

ПРИМЕНЕНИЕ ПРИНЦИПОВ ФОРМАЛЬНОГО МОДЕЛИРОВАНИЯ ДЛЯ КЛАССИФИКАЦИИ МЕДИЦИНСКИХ ИЗОБРАЖЕНИЙ

Беляев И. А., Кучерявский С. В.

(Россия, Барнаул)

Рассмотрены методы анализа и классификации медицинских изображений. Предложен подход для классификации лейкоцитов, основанный на принципах формального моделирования. Для построения вектора признаков используется метод среднеугловых спектров, а для построения модели и классификации — метод проекций на латентные структуры. Метод был протестирован на низкокачественных изображениях мазков крови и показал высокие результаты распознавания.

Введение. На сегодняшний день существуют различные подходы для анализа и классификации изображений гистологических препаратов [1–4], 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]. Кроме того, ввод изображений в компьютер позволяет не только хранить и сортировать, но и обмениваться ими с другими исследователями, используя для этого современные телекоммуникационные средства.

В большинстве предлагаемых на рынке программных реше-

ний 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 лейкоцитов, который бы позволил получить приемлемый для медицинской диагностики уровень распознавания при работе с изображениями, полученными при помощи лабораторного микроскопа.

Анализ последних работ по анализу, классификации и распознаванию изображений показал, что наиболее эффективным в этом направлении является использование методов формального моделирования, когда работа классификатора или анализатора строится на основе значений различных характеристик всего изображения, а не набора признаков изучаемых объектов.

Теория. Классический подход, применяемый для анализа и классификации изображений, широко используемый, в том числе и в медицине и биологии, может быть представлен в виде последовательности, состоящей из четырех этапов: получение изображения, его предварительная обработка с возможным последую-

щим выделением объектов, получение вектора признаков и классификация изображений на их основе.

В настоящее время первые два этапа не представляют научного интереса в связи с широким распространением систем получения цифровых снимков, а существующий инструментарий предобработки и сегментации изображений, разрабатываемым с 70-х годов прошлого века, позволяет решать подавляющее большинство возникающих задач.

Получение признаков — набор процедур, в результате действия которых происходит сопоставление изображению набора числовых характеристик, называемого вектором признаков. В качестве признаков наиболее часто используют морфологические и цветовые (яркостные) характеристики изучаемых объектов. Но при построении формальных моделей вектор признаков строится на основе различных характеристик изображения в целом, а не объектов на нем, например, метрики коэффициентов вейвлет-преобразования, гистограмма изображения, цветовые моменты. В настоящей работе для получения вектора признаков использовался метод среднеугловых спектров (Angle Measure Technique – AMT) [6], разработанный в 1994 г. для количественного описания и масштабного анализа сложных геоморфных линий. Позднее этот метод был предложен как обобщенный подход к анализу одномерных и двумерных сигналов.

Калибровка модели и классификация новых образцов проводились с помощью метода проекций на латентные структуры [7]. Этот метод лишен ряда существенных недостатков, характерных для искусственных нейронных сетей [8].

Эксперимент и результаты. 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). Затем для каждого изображения рассчитывался АМТ-спектр.

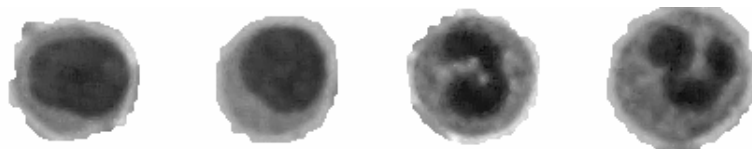


Рис. 1. Изображения клеток крови после сегментации

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.

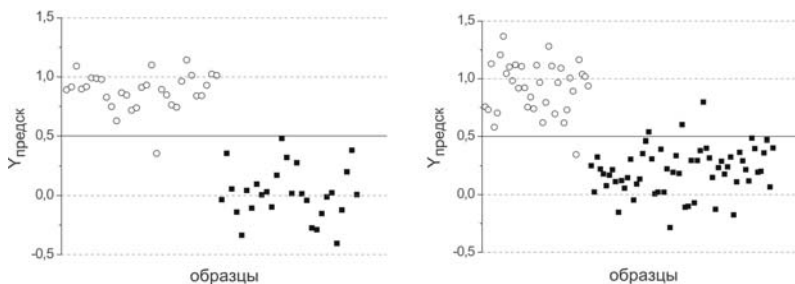


Рис. 2. Результаты классификации для калибровочного (слева) и тестового (справа) наборов данных.

Качество распознавания определялось как отношение количества неправильно распознанных клеток к общему количеству образцов каждого типа клеток, выраженное в процентах. Результаты анализа представлены на рис. 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 лимфоцитов. На графиках лимфоциты изображены окружностями, а нейтрофилы — квадратами. График результатов классификации слева соответствует калибровочному набору, а график в правой части того же рисунка — тестовому набору. Точность распознавания неизвестных образцов составила 97% для одного класса и 96% для другого.

В результате анализа ошибок классификации было выдвинуто предположение, что их причиной является алгоритм получения вектора признаков. Так, используемый на этапе развертки изображения, которая предшествует расчету АМТ-спектра, алгоритм, сохраняя информацию о радиальной симметрии объекта на изображении, вносил значительные искажения.

