

Шеретов М.А., Яруллин Д.В.

СОВРЕМЕННЫЕ МЕТОДЫ АНАЛИЗА И КЛАССИФИКАЦИИ КЛЕТОК КРОВИ: НЕОБХОДИМОСТЬ ИНТЕРПРЕТИРУЕМЫХ АЛГОРИТМОВ

В данной статье рассмотрены различные модели для анализа, подсчёта и классификации клеток крови. Традиционные методы, такие как метод опорных векторов и преобразование Хафа, показали свою эффективность, но требуют дорогостоящего оборудования и сложных процедур преданализа, что затрудняет их применение в современных клиниках. Современные подходы с использованием свёрточных нейронных сетей предлагают автоматизацию и снижение требований к преданализу за счёт автоматического выделения признаков. Однако основной недостаток свёрточных нейронных сетей заключается в сложности интерпретации результатов. Поднимается вопрос о необходимости разработки инструмента для интерпретации результатов работы свёрточных нейронных сетей, выявления причинно-следственной связи.

Ключевые слова: классификация клеток крови, метод опорных векторов, преобразование Хафа, свёрточные нейронные сети, преданализ.

SHERETOV M.A., Yarullin D.V.

MODERN METHODS OF BLOOD CELL ANALYSIS AND CLASSIFICATION: THE NEED FOR INTERPRETABLE ALGORITHMS

This article reviews various models for analyzing, counting and classifying blood cells. Traditional methods, such as support vector methods and Hough transforms, have been shown to be effective but require expensive equipment and complex pre-analytic procedures, making them difficult to implement in modern clinics. Current approaches using convolutional neural networks offer automation and reduced pre-analytics requirements through automatic feature extraction. However, the main drawback of convolutional neural networks is the difficulty in interpreting the results. The question of the need to develop a tool for interpreting the results of convolutional neural networks and identifying causality is raised.

Keywords: Blood cell classification, support vector method, Hough transform, convolutional neural networks, pre-analytics.

Существует множество моделей для анализа, подсчёта и классификации клеток крови, каждая из которых имеет свои особенности и области применения. Например, Lou и коллеги применяли метод опорных векторов на основе данных,

полученных с помощью спектрального анализа, для точного подсчёта эритроцитов [1]. Спектральный анализ позволяет получить высокоточные данные, однако этот метод требует дорогостоящего оборудования, недоступного большинству современных клиник. Таким образом, его использование ограничено специализированными лабораториями и крупными медицинскими центрами. В качестве альтернативы, Cruz и соавторы разработали систему, основанную на пороговом значении насыщенности оттенка и связной маркировке, для независимого подсчёта эритроцитов, лейкоцитов и тромбоцитов [2]. Этот подход менее затратен и может быть реализован с использованием стандартного лабораторного оборудования, что делает его более доступным для широкого круга медицинских учреждений. Кроме того, система Cruz и соавторов позволяет автоматизировать процесс подсчёта клеток, снижая вероятность ошибок, связанных с человеческим фактором, и повышая общую эффективность работы лабораторий. Также Acharya и Kumar разработали систему подсчёта эритроцитов, используя метод разделения вокруг медоида и преобразование Хафа [3]. Этот метод сочетает в себе кластеризацию данных и геометрический подход, что позволяет точно идентифицировать и подсчитывать эритроциты в образцах крови. Разделение вокруг медоида помогает сгруппировать похожие клетки, а преобразование Хафа используется для обнаружения и анализа круглых объектов, таких как эритроциты. Singhal и Singh занимались классификацией злокачественных и доброкачественных лейкоцитов, используя 256 локальных бинарных шаблонов [4]. Для улучшения результатов классификации они увеличили количество используемых шаблонов до 4096 [5]. Этот метод, основанный на локальных бинарных шаблонах (LBP), позволяет эффективно анализировать текстуры изображений клеток, что играет ключевую роль в распознавании и дифференциации типов лейкоцитов. Увеличение количества шаблонов с 256 до 4096 позволило значительно повысить точность классификации за счет более детального анализа структурных особенностей клеток. И наоборот, Bhattacharjee и Saini применили метод k -ближайших соседей (k -NN) для классификации лейкоцитов, используя всего 8 локальных бинарных

шаблонов [6]. Этот минималистичный подход демонстрирует, что даже с ограниченным числом шаблонов можно достичь приемлемой точности классификации, благодаря эффективности алгоритма k-NN в анализе близости данных. Метод k-NN работает на основе измерения сходства между образцами, что позволяет ему успешно классифицировать клетки, несмотря на небольшое количество используемых шаблонов. Такой подход может быть полезен в ситуациях, когда ресурсы ограничены или требуется быстрая и простая реализация. Помимо этого, использование малых наборов шаблонов снижает вычислительную нагрузку, что делает метод более доступным для широкого применения, включая реализацию в менее мощных компьютерных системах, таких как мобильные устройства или портативные анализаторы.

