

В.Г. НИКИТАЕВ, Т.Н. ЗАБОТИНА, А.Н. ПРОНИЧЕВ, А.А. БОРУНОВА, М.С. ТОРОСЯН

*Национальный исследовательский ядерный университет «МИФИ», Москва; Российский онкологический научный центр им. Н.Н. Блохина, Москва, Россия*

*Российский онкологический научный центр им. Н.Н. Блохина Министерства здравоохранения Российской Федерации, Москва, Россия*

## МЕТОД РАСПОЗНАВАНИЯ ПОПУЛЯЦИЙ ЛЕЙКОЦИТОВ В ДАННЫХ ПРОТОЧНОЙ ЦИТОФЛУОРИМЕТРИИ

Представлен разработанный метод преобразования данных и распознавания популяций лейкоцитов при проточной цитометрии. Ядром метода являются математические методы и модели машинного обучения, реализующие преобразование табличных данных в изображения, аугментацию и распознавание популяций лейкоцитов. Эксперимент продемонстрировал точность распознавания популяций лейкоцитов, на примере моноцитов, равную 89 %.

*V.G. NIKITAEV, T.N. ZABOTINA, A.N. PRONICHEV, A.A. BORUNOVA, M.S. TOROSYAN*

*National Research Nuclear University MEPH, Moscow; Blokhin Russian Cancer Research Center, Moscow, Russia*

*Blokhin Russian Cancer Research Center of the Ministry of Health of the Russian Federation, Moscow, Russia.*

## METHOD FOR RECOGNIZING LEUKOCYTE POPULATIONS IN FLOW CYTOFLUOROMETRY DATA

The developed method of data transformation and recognition of leukocyte populations during flow cytometry is presented. The core of the method is mathematical methods and models for converting tabular data into images, augmentation and recognition of leukocyte populations. The experiment demonstrated 89% accuracy of recognition of leukocyte populations, using monocytes as an example.

В клинической практике данные проточного цитометра содержат координаты десятков и даже сотен тысяч клеток по нескольким маркерам. На сегодняшний день, в России, анализ столь объемных данных осуществляется традиционным методом, основанный на последовательном ручном гейтировании, используя точечные графики [1]. Данный процесс анализа является трудозатратным для врачей лабораторной диагностики. Целью работы была разработка программного обеспечения, которая позволит автоматизировать поиск и распознавание клеточных популяций лейкоцитов на основе математических методов и моделей. Данные для эксперимента были взяты из открытого источника в интернете [2].

Предлагаемый метод распознавания популяций лейкоцитов основан на следующих этапах:

1. Математическое преобразование и визуализация данных. Исходные данные нормализуются методом преобразования  $x_{norm}$ , после чего координаты маркеров визуализируются в виде плотностных точечных графиков (dot plot) размером  $1000 * 1000$  пикселей;

2. Аугментация и разметка данных. Для увеличения и вариативности обучающей выборки применяется алгоритм случайного смещение координат точек  $x_{new}$  и  $y_{new}$  [3]. Полученные изображения вручную размечаются экспертом;

3. Этап распознавания основан на применении моделей глубокого обучения и алгоритмов машинного обучения. Для автоматического распознавания используется комплекс алгоритмов из программных библиотек TensorFlow и sklearn.cluster, которые реализуют классификацию изображений и обучение нейросетей.

В результате разработки метода, основанного на математических моделях и алгоритмах машинного обучения, на его основе создано программное обеспечение, реализованное на языке Python с использованием библиотек Cytoflow, Scanpy, TensorFlow и scikit-learn. Данное программное обеспечение позволяет проводить автоматическую нормализацию больших объемов данных, визуализировать их с помощью многомерных точечных графиков и выполнять автоматический поиск и классификацию популяций лейкоцитов. Эксперимент подтвердил эффективность рассмотренного метода: точность распознавания популяции лейкоцитов, на примере моноцитов, составила 89 %, а также скорость анализа, поиска и выделения популяций лейкоцитов возросла в 10 раз.

### Список литературы

1. Заботина Т.Н., Черткова А.И., Борунова А.А., Захарова Е.Н., Шоуа Э.К., Артамонова Е.В., Коваленко Е.И., Хорошилов М.В., Кадагидзе З.Г. Взаимосвязь субпопуляций лимфоцитов больных раком молочной железы с результатами лечения. // Российский биотерапевтический журнал. 2021. Т. 20. № 3. С. 25-33;

2. Flow Cytometry datasets consisting of peripheral blood and bone marrow samples for the evaluation of explainable artificial intelligence methods. Mendeley Data. [Электронный ресурс]. – Режим доступа: <https://data.mendeley.com/datasets/jk4dt6wprv/1> (дата обращения 25.12.2025);

3. Никитаев В.Г., Тупицын Н.Н., Проничев А.Н. и др. Технологии искусственного интеллекта в диагностике острых лимфобластных лейкозов и минимальной остаточной болезни. // Медицинская техника. 2020. № 5. С. 42-44.