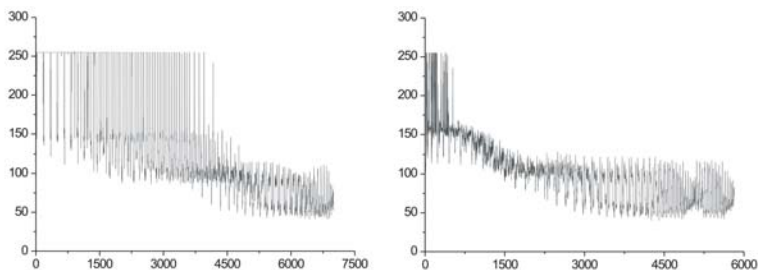


Рис. 3. Профили изображения нейтрофила. Профиль справа получен при помощи модифицированного алгоритма цепных кодов.

С целью улучшения классификации был предложен новый алгоритм развертки, основанный на методе получения цепных кодов [9]. На рис. 3 представлен два профиль для изображения нейтрофила. Профиль справа получен при помощи предложенного алгоритма развертки. Видно, что он визуально более информативен, чем представленный справа.

После этого вновь рассчитывались АМТ-спектры и строилась модель с теми же предопределенными значениями откликов. Затем проводилось тестирование полученной модели на “неизвестных” образцах.

Результаты анализа представлены на рис. 4. На графиках лимфоциты изображены окружностями, а нейтрофилы — квадра-

тами. График результатов классификации слева соответствует калибровочному набору, а график в правой части того же рисунка — тестовому набору. Точность распознавания для калибровочного набора составила 100% для клеток рассматриваемых типов.

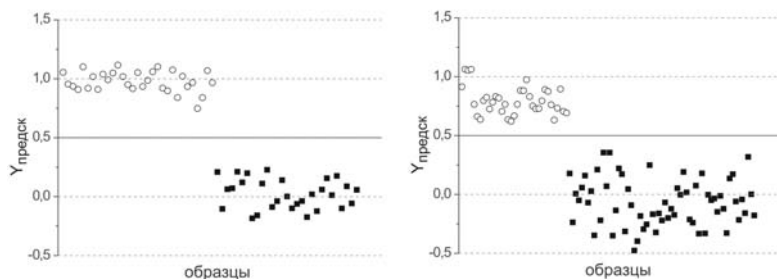


Рис. 4. Результаты классификации для калибровочного (вверху) и тестового (внизу) наборов данных. Разные типы клеток четко разделены.

Закключение. Таким образом, комбинация метода среднеугловых спектров, для получения вектора признаков, и метода проекций на латентные структуры для построения классификационной модели позволил распознавать клетки даже на изображениях мазков крови, полученных на обычном, недорогом оборудовании. Очевидно, что даже увеличение разрешения цифровой камеры от VGA до 1–2М позволит получить еще более надежные результаты и распознать более тонкие отличия в кровяных клетках.

СПИСОК ЛИТЕРАТУРЫ

1. Roberto Rodríguez, Teresa E. Alarcón and Juan J. Abad. Blood Ves-
sel Segmentation via Neural Network in Histological Images. // Journ.
of Intelligent and Robotic Systems, 2003. V. 36 № 4. P. 451–465.
2. Kyungsu Kim, Jeonghee Jeon, Wan Kyoo Choi, Pankoo Kim, Yo-
Sung Ho. Automatic Cell Classification in Human's Peripheral
Blood Images Based on Morphological Image Processing. AI
2001. // LNAI, 2001. V. 2256. P. 225–236.
3. Swolin B., Simonson P., Backman S., Lofqvist I., Bredin I., Johns-
son M. Differential counting of blood leukocytes using automated
microscopy and a decision support system based on artificial neu-

- ral networks – evaluation of DiffMaster Octavia. // Clin. Lab. Haem., 2003. V. 25. P. 139–147.
4. Соколинский Б.З., Демьянов В.Л., Медовый В.С., Парпара А.А., Пятницкий А.М. Автоматическая сортировка лейкоцитов мазка крови с использованием методов обучаемых нейронных сетей и watershed. // Здоровоохранение и медицинская техника, 2005. №4 (18). С. 35.
 5. Медовый В.С., Пятницкий А.М. и др. Автоматизированная микроскопия биоматериалов. // Здоровоохранение и медицинская техника, 2005. № 4 (18). С. 42-43.
 6. Andrie R. The angle measure technique: a new method for characterizing the complexity of geomorphic lines. // Mathematical Geology, 1996. V. 26. P. 83-97.
 7. Эсбенсен К. Анализ многомерных данных. / сокр. пер. с англ. под ред. О.Родионовой. — Из-во ИПХФ РАН, 2005. [К.Н. Esbensen. Multivariate Data Analysis – In Practice 4-th Ed., CAMO, 2000]
 8. Осовский. С. Нейронные сети для обработки информации. — Финансы и статистика, 2004 г. — 344 с.
 9. Glasbey C.A. and Horgan G.W. Image Analysis for the Biological Sciences. — John Wiley & Sons Ltd, 1995. — 218 p.

USING PRINCIPLES OF SOFT MODELING FOR CLASSIFICATION OF MEDICAL IMAGES

Belyaev I. A., Kucheryavski S. V.

(Russia, Barnaul)

Methods for processing and analysis of medical images are observed. A method for recognition and classification of white blood cells, based on principles of soft modeling is proposed. Angle Measure Technique are used for construction of feature vectors and PLS discriminant analysis is used for model construction and classification. The method is tested on the low-quality images of blood smears and shows high level of cell type recognition.