



Silesian
University
of Technology

Silesian University of Technology

Faculty of Automatic Control, Electronics and Computer Science

Developing a system of automatic identification of cellular subpopulations in data from single-cell mass cytometry with the use of algorithms for grouping of high dimensional data

Doctoral Dissertation

by

Aleksandra Suwalska

Supervisor

Professor Joanna Polańska, PhD, DSc

2023

Gliwice, Poland

Streszczenie

Rozwój technologii badających pojedyncze komórki, jak cytometria masowa, pozwolił na dogłębne zrozumienie procesów komórkowych i odkrycie nowych subpopulacji oraz ich roli w organizmie. Wiedza ta ma szansę przyczynić się do rozwoju nowych terapii, leków i metod zapobiegania różnym chorobom. Cytometria masowa umożliwia jednoczesny pomiar dziesiątek markerów wykorzystywanych do identyfikacji typów komórek czego wynikiem jest macierz zawierająca wartości ekspresji dla każdej komórki i markera. Jednakże wysoka wymiarowość danych sprawia, że ich analiza jest wyzwaniem.

Gruźlica, czyli infekcja spowodowana przez bakterie *Mycobacterium Tuberculosis*, każdego roku zabija miliony ludzi na świecie. Leczenie jest drogie, zwłaszcza gdy bakterie uodpornią się na jeden lub więcej głównych leków. Wybuch pandemii COVID-19 doprowadził do regresji w postępach ostatnich lat w powstrzymywaniu roznoszenia się gruźlicy. Naukowcy próbują zmniejszyć występowanie nowo zdiagnozowanych przypadków do minimalnego poziomu poprzez pracę nad nowymi terapiami i lekami, wykorzystując, między innymi, cytometrię masową.

Rozprawa doktorska skupia się na analizie wysokowymiarowych danych z cytometrii masowej. Publicznie dostępne zbiory danych posłużyły do zaproponowania schematu analizy, który rozwiązuje problemy występujące przy istniejących rozwiązaniach. Zaimplementowane metody zostały wykorzystane do przetworzenia zbioru danych z badań nad gruźlicą, zapewnionego przez partnerów naukowych z Stellenbosch University, RPA. Analiza obejmuje nową, w pełni automatyczną metodę do wstępnego bramkowania danych z cytometrii masowej. Dodatkowo, istniejące rozwiązania do korekty efektu paczki zostały porównane pod względem usuwania technicznej wariacji i wpływu na końcowe wyniki identyfikacji subpopulacji. Zaproponowana technika do identyfikacji typów komórek bierze pod uwagę istnienie heterogeniczności grup komórek podczas ewaluacji modelu, co jest podejściem zupełnie nowym. Wprowadzenie rozszerzonej przestrzeni cech i dwukrokowej techniki klasteryzacji pozwala uzyskać dobrze zdefiniowane i spójne populacje.

Przeprowadzone analizy wskazują na potencjał wprowadzonych metod w identyfikacji typów komórek i odpowiedniej weryfikacji ich wyników.