych metod analitycznych w celu identyfikacji odmiennie metylowanych sekwencji CpG. Autor wykorzystał innowacyjny algorytm działający w dwóch wersjach CTMeth-hh oraz CTMeth-hhi. Wersja CTMeth-hh skupia się na wyodrębnieniu zestawów danych, gdzie wartości w jednej grupie są hipometylowane, a w drugiej hipermetylowane, lub odwrotnie. Wersja CTMeth-hhi wybiera zestawy, w których jedna z grup kontrolna lub badana zawiera wartości skrajne, a druga albo wartości nieokreślone/częściowo-metylowane, albo skrajne w przeciwnym zakresie. Metody CTMeth-hhi, która wykazuje większe znaczenie biologiczne w porównaniu do tradycyjnych metod. Metoda CTMeth-hhi wykazuje potencjał jako alternatywne i skuteczne narzędzie w analizie metylacji DNA, co może mieć szerokie zastosowanie w różnych dziedzinach badań biomedycznych. Badania zostały przeprowadzone z użyciem baz danych FUMA, DICE oraz BloodSpot, co pozwoliło na kompleksową ocenę wyników.

Praca ta nie tylko wprowadza nowe narzędzie analityczne, ale również otwiera inne możliwości w kontekście dalszych badań i modyfikacji modelu wraz z postępem wiedzy na temat epigenetyki.

Rozprawa doktorska lek. Tomasza Falgowskiego została przedstawiona w sposób kompleksowy, obejmując kluczowe sekcje wymagane w pracach naukowych. Praca składa się z następujących rozdziałów: Wstęp, Cel Pracy, Materiały i Metodyka, Wyniki, Wnioski, Dodatkowe moduły i dalszy plan rozwoju, Bibliografia, Streszczenie w języku polskim i angielskim.

Układ pracy jest nietypowy, co wpływa na jej czytelność i zrozumienie, czyniąc je mniej klarownymi. Brak wyodrębnionej sekcji "Dyskusja" może być mylący, a przemieszczenie streszczeń na początek poprawiłoby czytelność. Układ pracy, choć logiczny, po prowadzeniu pewnych zmian (numeracji rozdziałów i podrozdziałów), mógłby ułatwić odbiór i zwiększyć przejrzystość.

W poniższej ocenie przedstawiono szczegółową analizę poszczególnych sekcji pracy, ich adekwatności do tematyki oraz sposób, w jaki przyczyniają się do klarowności i spójności całego dokumentu.

1. Wstęp

Wstęp pracy doktorskiej skutecznie wprowadza czytelnika w tematykę epigenetyki, szczególnie podkreślając rolę metylacji DNA oraz jej znaczenie w biologii i medycynie. Autor jasno przedstawia wyzwania związane z analizą danych metylacyjnych, wskazując na trudności klasycznych metod statystycznych i korzyści wynikające z zastosowania technik uczenia maszynowego i sieci neuronowych. Opisane zostało również zastosowanie tej technologii w badaniach medycznych, co świadczy o dogłębnym zrozumieniu tematu przez doktoranta.

Rozwinięcie wstępu o konkretne przykłady zastosowań technik uczenia maszynowego w analizie danych metylacyjnych mogłoby dodatkowo uwypuklić praktyczne znaczenie proponowanych metod badawczych. Całość powinna być także bardziej spójna językowo, unikając zbyt technicznych zwrotów, które mogą być trudne do zrozumienia dla osób spoza dziedziny.

Uwagi:

