

Проект по Биоинформатика

Тема: Предвидување на молекуларен подтип на рак на дојка со машинско учење врз основа на податоци од генска експресија.

Опис на проектот:

Проектот има за цел развој на модел кој автоматски ќе го класифицира молекуларниот подтип на рак на дојка (Luminal A, Luminal B, HER2-enriched и Basal-like) користејќи податоци од генска експресија добиени со RNA-Seq или microarray технологии. Ракот на дојка претставува хетерогена болест со различни биолошки карактеристики и клинички исходи, па точната идентификација на подтипот е клучна за избор на соодветна терапија и проценка на дијагнозата.

Во проектот ќе се користат јавно достапни геномски бази на податоци (на пр. TCGA или METABRIC), од кои ќе се извлечат експресионите профили на илјадници гени по пациент. Ќе се применат техники за претпроцесирање на податоците (нормализација, редукција на димензионалност, селекција на карактеристики), по што ќе се обучат и евалуираат повеќе модели на машинско учење (SVM, Random Forest, XGBoost, Neural Networks).

Целта е да се постигне висока точност на класификација и да се идентификуваат најрелевантните гени кои придонесуваат за разликување на подтиповите. Ваквиот пристап има значење за персонализирана медицина и подобрување на терапевтските одлуки.