Учёные из Университета Джорджа Мейсона предложили использовать метод опорных векторов (SVM) для классификации клеток крови, опираясь на легко интерпретируемые признаки, извлечённые с помощью алгоритмов компьютерного зрения [7]. Этот подход позволяет эффективно анализировать изображения клеток, выделяя ключевые характеристики, которые существенно облегчают процесс классификации. Использование SVM в сочетании с алгоритмами компьютерного зрения обеспечивает высокую точность и надёжность, поскольку метод опорных векторов хорошо работает с высокоразмерными данными и помогает находить оптимальные гиперплоскости для разделения классов. Алгоритмы компьютерного зрения, в свою очередь, позволяют автоматически выделять значимые признаки из изображений, что минимизирует необходимость в ручной обработке данных и снижает вероятность ошибок. Эти признаки могут включать такие параметры, как форма, размер, текстура и интенсивность пикселей, что даёт возможность более детально анализировать клетки и улучшать точность диагностики. Более того, применение таких методов в медицинской практике способствует разработке автоматизированных систем, которые могут значительно ускорить и упростить процесс диагностики, делая его более доступным и надёжным.

Но каждый из представленных учеными методов очень требователен к набору данных, на котором он производится, увеличивая требования к ресурсам для процедуры преданализики — сбору и окраске исходного материала, что делает их трудно воспроизводимыми. При нарушении чётких правил пробоподготовки, результат сильно снижается. Однако каждый из представленных учёными методов весьма требователен к набору данных, на котором он производится, что увеличивает требования к ресурсам для процедуры преданализики, включая сбор и окраску исходного материала. Это делает их трудновоспроизводимыми и может значительно усложнить процесс подготовки образцов. Если чёткие правила пробоподготовки нарушаются, точность и надёжность результатов существенно снижаются. Процедура преданализики является критически важной, так как малейшие отклонения в сборе, обработке или окраске образцов могут привести к искажению данных, что негативно скажется на результатах анализа. Например, неправильная окраска клеток или использование некачественных реагентов может затруднить выделение значимых признаков и ухудшить качество классификации.

Одним из популярных методов анализа и классификации изображений, который способен снизить требования к этапу преданализики, является использование сверточных нейронных сетей (CNN) [8]. Этот тип модели вычислений широко применяется в анализе медицинских изображений [9,10]. Сверточные нейронные сети способны автоматически выделять и определять важные признаки изображений, не требуя ручного определения характеристик [8]. Kutlu и его коллеги, а также Long' исследовали применение различных архитектур сверточных нейронных сетей для классификации различных типов лейкоцитов [11,12]. Существует множество различных архитектур для CNN, включая VGGNet [13], ResNet [14] и YOLO [15]. Трансферное обучение особенно полезно в случаях, когда у исследователя есть ограниченный набор данных или недостаточно обучающих примеров. При таком подходе предварительно обученная сверточная нейронная сеть, сначала обученная на большом наборе данных, доучивается на задаче, которую исследует аналитик [8]. Это позволяет

эффективно использовать знания, полученные из крупных наборов данных, для решения конкретной задачи классификации клеток крови, даже при ограниченном доступе к обучающим данным.

Один из основных недостатков использования сверточных нейронных сетей заключается в сложности интерпретации признаков, извлекаемых этими моделями, независимо от их архитектуры. Некоторые исследователи стремятся снизить количество признаков, используемых для классификации, и добиваются успеха в этом. Например, Sahlol и его коллеги применили метод трансферного обучения с моделью VGGNet в сочетании с улучшенным статистическим алгоритмом "Salp Swarm Algorithm" (SESSA) для извлечения более информативных признаков при классификации здоровых и злокачественных лейкоцитов [10]. В результате отбора было выбрано более 1000 признаков из 25088 возможных, которые затем использовались в модели классификации на основе метода опорных векторов [10]. Однако этот подход по-прежнему не обеспечивает достаточного представления о ключевых признаках, определяющих доброкачественные и злокачественные лейкоциты. Кроме того, данная система в основном ориентирована на классификацию лейкоцитов и не способна различить их от эритроцитов или тромбоцитов. Это указывает на необходимость дальнейших исследований для разработки более эффективных методов анализа и классификации клеток крови, способных учитывать различные типы клеток и их характеристики.

В заключение можно отметить, что анализ, подсчет и классификация клеток крови остаются сложными задачами, требующими высокоточных и эффективных методов. Традиционные подходы, такие как метод опорных векторов, пороговое значение насыщенности оттенка и преобразование Хафа, хоть и эффективны, но часто требуют дорогостоящего оборудования и сложных процедур преданализики. Современные методы, такие как сверточные нейронные сети, предлагают более гибкие и автоматизированные решения, способные снизить требования к этапу преданализики за счет автоматического выделения признаков. Однако основным недостатком использования сверточных нейронных сетей является сложность интерпретации результатов их работы. Несмотря на достижения в области трансферного обучения и использования различных архитектур сверточных нейронных сетей, таких как VGGNet, ResNet

и YOLO, остаются проблемы с выделением и интерпретацией ключевых признаков, особенно при классификации различных типов клеток крови.

