

УДК 004.931+004.932

С.В. Кучерявский, И.А. Беляев

Применение методов анализа многомерных данных для классификации клеток крови

На сегодняшний день существуют различные подходы для анализа и классификации изображений гистологических препаратов [1, с. 451–465; 2, с. 225–236; 3, с. 139–147; 4, с. 35], в том числе для классификации клеток крови [2–4]. Некоторые из них [3, 4] получили реализацию в коммерческих программных продуктах, входящих в состав систем автоматизированной микроскопии.

Использование таких систем позволяет повысить точность результатов исследований, а также провести сопоставление между особенностями изображений и результатами других методов исследования (биохимических, биофизических, физиологических и прочих), сделать результаты исследований более объективными [5, с. 42–43]. Кроме того, ввод изображений в компьютер позволяет не только их хранить и сортировать, но и обмениваться ими с другими исследователями, используя для этого современные телекоммуникационные средства.

Классический подход, применяемый для анализа и классификации изображений, широко используемый, в том числе в медицине и биологии, может быть представлен в виде последовательности, состоящей из четырех этапов: получение изображения, предварительная обработка, получение вектора признаков и анализ полученных признаков.

В большинстве предлагаемых на рынке программных решений для классификации используются формальные признаки, в первую очередь, морфологические, характеризующие положение и ориентацию объектов. Наряду с точностью и надежностью результатов (доля распознаваемых клеток ~95%), такие системы имеют и ряд существенных недостатков, и, в первую очередь, – высокую стоимость. Более того, приобретение только программного обеспечения не гарантирует высоких результатов на имеющемся в обычной лаборатории оборудовании и при ручном изготовлении мазков, которые часто бывают более низкого, чем необходимо для правильной работы программы, качества.

В связи с этим представляется актуальной разработка алгоритма классификации, способного давать приемлемый результат при работе

с образцами различного качества и не зависящего от аппаратной части анализатора. Инициатором исследований, результаты которых представлены в данной работе, явилась лаборатория биохимии Алтайской краевой детской клинической больницы. В ходе предварительных консультаций было предложено создать автоматический гематологический анализатор на основе имеющегося в лаборатории оборудования. На первом этапе предполагалась разработка универсального алгоритма классификации белых кровяных клеток, который бы позволил получить приемлемый для медицинской диагностики уровень распознавания при работе с изображениями, полученными при помощи лабораторного микроскопа.

Последние работы по анализу, классификации и распознаванию изображений показывают, что наиболее эффективным в этом направлении является использование подходов на основе содержательного моделирования, когда работа классификатора или анализатора строится на основе значений различных характеристик всего изображения, а не формальных признаков изучаемых объектов.

В данной работе использована классическая структура алгоритма классификации, описанная выше. Изображения мазков крови получали с помощью лабораторного оптического микроскопа, с конденсором светлого поля, снабженного VGA-камерой.

Предварительная обработка изображений заключалась в выделении на исходном изображении изображений лейкоцитов. В основе процедуры выделения интересующих объектов лежал алгоритм сегментации цветных изображений в цветовом пространстве *Lab*. Каждая выделенная клетка помещалась на белое поле размером 128x128, что достаточно для всего диапазона возможных размеров изображений кровяных клеток, определенного экспериментально (рис. 1).

Для получения вектора признаков использовался метод среднеугловых спектров (*Angle Measure Technique – АМТ*) [6, с. 83–79], разработанный в 1994 г. для количественного описания и масштабного анализа сложных геоморфных линий. Калибровка модели и классификация но-



