

# Применение нейронных сетей для классификации клеток крови при патологии

А.А. Трубников  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Самара, Россия  
anteii3mer@gmail.com

Д.А. Савельев  
Самарский национальный  
исследовательский университет им.  
академика С.П. Королева  
Самара, Россия  
dmitrey.savelyev@yandex.ru

**Аннотация**— В данной работе представлен подход к классификации цифровых изображений клеток крови при остром лимфобластном лейкозе с использованием сверточных нейронных сетей. Также проведен сравнительный анализ различных архитектур сверточных сетей и рассмотрены особенности работы с клеточными изображениями. Исследование проводилось с использованием языка программирования Python и библиотеки tensorflow.

**Ключевые слова**— глубокое обучение, острый лимфобластный лейкоз, сверточные нейронные сети

## 1. ВВЕДЕНИЕ

Острый лимфобластный лейкоз (ОЛЛ) – это злокачественное заболевание кроветворной системы, выражающееся в чрезмерной выработке лимфобластов (незрелых лимфоцитов), входит в обширную группу лейкозов [1]. Является быстро прогрессирующим заболеванием, при отсутствии соответствующего и своевременного лечения может привести к летальному исходу в течение нескольких месяцев. Около 85% случаев регистрируется среди детей и подростков младше 15 лет [2].

Отличительная особенность ОЛЛ – это сильное морфологическое сходство между нормальными лимфоидными клетками и аномалиями [3]. Ранние методы диагностики включали в себя морфологическое исследование мазков периферической крови, однако сейчас, основными инструментами диагностики стали: цитогенетическое исследование, биопсия костного мозга и люмбальная пункция [2]. Для проведения таких исследований требуются врачи и лаборанты с высокой квалификацией, специальные условия и оборудование. Поэтому такие исследования обычно доступны не всем.

Использование нейронных сетей делает возможным построение автоматических систем классификации цифровых изображений клеток крови, без участия экспертов. Возможности применения машинного обучения широко исследуются в смежных сферах и при решении конкретно этой задачи [4, 5].

Цель работы – рассмотреть особенности анализа клеток крови и провести сравнительный анализ различных архитектур нейронных сетей при классификации изображений клеток крови. Набор данных был взят из открытого источника [6] и состоит из 10581 фотографий лимфоидных клеток крови, среди которых представлены два класса: нормальная клетка и аномальная (лимфобласт). Файлы имеют расширение .bmp и разрешение 600x600.

Научная новизна данного исследования заключается в том, что при проведении работы были учтены

особенности клеточных изображений и был проведен анализ результатов работы обученных моделей, чего не проводилось в аналогичных работах [7].

## 2. КЛАССИФИКАЦИЯ ЦИФРОВЫХ КЛЕТОЧНЫХ ИЗОБРАЖЕНИЙ

Данная работа разделена на три этапа: предварительный анализ и аугментация данных, обучение нейронной сети, сравнение результатов и их анализ.

При анализе микроскопических фотографий клеток стоит учитывать следующие особенности:

- 1) Небольшое количество данных.
- 2) Несбалансированность выборки.
- 3) Специфичные ошибки, вносимые освещением.
- 4) Предварительная сегментация.
- 5) Требуется высокая квалификация для конструирования ручных признаков.
- 6) Высокая внутриклассовая вариативность и вариативность среди субъектов исследования.

В частности, из приведенных выше особенностей видно преимущество нейронных сетей, по сравнению с классическими алгоритмами машинного обучения, требующие ручного выделения признаков.

Специфичные ошибки могут быть внесены освещением в лаборатории, оптической системой микроскопа и камеры или при окраске препарата. В анализируемых данных ошибки освещения и окраски были минимизированы путем применения специального метода нормализации цветов окраски.

Специфику данных (инвариантность класса к вращению изображения и внесению небольших шумов) можно использовать для их аугментации (поворота на случайный угол, внесение случайного шума). Такой подход позволит частично преодолеть изначальное ограничение, связанное с доступностью данных.

Важно также учитывать, что свойства клетки, в частности морфологические, могут сильно варьироваться среди субъектов (пациентов) из-за чего важно правильно разделять исходную выборку на обучающее и тестовое подмножества.

В эксперименте участвовали несколько архитектур нейронных сетей, которые были предварительно обучены на датасете ImageNet. Также с нуля была обучена небольшая сеть SimpleNet (представленная простой последовательностью из нескольких слоев). В ходе эксперимента модели обучались в течение 200

эпох. Результаты работы обученных моделей на тестовой выборке представлены в таблице I.

