

Ж.В. Штадельманн, И.Н. Спиридонов

МЕТОД АВТОМАТИЗИРОВАННОГО ОБНАРУЖЕНИЯ ЛЕЙКОЦИТОВ НА ИЗОБРАЖЕНИЯХ МАЗКОВ КРОВИ НА ОСНОВЕ БУСТИНГА

Аннотация

В статье рассмотрены морфологические параметры клеток белой крови, этапы автоматизированного анализа лейкоцитов в мазках крови. Реализован алгоритм обнаружения лейкоцитов на изображениях. Результаты экспериментальных исследований показали его пригодность к практическому использованию в системах автоматизированной микроскопии.

Клинический анализ крови, включающий в себя определение лейкоцитарной формулы, широко используется в диагностике большинства заболеваний [1], [2]. Дифференциация лейкоцитов на все типы, встречающиеся в периферической крови в норме и при патологии, возможна только по мазкам крови, анализ которых в настоящее время выполняется визуально врачом-гематологом. Число лейкоцитов, учитываемых при визуальном анализе (до 200 клеток) [3], не обеспечивает статистической достоверности результатов, а его увеличение приводит к существенному росту трудоемкости визуального анализа.

Методы проточной цитометрии позволяют автоматизировать анализ клеток крови за счет оценки их оптических и электрических характеристик. Однако аппараты проточной цитометрии не могут распознать все возможные морфологические типы клеток, поэтому данные приборы используются в первую очередь для выявления отклонений параметров крови от нормы, что требует визуального анализа мазков и дифференцирования возникших изменений [4], [5]. Поэтому является актуальным создание автоматизированных систем анализа изображений мазков крови, обеспечивающих возможность численного анализа морфологии клеток.

Формула белой крови

В организме человека присутствует примерно 5 л крови, состоящей из 3 л плазмы и 2 л форменных элементов: эритроцитов, лейкоцитов и тромбоцитов. Лейкоциты являются главными агентами

иммунной системы и делятся, в зависимости от присутствия или отсутствия гранул в их цитоплазме, на агранулоциты, включающие в себя лимфоциты и моноциты, и гранулоциты, включающие в себя нейтрофилы, эозинофилы и базофилы. Разные типы лейкоцитов отличаются морфологией, относительной концентрацией в крови и функцией [1]. При определении лейкоцитарной формулы учитывают 5 типов [1] (рис. 1):

- эозинофилы, составляющие до 5 % лейкоцитов;
- лимфоциты, составляющие от 18 до 40 % лейкоцитов;
- моноциты, составляющие от 2 до 9 % лейкоцитов;
- нейтрофилы, составляющие от 45 до 70 % лейкоцитов;
- базофилы, составляющие менее 1 % лейкоцитов.

Измерения пропорции различных типов лейкоцитов недостаточно, чтобы установить диагноз. Например, увеличение количества эозинофилов может быть связано как с аллергией, так и с наличием паразитов. Между тем состояние белой крови учитывается при постановке большого числа диагнозов, так как лейкоцитарная формула интегрально отражает состояние здоровья человека [6].

Этапы автоматизированного анализа изображений лейкоцитов

Автоматизация анализа изображений мазков крови основывается на использовании методов и алгоритмов компьютерного зрения. Анализ начинается с оценки качества изображений препаратов [7]. Для мазков крови критерием качества яв-

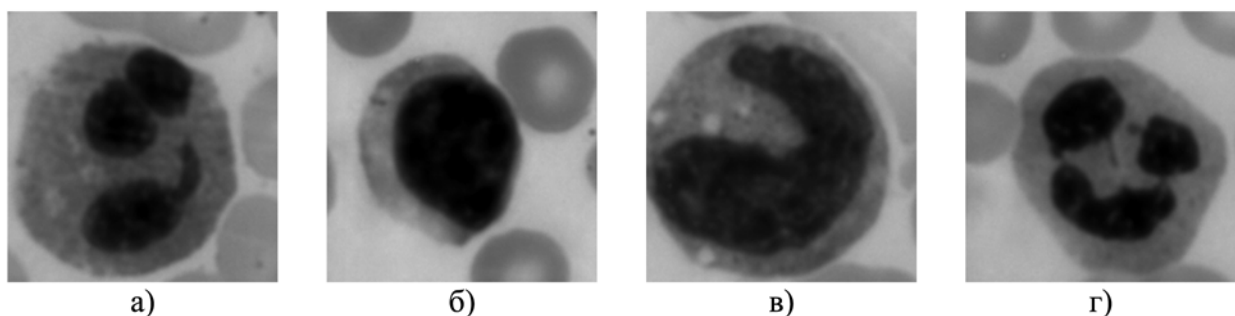


Рис. 1. а) Эозинофил; б) лимфоцит; в) моноцит; г) нейтрофил

ляется наличие области монослоя клеток. Далее решается задача обнаружения лейкоцитов, после чего выполняется сегментация обнаруженных лейкоцитов с выделением областей ядра и цитоплазмы. По сегментированному изображению лейкоцитов рассчитывается комплекс морфологических характеристик, который используется для классификации.

