

## Сравнение методов отбора признаков для идентификации объектов на цифровых изображениях микроскопических препаратов

*А.Н. Наркевич<sup>1</sup>, Д.В. Плотников<sup>2</sup>, К.А. Виноградов<sup>1</sup>, А.В. Катаева<sup>3</sup>*

<sup>1</sup>*Красноярский государственный медицинский университет им. проф. В.Ф. Войно-Ясенецкого, Красноярск*

<sup>2</sup>*Сибирский государственный университет науки и технологий имени академика М.Ф. Решетнева, Красноярск*

<sup>3</sup>*Сибирский федеральный университет, Красноярск*

**Аннотация:** Проведено сравнение различных методов отбора признаков для идентификации объектов на цифровых изображениях микроскопических препаратов мокроты, окрашенной по методу Циля-Нильсена. Рассматривались следующие методы: метод пересечений, Шеннона, Кульбака и накопленных частот. Сделано заключение о том, что метод пересечений позволяет осуществлять отбор признаков из всего признакового пространства так, что классификационные модели дают максимальную точность классификации при наименьшем числе входных параметров.

**Ключевые слова:** метод Циля-Нильсена, распознавание объектов, распознавание изображений, отбор признаков, метод накопленных частот, метод Шеннона, метод Кульбака, метод пересечений, логистическая регрессия, дерево классификации, дискриминантный анализ.

Современные информационные и математические технологии позволяют осуществлять автоматизированный анализ различных изображений [1 – 5]. Это требует, как правило, использования методов и алгоритмов предобработки изображений, их сегментации для выделения объектов и дальнейшего изучения данных объектов для идентификации или распознавания [6, 7].

Одной из проблем распознавания объектов на изображении является их параметризация, то есть описание объектов с помощью различных параметров, которые будут использоваться для последующей идентификации [8]. Необходимо соблюдать баланс между числом таких параметров и их ценностью для идентификации при построении классификационных моделей. В настоящее время существует довольно большое число различных методов, с помощью которых можно оценить ценность параметров для их использования в процессе распознавания объектов.

---

Целью настоящего исследования является сравнение различных методов отбора признаков для идентификации объектов на цифровых изображениях микроскопических препаратов.

### **Материал и методы**

Материалом исследования явились 177 393 объекта, выделенных с использованием одного из разработанных нами алгоритмов, на цифровых изображениях микроскопических препаратов мокроты, окрашенной по методу Циля-Нильсена [9 – 11], полученных с помощью тринокулярного микроскопа Микромед 1 вар. 3-20 при увеличении  $10\times 60$  с установленной цифровой камерой TopCam UCMOS01300KPA с разрешением 0,3 МР. Каждое изображение имело разрешение 572 x 422 пикселей. Среди выделенных объектов – 6 708 являются кислотоустойчивыми микобактериями, а 170 685 – иные объекты (артефакты, скопления мокроты, части клеток и т.д.).

При оценке ценности параметров для идентификации объектов использовались 240 цветовых и морфометрических признаков [12]. Данные признаки разделены на 3 группы, которые подразделяется на подгруппы. Использована следующая классификация признаков:

1. Основные морфометрические признаки объектов:
  - а) попиксельная площадь объектов (в пикселях);
  - б) размер объектов по оси X (в пикселях);
  - в) размер объектов по оси Y (в пикселях).
2. Радиальные размеры объектов и их соотношения:
  - а) размеры от центра объекта до его края (всего 72 параметра);
  - б) соотношения противоположащих размеров от центра объекта до его края (всего 36 параметров);

в) соотношения перпендикулярных размеров от центра объекта до его края (всего 72 параметра);

г) соотношения перпендикулярных размеров от одного края объекта до его противоположного края (всего 36 параметров).

### 3. Цветовые признаки объектов:

а) характеристики цвета объектов в цветовой схеме RGB (всего 9 параметров);

б) характеристики цвета объектов в цветовой схеме HSV (всего 9 параметров);

в) характеристики цвета объектов в оттенках серого (всего 3 параметра).

Основные морфометрические признаки объектов, радиальные размеры объектов и их соотношения измерялись в пикселях. Измерение цветовых признаков осуществлялось в безразмерных единицах. Интервал возможных значений отличался в зависимости от цветовой схемы. Так, цвета объектов в цветовой схеме RGB и в оттенках серого измерялись в интервале от 0 до 255, параметры, отражающие H в цветовой схеме HSV, измерялись в интервале от 0 до 359, а параметры, характеризующие S и V, измерялись в интервале от 0 до 100.

Ценность приведенных параметров для идентификации объектов оценивалась с использованием методов оценки информативности: метод «пересечений» [12], накопленных частот, Шеннона и Кульбака [13]. На первом этапе исследования для каждого параметра объектов были рассчитаны значения информативности.

