

Hierarchické klastrovacie algoritmy sú bežnými nástrojmi na zjednodušenie, skúmanie a analýzu datasetov v mnohých oblastiach výskumu. Pre prietokovú cytometriu bol navrhnutý špecifický variant aglomeračného zhlukovania, ktorý využíva klastrovanie založené na Mahalanobisovskej vzdialenosti, aby sa dosiahli výsledky, ktoré sa pre danú doménu lepšie hodia. Uplatniteľnosť tohto klastrovacieho algoritmu je v súčasnosti obmedzená jeho pomerne vysokou výpočtovou náročnosťou, ktorá mu neumožňuje škálovať bežné cytometrické datasety. Táto práca popisuje špecializovanú, na GPU-zrýchlenú verziu hierarchického klastrovania založeného na Mahalanobisovskej vzdialenosti, ktorá zlepšuje výkon algoritmu o niekoľko rádov, čo mu umožňuje počítať oveľa väčšie datasety. Práca poskytuje prehľad súčasných hierarchických klastrovacích algoritmov a podrobne popisuje implementáciu algoritmu na GPU. Výsledok sa porovnáva s verejne dostupnými vysokorozmernými údajmi z hmotnostnej cytometrie.