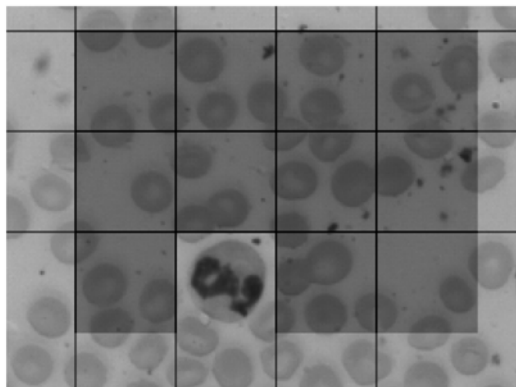


Рис. 2. Формирование фрагментов изображения: фрагмент, содержащий лейкоцит, и окружающие его фрагменты, содержащие другие клетки

Обнаружение лейкоцитов является одной из наиболее важных задач автоматизированного анализа изображений. Качество решения задачи обнаружения лейкоцитов во многом определяет скорость автоматического сканирования препарата, а также выборку клеток, по которой будет рассчитываться лейкоцитарная формула [8]. Требованиями, предъявляемыми к алгоритму обнаружения, являются минимизация вероятности пропуска клеток при невысокой вероятности ложного обнаружения и, одновременно, минимизация времени, затрачиваемого на анализ изображений.

Обнаружение лейкоцитов с использованием алгоритма AdaBoost

В настоящее время в задачах распознавания и классификации успешно применяются алгоритмы boosting (от англ. boost – поднимать, усиливать), в частности для решения задачи поиска лиц на изображении в реальном времени [9], [10]. Эти алгоритмы основываются на усилении слабых классификаторов путем их комбинирования в один сильный. В процессе обучения классификатора на выборке изображений итеративно рассчитываются веса всех классификаторов, комбинация которых позволяет минимизировать ошибки первого и второго рода [11]. В описываемой работе для обнаружения лейкоцитов на изображении препаратов был применен алгоритм Gentle AdaBoost.

База данных изображений, использованная для исследования алгоритма обнаружения, содержала цветные изображения размером 1536x2048 пикселей, без сжатия, содержащие от одного до трех лейкоцитов. Изображения регистрировались иммерсионным объективом 100x, шаг дискретизации изображений составлял 0,04 мкм/пиксел. Всего было зарегистрировано 703 изображения мазков крови, окрашенных по методу Романовского-Гизмы, для 12 пациентов. Все лейкоциты были классифицированы визуально врачом-гематологом. Изображения содержали в общей сложности 37 эозинофилов, 285 лимфоцитов, 48 моноцитов и 333 лейкоцита.

База данных изображений была разделена на обучающую выборку, состоящую из изображений клеток 10 пациентов, и контрольную выборку, состоящую из изображений клеток двух пациентов и содержащую в общей сложности 164 лейкоцита.

Из изображений обучающей выборки, с помощью специально написанной программы, были в

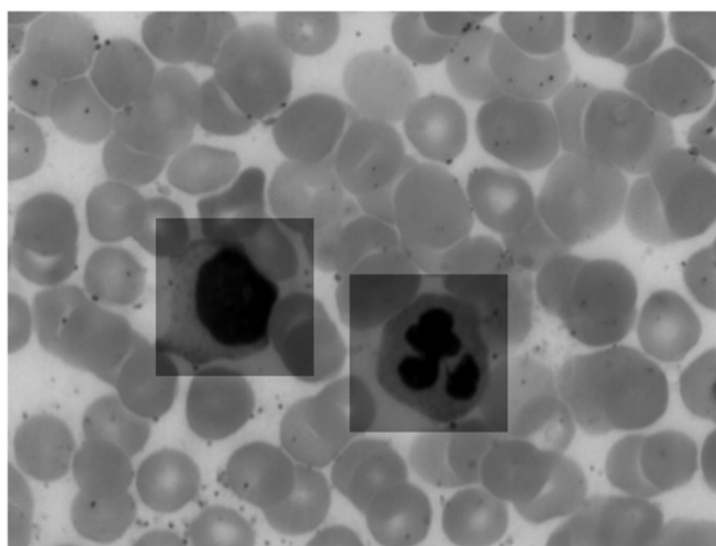


Рис. 3. Результат работы программы: обнаружение двух лейкоцитов на изображении

интерактивном режиме сформированы фрагменты изображений размером 17,6x17,6 мкм, содержащие лейкоциты, и фрагменты, содержащие другие клетки крови (рис. 2).