- Styl i język: Miejscami zdania są zbyt złożone i długie, co może utrudniać płynność czytania.
- Powtórzenia: W tekście występują pewne powtórzenia, np. wyjaśnienie, że elementy cis-regulacyjne wpływają na ekspresję genów, pojawia się kilka razy.
- Brak spójności: Wstęp nie jest spójny w zakresie wprowadzenia do tematu pracy. Podrozdziały nie przechodzą płynnie, co utrudnia czytelnikowi zrozumienie głównego tematu badań. Wstęp mógłby być bardziej ukierunkowany, zawężając tematykę i jednocześnie dodając bardziej szczegółowe informacje dotyczące stricte metylacji DNA oraz zastosowania sztucznych sieci neuronowych (ANN) w tym zakresie.
- Logika i płynność: Podrozdziały powinny być logicznie ze sobą powiązane, aby płynnie wprowadzać czytelnika w kolejne zagadnienia. Każdy podrozdział powinien kończyć się krótkim podsumowaniem, które prowadzi do kolejnego zagadnienia.
- Braki: W niniejszej rozprawie brakuje szczegółowego opisu metod analizy metylacji DNA, które są istotne dla zrozumienia dynamiki metylacji genomowej. Dwa podejścia dostarczają cennych informacji na temat stanu epigenetycznego genomu, ale różnią się pod względem szczegółowości i zakresu. Analiza pojedynczych miejsc CpG pozwala na dokładne określenie stopnia metylacji poszczególnych nukleotydów. Metylacja jest tutaj wyrażana jako procent metylowanych cytozyn w danym miejscu CpG. Przykładowo, jeśli 70% cytozyn w określonym miejscu CpG jest metylowanych, mówi się o 70% metylacji w tym miejscu. Taka precyzyjna analiza jest szczególnie przydatna w identyfikacji specyficznych biomarkerów metylacyjnych związanych z różnymi stanami chorobowymi lub procesami biologicznymi. Z kolei analiza całych fragmentów DNA obejmuje ocenę średniego poziomu metylacji w większych regionach genomowych, takich jak promotory genów, regiony międzygenowe czy wyspy CpG. W tym przypadku metylacja jest również wyrażana w procentach, lecz odnosi się do średniego poziomu metylacji wszystkich miejsc CpG w danym fragmencie. Na przykład, jeśli w analizowanym fragmencie DNA średni poziom metylacji wynosi 45%, oznacza to, że 45% wszystkich miejsc CpG w tym regionie jest metylowanych. Porównanie metylacji na poziomie pojedynczych miejsc CpG i całych fragmentów DNA pozwala na bardziej kompleksowe zrozumienie dynamiki metylacji genomowej. Oba podejścia są komplementarne: analiza pojedynczych miejsc dostarcza szczegółowych informacji, podczas gdy analiza całych fragmentów pozwala na uchwycenie ogólnych trendów metylacyjnych w szerszym kontekście genomycznym.
- Przykłady i studia przypadków: Dodanie konkretnych przykładów i studiów przypadków dotyczących zastosowania uczenia maszynowego w analizie metylacji DNA mogłoby uczynić wstęp bardziej angażującym i zrozumiałym dla czytelnika.

Podsumowując:

Wstęp rozprawy doktorskiej jest dobrze napisany i merytoryczny, ale można go jeszcze ulepszyć poprzez uproszczenie zdań, redukcję powtórzeń oraz dodanie szczegółów w niektórych miejscach. Poprawa płynności przejść między podrozdziałami oraz bardziej szczegółowe skupienie się na metylacji DNA i sieci neuronowej zwiększyłoby czytelność i zrozumienie tekstu. Dodanie konkretnych przykładów zastosowań technologii omówionych we wstępie również przyczyniłoby się do lepszego zrozumienia tematu przez czytelnika.

2. Cel Pracy

Cel pracy został sformułowany w sposób jasny i zrozumiały. Wskazuje na konkretne technologie (sieci neuronowe konwolucyjne i transformers) oraz specyficzne zastosowanie -

analiza metylacji DNA, jest określony precyzyjnie, co pozwala na jednoznaczną ocenę, czy został osiągnięty. Ograniczenie celu do oceny porównawczej metody na podstawie syntetycznych danych jest konkretne i mierzalne, jak również wykonalne. Wymaga zaawansowanej znajomości technik uczenia maszynowego oraz zrozumienia procesów biologicznych związanych z metylacją DNA. Przetrenowanie modeli na syntetycznych danych jest praktycznym podejściem do oceny ich skuteczności.

Podsumowując:

Cel pracy jest dobrze sformułowany, jasny i konkretny, co ułatwia jego realizację i ocenę końcowych wyników. Tematyka jest aktualna i nowoczesna, a podejście oryginalne, co zwiększa wartość naukową pracy. Wykorzystanie zaawansowanych algorytmów AI w analizie metylacji DNA ma potencjał, aby przyczynić się do istotnych postępów w naukach biomedycznych.