Рис. 1. Изображения клеток крови после сегментации

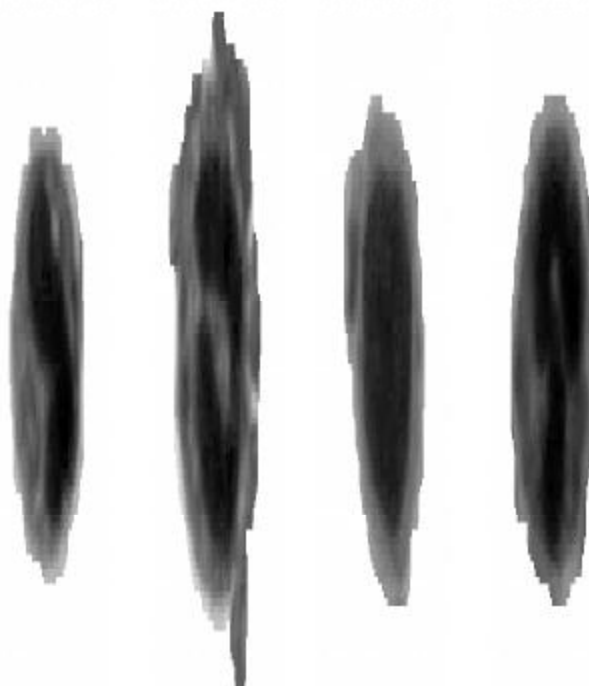


Рис. 2. Результаты классификации для калибровочного (вверху) и тестового (внизу) наборов данных

вых образцов проводились с помощью метода проекций на латентные структуры [7].

Для обучения классификатора был сформирован калибровочный набор, состоящий из 60 изображений клеток нейтрофильного и лимфатического рядов, для оценки качества распознавания использовался тестовый набор, состоящий из 96 изображений клеток этих типов.

Качество распознавания определялось как отношение количества неправильно распознанных клеток к общему количеству образцов каждого типа клеток, выраженное в процентах. Ре-

зультаты анализа представлены на рисунке 2. В качестве отклика использовалась качественная переменная y , принимающая значения 0 для нейтрофилов и 1 – для лимфоцитов. Таким образом, объекты, имеющие предсказанное значение $-0,5 \leq y < 0,5$ (лежащие ниже прямой $y = 0,5$) классифицируются как нейтрофилы, а при $0,5 \leq y < 1,5$ как лимфоциты. На графиках лимфоциты изображены окружностями, а нейтрофилы – квадратами. График классификации в верхней части рисунка 2 соответствует калибровочному набору, а классификационный график в ниж-

ней части этого рисунка – тестовому набору. Точность распознавания неизвестных образцов составила 97% для одного класса и 96% для другого.

Таким образом, комбинация метода средних угловых спектров для получения вектора при-

знаков и метода проекций на латентные структуры для построения классификационной модели позволила распознавать клетки даже на изображениях мазков крови, полученных на обычном, недорогом оборудовании.

Литература

1. Roberto Rodriguez, Teresa E. Alarcon and Juan J. Abad. Blood vessel segmentation via neural network in histological images. *Journal of Intelligent and Robotic Systems*, 2003, 36.
2. Kyungsu Kim, Jeonghee Jeon, WanKyoo Choi, Pankoo Kim, Yo-Sung Ho. Automatic Cell Classification in Human's Peripheral Blood Images Based on Morphological Image Processing. AI 2001, LNAI 2256, 2001.
3. Swolin B., Simonson P., Backman S., Lofqvist I., Bredin I., Johnsson M. Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks – evaluation of DiffMaster Octavia. *Clin. Lab. Haem.*, 2003. V. 25.
4. Соколинский Б.З. Классификация лейкоцитов с использованием методов обучающихся нейронных сетей и Watershed. / Б.З. Соколинский, В.Л. Демьянов, В.С. Медовый, А.М. Пятницкий, А.А. Парпра // *Здравоохранение и медицинская техника*. – 2005. – №4(18).
5. Медовый В.С. Автоматизированная микроскопия биоматериалов. / В.С. Медовый, А.М. Пятницкий и др. // *Здравоохранение и медицинская техника*. – 2005. – №4.
6. Andrie R., *Math. Geol.*, 1996, 16.
7. Эсбенсен К. *Анализ многомерных данных* / сокр. пер. с англ.: под ред. О.Родионовой. – М., 2005 [К.Н. Esbensen. *Multivariate Data Analysis – In Practice 4-th Ed.*, СМО, 2000]