Таблица I. РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

		Показатели эффективности				
		Время, мс	Precision	Recall	F1	$\Delta F1$ , %
Алгоритмы	<i>Xception</i>	410	0,81	0,88	0,79	1,2
	<i>VGG16</i>	564	0,84	0,86	0,80	2,5
	<i>ResNet15 2V2</i>	549	0,84	0,88	0,81	3,8
	<i>EfficientNet B5</i>	790	0,84	0,92	0,83	5,1
	<i>EfficientNet V2M</i>	654	0,86	0,91	0,85	9
	<i>SimpleNet</i>	438	0,69	0,98	0,62	-19,2

Лучшие результаты показала модель архитектуры EfficientNetV2M, при этом время её работы мало отличается от других моделей (больше среднего на 86,5 мс). Архитектура SimpleNet показала худшие результаты, чем продемонстрировала невозможность применения небольших и примитивных архитектур на рассматриваемом наборе данных.

На Рис. 1 также представлены карты значимости для двух образцов: нормальной клетки и аномалии. Можно заметить, что внимание модели распределено по площади клетки неравномерно (точки за пределами границы допустимы, т.к. использованный метод построения карт значимости допускает возникновение неточностей [8]). Это значит, что на принятие решения моделью некоторые участки клетки оказали большее влияние, соответственно информация, представленная на этих участках изображения может оказаться полезной при построении системы поддержки принятия решений

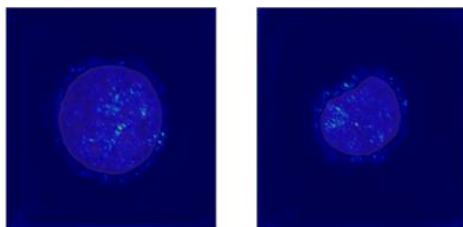


Рис. 1. Карты значимости построенные для EfficientNetV2M

### 3. ЗАКЛЮЧЕНИЕ

Появление моделей, обеспечивающих приемлемый уровень точности, даст возможность построения большого количества инновационных систем: диагностических, систем принятия решений или

поддержки принятия решений. Кроме того, станет возможна централизованная обработка результатов медицинской диагностики, не привязанная к месту сбора анализов.

В ходе проведенной работы были выделены особенности анализа клеточных изображений, проведена аугментация данных и их разбиение на обучающую и тестовую выборки с учетом распределения субъектов, проведены эксперименты и интерпретация результатов работы обученных моделей. Было установлено, что лучшие результаты на исследуемом наборе данных показала архитектура EfficientNetV2M (значение целевой метрики F1 на 9% больше среднего среди рассматриваемых архитектур, при этом, разница между худшим и лучшим значением метрики F1 составляет 0,23).

### ЛИТЕРАТУРА

- [1] Thomas, X. Acute promyelocytic leukemia: a history over 60 years - from the most malignant to the most curable form of acute leukemia / X. Thomas // *Oncology and therapy*. – 2019. – Vol. 7(1). – P. 33-65. DOI: 10.1007/s40487-018-0091-5
- [2] Overview Acute lymphoblastic leukaemia [Electronic resource]. — Access mode: <https://www.nhs.uk/conditions/acute-lymphoblastic-leukaemia> (06.10.2022)
- [3] Ladines-Castro, W. Morphology of leukaemias / W. Ladines-Castro, G. Barragán-Ibañez, M.A. Luna-Pérez, A. Santoyo-Sánchez, J. Collazo-Jaloma, E. Mendoza-García, C.O. Ramos-Peñañiel // *Revista Médica del Hospital General de México*. – 2016. – Vol. 79(2). – P. 107-113. DOI: 10.1016/j.hgmx.2015.06.007
- [4] Gu, Y. Deep Learning based Cell Classification in Imaging FlowCytometer / Y. Gu, A. Chen, X. Zhang, C. Fan // *Asp Transactions on Pattern Recognition and Intelligent Systems*. – 2021. – Vol. 1(2). – P. 18-27. DOI: 10.52810/TPRIS.2021.100050
- [5] Shafique, S. Acute lymphoblastic leukemia detection and classification of its subtypes using pretrained deep convolutional neural networks // S. Shafique, S. Tehsin // *Technology in cancer research & treatment*. – 2018. – Vol. 17. – P. 1-7. DOI: 10.1177/1533033818802789
- [6] Leukemia Classification [Electronic resource]. – Access mode: <https://www.kaggle.com/datasets/andrewmvd/leukemia-classification> (04.09.2022)
- [7] Pansombut, T. Convolutional Neural Networks for Recognition of Lymphoblast Cell Images / T. Pansombut, S. Wikaisuksakul, K. Khongkraphan, P. Aniruth // *Computational Intelligence and Neuroscience*. – 2019. – Vol. 20. – P. 1-12. DOI: 10.1155/2019/7519603
- [8] Chenlei, G. A Novel Multiresolution Spatiotemporal Saliency Detection Model and Its Applications in Image and Video Compression / G. Chenlei, L. Zhang // *IEEE Transactions on Image Processing*. – 2010. – Vol. 19(1). – P. 185-198. DOI: 10.1109/TIP.2009.2030969