



Silesian
University
of Technology

SILESIAN UNIVERSITY OF TECHNOLOGY
Faculty of Automatic Control, Electronics
and Computer Science

**Machine learning methods in support
of multiomics signature identification
for breast cancer patient
subpopulations**

Doctoral Dissertation

Joanna Tobiasz

Supervisor:
Professor Joanna Polańska, PhD, DSc

Co-supervisor:
Christos Hatzis, PhD

2023
Gliwice, POLAND

Streszczenie

Nowotwór piersi jest chorobą mocno zróżnicowaną o wysoce heterogenicznym obrazie molekularnym. Stosowana powszechnie klasyfikacja kliniczna oparta jest na poziomie kilku białek markerowych, natomiast klasyfikacja molekularna powstała na podstawie profili ekspresji genów. Oba te podziały pozostają niezmiennie od lat i nie odzwierciedlają dostatecznie złożonej struktury tej choroby oraz zróżnicowania zachowań klinicznych. Określenie podtypu raka piersi jest kluczowe przy wyborze terapii. W obliczu szybkiego rozwoju biologii molekularnej, dokładniejsze scharakteryzowanie podtypów nowotworu piersi może wesprzeć poszukiwanie nowych celów terapeutycznych. Celem tej pracy było zastosowanie metod uczenia maszynowego do identyfikacji i klinicznego oraz molekularnego scharakteryzowania podpopulacji pacjentek z rakiem piersi.

Przetestowane kombinacje metod inżynierii cech i klastrowania oraz zaproponowane sposoby ich porównania pozwoliły na pogrupowanie pacjentek w oparciu o profil białkowy. Zidentyfikowano sześć podpopulacji pacjentek, które oceniono pod kątem demograficznym, klinicznym i molekularnym na podstawie profili białkowych i transkryptomocnych. Dobrane metody klasycznej analizy statystycznej wspartej miarą wielkości efektu i uczeniem maszynowym pozwoliły zmierzyć się z problemem różnych, czasem niedostatecznych, rozmiarów grup.

Trzy z uzyskanych na podstawie profilu białkowego podpopulacji wykazały dużą zgodność ze stosowanymi podtypami opartymi o poziom mRNA: podstawnym, HER2-wzbogaconym oraz luminalnym B. Transkryptomocny podtyp luminalny A okazał się wysoce zróżnicowany i został podzielony w tej pracy na trzy podgrupy. Otrzymane podpopulacje różnią się przeżywalnością oraz profilami białkowymi i transkryptomocnymi. Uzyskano zestawy markerów specyficznych dla podpopulacji oraz sygnaturę pozwalającą na rozróżnienie wszystkich podtypów. Nowe podtypy luminalne są mniej zróżnicowane w przestrzeni transkryptomocnej niż w białkowej.

Profile wykrytych podpopulacji, zwłaszcza białkowy, mogą uzupełnić stosowane klasyfikacje raka piersi i wesprzeć poszukiwania nowych terapii celowanych w rozwoju medycyny spersonalizowanej. Niemniej jednak, do dalszej oceny przydatności klinicznej otrzymanych wyników niezbędna jest ich niezależna walidacja.