3. Materiały i Metodyka

W pracy zastosowano nowoczesne metody analizy danych, w tym algorytmy uczenia maszynowego oraz sieci neuronowe. Autor szczegółowo opisał procesy przygotowania danych, wybór odpowiednich algorytmów oraz parametryzację modeli. Metodyka badań została przeprowadzona na wysokim poziomie i spełnia wszystkie wymagania współczesnych standardów naukowych.

Zbiory danych zostały dobrane w sposób umożliwiający stopniowanie trudności analizy, co pozwala na kompleksową ocenę skuteczności zaproponowanej metody w różnych kontekstach biologicznych. Jest to podejście adekwatne do oceny, czy zaproponowana architektura sieci neuronowych jest lepsza od standardowych metod. Wykorzystanie sieci konwolucyjnych i transformers jest właściwe dla analizy sekwencji DNA, które mają złożone wzorce metylacji. Kombinacja tych technik pozwala na uchwycenie zarówno lokalnych, jak i globalnych zależności w danych, co jest kluczowe dla dokładnej analizy metylacji. Wykorzystanie syntetycznych danych do przetrenowania modeli pozwoliło na uzyskanie większej liczby próbek treningowych, co jest kluczowe dla skuteczności technik uczenia maszynowego.

Podsumowując:

Materiały i metodyka zastosowane w pracy doktorskiej są adekwatne do założonego celu. Autor wybrał odpowiednie zbiory danych o różnym stopniu trudności, co umożliwia wszechstronne przetestowanie skuteczności zaproponowanej metody. Wykorzystanie zaawansowanych technik uczenia maszynowego, takich jak kombinacja sieci konwolucyjnych i transformers, jest nowatorskie i dobrze dopasowane do analizy metylacji DNA. Wykonane działania, w tym implementacja i porównanie z metodami standardowymi, zostały przeprowadzone rzetelnie i świadczą o dużym wkładzie pracy oraz zaawansowanej wiedzy technicznej autora.

4. Wyniki

Wyniki uzyskane przez lek. Tomasza Falgowskiego wnoszą nowe, istotne informacje do wiedzy na temat statystycznej analizy metylacji sekwencji CpG. Autor przedstawił wyniki w sposób klarowny, z wykorzystaniem odpowiednich narzędzi AI oraz wizualizacji danych. Analizy potwierdziły hipotezy badawcze i wskazały na potencjalne kierunki dalszych badań. Wyniki uzyskane przy użyciu kombinacji sieci konwolucyjnych i transformerów są

nowatorskie i pokazują potencjał tej technologii w analizie metylacji DNA. Lepsze wyniki w bardziej złożonych zbiorach danych sugerują, że metoda może być szczególnie użyteczna w analizie skomplikowanych wzorców metylacji, co jest ważnym wkładem w dziedzinę bioinformatyki.

Podsumowując:

Uzyskane wyniki pokazują, że zaproponowana przez autora metoda jest skuteczna w analizie metylacji DNA. Metoda jest porównywalna lub lepsza od standardowych technik, szczególnie w bardziej złożonych zbiorach danych, co podkreśla jej nowatorski charakter. Wyniki mają duże znaczenie biologiczne i praktyczne, co zwiększa wartość naukową i potencjalne zastosowania tej metody w bioinformatyce i medycynie.

○

5. Wnioski

Wnioski pracy są jasno sformułowane i wynikają bezpośrednio z przeprowadzonych badań. Autor precyzyjnie przedstawia, jakie cele zostały osiągnięte oraz jakie są najważniejsze odkrycia wynikające z przeprowadzonych badań. Każdy wniosek jest przedstawiony w sposób jednoznaczny, co ułatwia czytelnikowi zrozumienie kluczowych rezultatów pracy. Wnioski są logiczną kontynuacją wyników przedstawionych w pracy. Autor konsekwentnie odnosi się do celów, które zostały przedstawione na początku rozprawy. Wnioski kończą się podsumowaniem najważniejszych osiągnięć pracy, co pozwala czytelnikowi na szybkie zrozumienie, jakie były główne rezultaty badań i jakie mają one znaczenie dla dalszego rozwoju nauki i praktyki klinicznej.