Таким образом, из описанной выше базы данных было создано 5832 фрагментов изображений, не содержащих лейкоциты, и 546 фрагментов изображений с лейкоцитами, по которым было проведено обучение классификатора Gentle AdaBoost. В качестве признаков фрагментов изображений были использованы средние значения шести цветовых координат в пространствах RGB и HSV.

Алгоритм обнаружения лейкоцита на изображении заключается в последовательном сканировании изображения окном, размер которого соответствует размеру моноцита, являющегося самым крупным типом лейкоцита, и определении наличия лейкоцита в выделенном фрагменте изображения по результатам классификации с использованием алгоритма Gentle AdaBoost.

Так как в процессе сканирования изображения окном его соседние положения перекрываются, может происходить повторное обнаружение одной и той же клетки. Поэтому после такого сканирования выполняется постобработка, заключающаяся в анализе расстояний между положениями окна, в которых был обнаружен лейкоцит, и объединении результатов обнаружения, относящихся к одному лейкоциту (рис. 3).

Результаты экспериментальных исследований алгоритма обнаружения показали, что полный процесс анализа одного изображения занимает в среднем 0,9 с. Однако большая часть этого времени приходится на открытие и загрузку изображения. Этап полного анализа загруженного изображения занимает менее 0,2 с, что обеспечивает возможность использования данного алгоритма в режиме реального времени при автоматическом сканировании мазка крови. Для 164 клеток в тестовой выборке вероятность пропуска составила 0 %, вероятность ложного обнаружения – 15 %.

Заключение

Проведенные работы показали актуальность разработки системы автоматизированного анализа мазков крови, в том числе определения лейкоцитарной формулы. В работе реализован алгоритм AdaBoost обнаружения лейкоцитов на изображениях мазков крови. Результаты его экспериментальных исследований показали, что он может быть использован в системах автоматизированной микроскопии мазков крови для решения задач обнаружения лейкоцитов в ходе сканирования мазка крови в режиме реального времени.

Список литературы:

1. Козинец Г.И. Атлас клеток крови и костного мозга. – М.: Издательство «Триада – X», 1998.
2. Webster J.G. Medical instrumentation application and design. Fourth Edition. – Wiley and Sons, 2010.
3. Cornbleet J. Clinical Utility of the Band Count // Clinics in Laboratory Medicine. 2002. Vol. 22. № 1.
4. Bong H.H., Gulati G.L., Ashton J.K. Differential Leukocyte Count: Manual or Automated, What Should It Be? // Yonsei Medical Journal. 1991. Vol. 32. № 4.
5. O'Neil P., Vital E., Betancourt-Loria N., Dinah M. Performance Evaluation of the Complete Blood Count and White Blood Cell Differential Parameters on the AcT 5diff Hematology Analyzer // Laboratory Hematology. 2001. Vol. 7. PP. 116-124.
6. Гаркави Л.Х., Квакина Е.Б., Уколова М.А. Адаптационные реакции и резистентность организма. – Ростов н/Д.: Изд-во Ростовского ун-та, 1990. 224 с.
7. Самородов А.В. Оценка качества цитологических препаратов // Биомедицинская радиоэлектроника. 2008. № 10. С. 39-45.
8. Самородов А.В. Автоматизированный морфологический анализ цитологических препаратов // Биомедицинская радиоэлектроника. 2009. № 10. С. 35-41.
9. Viola P., Jones M.J. Robust Real-Time Face Detection // International Journal of Computer Vision. 2004. Vol. 57 (2). PP. 137-154.
10. Vezhnevets A. GML AdaBoost Matlab Toolbox / AdaBoost Library. MSU Graphics & Media Lab. Computer Vision Group. – М.: Изд-во МГУ, 2006.
11. Meir R., Rätsch G. An Introduction to Boosting and Leveraging. – Advanced lectures on machine learning. Springer-Verlag, New York. 2003.

Жоэль Валентин Штадельманн,
аспирант,

Игорь Николаевич Спиридонов,
д-р техн. наук, профессор,
зав. кафедрой,

кафедра «Биомедицинские
технические системы»,

МГТУ им. Н.Э. Баумана,
г. Москва,

e-mail: joel.stadelmann@gmail.com

* * * * *