Необходимо разработать алгоритмы, которые бы позволили интерпретировать результаты работы сверточных нейронных сетей, обеспечивая более четкое понимание важных признаков, используемых для классификации клеток. Такие алгоритмы способствовали бы дальнейшим исследованиям в области классификации клеток в пробах жидкостей человека.

В целом, дальнейшие исследования и разработки необходимы для создания более точных, доступных и интерпретируемых методов анализа клеток жидкостей человека. Это позволит улучшить диагностику заболеваний и разработать более эффективные методы лечения.

Библиографический список

1. J. Lou, M. Zhou, Q. Li, C. Yuan and H. Liu, "An automatic red blood cell counting method based on spectral images," 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 2016, pp. 1391-1396, doi: 10.1109/CISP-BMEI.2016.7852934.
2. D. Cruz *et al.*, "Determination of blood components (WBCs, RBCs, and Platelets) count in microscopic images using image processing and analysis," 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, Philippines, 2017, pp. 1-7, doi: 10.1109/HNICEM.2017.8269515.
3. Acharya, V., Kumar, P. Identification and red blood cell automated counting from blood smear images using computer-aided system. *Med Biol Eng Comput* 56, 483–489 (2018). <https://doi.org/10.1007/s11517-017-1708-9>
4. V. Singhal and P. Singh, "Local Binary Pattern for automatic detection of Acute Lymphoblastic Leukemia," 2014 Twentieth National Conference on Communications (NCC), Kanpur, India, 2014, pp. 1-5, doi: 10.1109/NCC.2014.6811261.
5. Singhal, V., Singh, P. (2016). Texture Features for the Detection of Acute Lymphoblastic Leukemia. In: Satapathy, S., Joshi, A., Modi, N., Pathak, N. (eds) Proceedings of International Conference on ICT for Sustainable Development.

Advances in Intelligent Systems and Computing, vol 409. Springer, Singapore.
https://doi.org/10.1007/978-981-10-0135-2_52.

6. R. Bhattacharjee and L. M. Saini, "Robust technique for the detection of Acute Lymphoblastic Leukemia," 2015 IEEE Power, Communication and Information Technology Conference (PCITC), Bhubaneswar, India, 2015, pp. 657-662, doi: 10.1109/PCITC.2015.7438079.

7. William Franz Lamberti, Blood cell classification using interpretable shape features: A Comparative study of SVM models and CNN-Based approaches, Computer Methods and Programs in Biomedicine Update, Volume 1, 2021, 100023, ISSN 2666-9900, <https://doi.org/10.1016/j.cmpbup.2021.100023>.

8. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, Tsuhan Chen, Recent advances in convolutional neural networks, Pattern Recognition, Volume 77, 2018, Pages 354-377, ISSN 0031-3203, <https://doi.org/10.1016/j.patcog.2017.10.013>.
(<https://www.sciencedirect.com/science/article/pii/S0031320317304120>)

9. Esteva, A., Chou, K., Yeung, S. *et al.* Deep learning-enabled medical computer vision. *npj Digit. Med.* 4, 5 (2021). <https://doi.org/10.1038/s41746-020-00376-2>.

10. Sahlol, A.T., Kollmannsberger, P. & Ewees, A.A. Efficient Classification of White Blood Cell Leukemia with Improved Swarm Optimization of Deep Features. *Sci Rep* 10, 2536 (2020). <https://doi.org/10.1038/s41598-020-59215-9>.

11. Hüseyin Kutlu, Engin Avci, Fatih Özyurt, White blood cells detection and classification based on regional convolutional neural networks, Medical Hypotheses, Volume 135, 2020, 109472, ISSN 0306-9877, <https://doi.org/10.1016/j.mehy.2019.109472>.

12. Fei Long, Jing-Jie Peng, Weitao Song, Xiaobo Xia, Jun Sang, BloodCaps: A capsule network based model for the multiclassification of human peripheral blood cells, Computer Methods and Programs in Biomedicine, Volume 202, 2021, 105972, ISSN 0169-2607, <https://doi.org/10.1016/j.cmpb.2021.105972>.

13. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556 [cs] (2015).ArXiv: 1409.1556, <http://arxiv.org/abs/1409.1556>.
14. K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.
15. M.M. Alam, M.T. Islam, Machine learning approach of automatic identification and counting of blood cells, Healthc Technol Lett 6 (4) (2019) 103–108, <https://doi.org/10.1049/htl.2018.5098>.<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6718065/>

Сведения об авторах

Шеретов Марк Алексеевич – магистрант кафедры ИТАС, Пермского национального исследовательского политехнического университета, группы АСУ8-23-1м, г. Пермь, e-mail: mark.sheretov@gmail.com

Яруллин Денис Владимирович — кандидат технических наук, доцент кафедры «Информационные технологии и автоматизированные системы», Пермский национальный исследовательский политехнический университет, г. Пермь, e-mail: d.v.yarullin@ya.ru