Podsumowując:

Wnioski w rozprawie doktorskiej Tomasza Falgowskiego są jasno sformułowane, spójne z przedstawionymi wynikami oraz dyskusją. Autor podkreśla nowatorski charakter swojej metody oraz jej znaczenie biologiczne i kliniczne. Wnioski są szczegółowe, kompleksowe i poparte solidnymi dowodami. Ogólnie, prezentacja wniosków jest mocnym elementem rozprawy i dobrze podsumowuje osiągnięcia autora.

6. Dodatkowe moduły i dalszy plan rozwoju

Rozdział „Dodatkowe moduły i dalszy plan rozwoju” w pracy doktorskiej Tomasza Falgowskiego stanowi krótkie przedstawienie planów na przyszłość związane z badaniami oraz ewentualne rozwiniecie zastosowanych metod. Autor omawia potencjalne kierunki rozwoju swojej metody oraz dodatkowe moduły, które mogą być wprowadzone w przyszłości, aby ulepszyć analizę metylacji DNA.

7. Bibliografia

Bibliografia rozprawy doktorskiej Tomasza Falgowskiego obejmuje szeroki zakres literatury naukowej, od klasycznych prac po najnowsze badania w dziedzinie epigenetyki, genetyki i biologii molekularnej oraz metod analizy bioinformatycznej. Zidentyfikowano łącznie 373 pozycje literaturowe, w tym wiele publikacji o wysokim wskaźniku wpływu (IF) oraz znaczących dla współczesnych badań naukowych.

Wiele z cytowanych publikacji pochodzi z renomowanych czasopism naukowych, takich jak Clinical Epigenetics, Genome Biology, Genes, Neural Comput & Applic oraz Proc Natl Acad Sci USA. Cytowanie takich źródeł zwiększa wiarygodność i aktualność pracy.

Podsumowując:

Praca zawiera około 120 pozycji opublikowanych w ciągu ostatnich pięciu lat (2019-2024). Jeśli całkowita liczba pozycji w bibliografii wynosi 373, to około 32% wszystkich cytowanych prac pochodzi z ostatnich pięciu lat. To stanowi dobry odsetek, który wskazuje na solidną aktualność badań i literatury wykorzystanej w pracy.

Bibliografia obejmuje szeroki zakres tematów, od mechanizmów epigenetycznych po zastosowania uczenia maszynowego w bioinformatyce. Ta różnorodność wskazuje, że praca jest dobrze zakorzeniona w szerokim kontekście współczesnych badań naukowych.

Analiza bibliografii obejmuje również klasyczne prace fundamentujące wiedzę w dziedzinie genetyki, koncentrujące się na metylacji DNA i jej roli w regulacji genów, co jest jednym z elementów tematyki rozprawy doktorskiej.

III. Ocena końcowa

Autora przedstawił rozwiązanie szeregu problemów związanych z analizą metylacji DNA przy użyciu zaawansowanych technik uczenia maszynowego, w tym sieci neuronowych konwolucyjnych i typu transformers. Praca dowodzi wysokiego poziomu wiedzy Doktoranta w zakresie bioinformatyki oraz umiejętności samodzielnego prowadzenia pracy naukowej w tej dziedzinie.

Autor przeprowadził szczegółowe badania, wykorzystując trzy różne zbiory danych metylacyjnych, co pozwoliło na wszechstronną ocenę skuteczności zaproponowanej metody. Uzyskane wyniki wskazują, że metoda opracowana przez Autora jest porównywalna lub lepsza od standardowych metod analizy metylacji DNA, szczególnie w przypadku bardziej złożonych zbiorów danych. Ponadto, praca wnosi istotny wkład w rozwój metod bioinformatycznych stosowanych w analizie epigenetycznej, co może mieć szerokie zastosowanie w badaniach biomedycznych.

Po zapoznaniu się z rozprawą stwierdzam, że spełnia ona wymagania stawiane pracom doktorskim określone w art. 187 Ustawy o szkolnictwie wyższym i nauce z dnia 20 lipca 2018 (Dz.U. 2023 poz. 742) i wnoszę o dopuszczenie Pana lek. Tomasza Falgowskiego do dalszych etapów przewodu doktorskiego.

Podpis recenzenta

dr hab. n. med. Magdalena Spólnicka
08.07.2024 r.