

Количественный анализ деления клеток человека с применением алгоритмов машинного обучения

Научный руководитель – Фахруллин Равиль Фаридович

Ишмухаметов Ильнур Ринатович

Аспирант

Казанский (Приволжский) федеральный университет, Институт фундаментальной
медицины и биологии, Казань, Россия

E-mail: irishmukhametov@gmail.com

Поиск новых путей применения наноматериалов в биомедицине требует тщательного анализа их свойств на клеточном уровне [1]. Техника корреляционной микроскопии, при которой исследуемый объект может быть визуализирован с применением нескольких типов микроскопии одновременно позволяет быстро получить информацию о пространственно-спектральных параметрах образцов. Однако ключевым этапом является подготовка и анализ больших массивов данных. Применение в подобных задачах алгоритмов машинного обучения может не только значительно сократить затрачиваемое на обработку время, но и способствует выявлению неочевидных исследователю закономерностей [2].

Целью данной работы является автоматизированный анализ митотического индекса клеток аденокарциномы лёгких человека A549 с применением алгоритмов машинного обучения. Для этого F-актин и ядра клеток, зафиксированных спустя 24 часа культивации, были окрашены ФИТЦ-конъюгированным фаллоидином и ДАФИ соответственно. Анализ образцов производился с применением системы высококонтрастной микроскопии темного поля в сочетании с гиперспектральной визуализацией CytoViva. Снимки цитоскелета и ядер были получены с использованием модуля двухрежимной флуоресценции. Сегментация единичных клеток на снимках была выполнена посредством программного обеспечения CellProfiler 4.0.0. Конфигурационный файл, содержащий информацию о параметрах клеток (площадь, интенсивность флуоресценции) был загружен в CellProfiler Analyst 2.2.1 для бинарной классификации с применением алгоритмов машинного обучения Random Forest и Fast Gentle Boosting. В качестве тренировочного набора данных вручную были отобраны 20 клеток, находящихся в интерфазе (отрицательные) и 10 клеток в состоянии деления (положительные). Точность классификации относительно ручного анализа клеток оценивалась посредством матрицы ошибок и подсчета метрик точности (полнота, аккуратность, F-мера).

В результате исследования был сформирован набор данных, состоящий из 28 групп снимков в трех каналах: темное поле и два канала флуоресценции. Общее количество клеток, подсчитанное вручную составило 528, из которых 63 находились в состоянии деления (митотический индекс - 12%). Автоматизированный процесс сегментации позволил выявить 99% клеток на снимках. Наиболее точные предсказания классов были получены с использованием алгоритма Fast Gentle Boosting, при котором общая точность была равна 95%, а значение F-меры составило 0,75.

Исследование выполнено за счет гранта Российского научного фонда (проект № 20-13-00247).

Источники и литература

- 1) Gautam A., van Veggel F.C. Synthesis of nanoparticles, their biocompatibility, and toxicity behavior for biomedical applications // Journal of Materials Chemistry B. 2016. Vol. 1. N. 39. P. 5186-5200.

- 2) Orth A., Schaak D., Schonbrun E. Microscopy, meet big data // Cell systems. 2017. Vol. 4. N. 3. P. 260-261.