Оценка результатов осуществлялась следующим образом. На первом шаге использовался один параметр объектов, который имел максимальную информативность по каждому из приведенных выше методов. С использованием данного параметра осуществлялось построение различных

---

классификационных моделей для идентификации объектов, по результатам которой оценивалась точность классификации (%). Точность классификации определялась как доля истинно положительных и истинно отрицательных результатов классификации среди всех результатов. На следующем шаге использовались два параметра, которые имели наибольшую информативность по каждому из методов оценки информативности с последующим повторением процедуры построения классификационных моделей. Таким образом, на завершающем этапе исследования использовались 15 параметров объектов, которые по результатам применения каждого из методов оценки информативности имели большую ценность для идентификации объектов.

В качестве классификационных моделей использовались методы: логистической регрессии, дерева классификации и дискриминантный анализ.

Лучшим методом отбора признаков для идентификации объектов считался метод, с использованием которого было получено максимальное значение точности классификации при минимальном числе параметров, входящих в классификационные модели.

### **Результаты и обсуждение**

Результаты классификации объектов с использованием логистического регрессионного анализа представлены на рис. 1. Как следует из данного рисунка худшие показатели точности классификации были получены с включением в модель параметров объектов, информативность которых определялась по методу Кульбака. Так при включении в логистическое регрессионное уравнение только одного параметра, имеющего максимальную информативность по данному методу, точность классификации составила 74,2%, а при включении 15 признаков – 79,8%.

Использование метода накопленных частот с целью отбора признаков для построения логистической регрессии также не позволило получить

---

максимально возможный результат классификации. На начальном этапе была получена точность классификации – 85,8%, а при включении в уравнение 15 признаков – 86,0%.

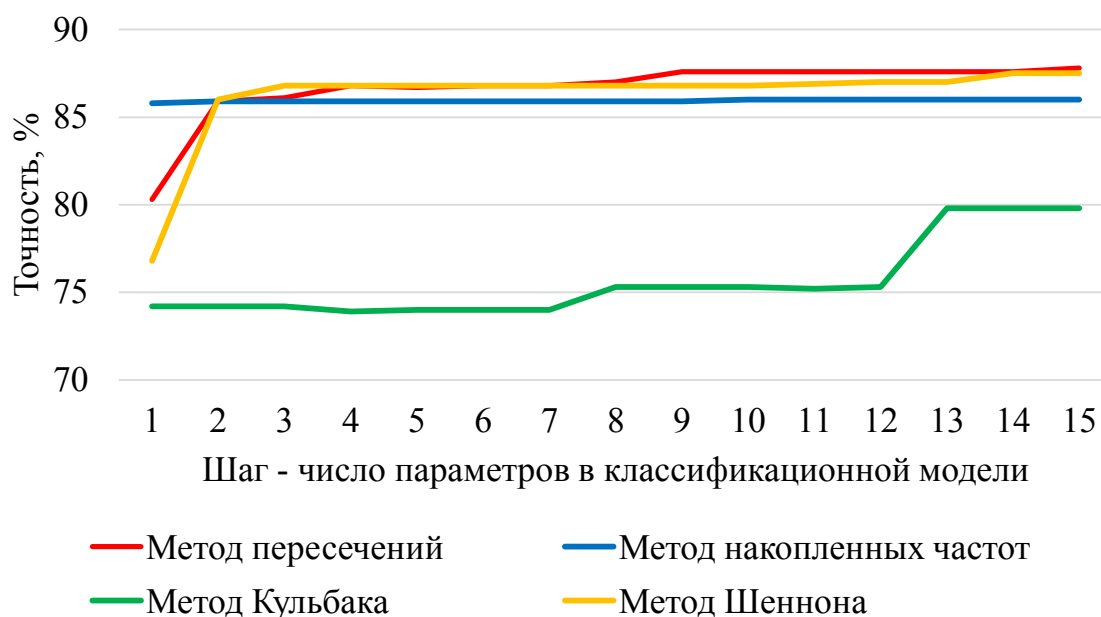


Рис. 1. – Результаты классификации с использованием логистической регрессии

Лучшие показатели точности классификации с использованием логистического регрессионного уравнения были получены при отборе признаков методами пересечений и Шеннона. Так, с использованием метода Шеннона на 3 шаге была получена точность 86,8%, что превышало точность при применении метода пересечений (86,1%). Однако, на 9 шаге с использованием метода пересечений была получена точность равная 87,6%, а на 15 шаге – 87,9%. Максимальная точность, полученная с использованием метода Шеннона, на всех 15 шагах построения логистического регрессионного уравнения составила 87,5%. Таким образом, лучшим методом отбора признаков для построения логистического регрессионного

уравнения в нашем исследовании является метод пересечений, позволяющий при включении минимального числа параметров объектов получить максимальный результат классификации.

Практически такие же результаты были получены и при использовании дерева классификации для идентификации объектов. Максимальная точность при отборе признаков по методу Кульбака была получена на 15 шаге и составила 80,7%, а при использовании метода накопленных частот точность практически не менялась и составила от 85,9% до 87,0%. Максимальная точность (88,2%) при использовании метода пересечений была получена на 11 шаге, то есть при включении 11 параметров объектов, имеющих наибольшую информативность при использовании данного метода отбора признаков. Аналогичная точность при использовании метода Шеннона получена только на 15 шаге.

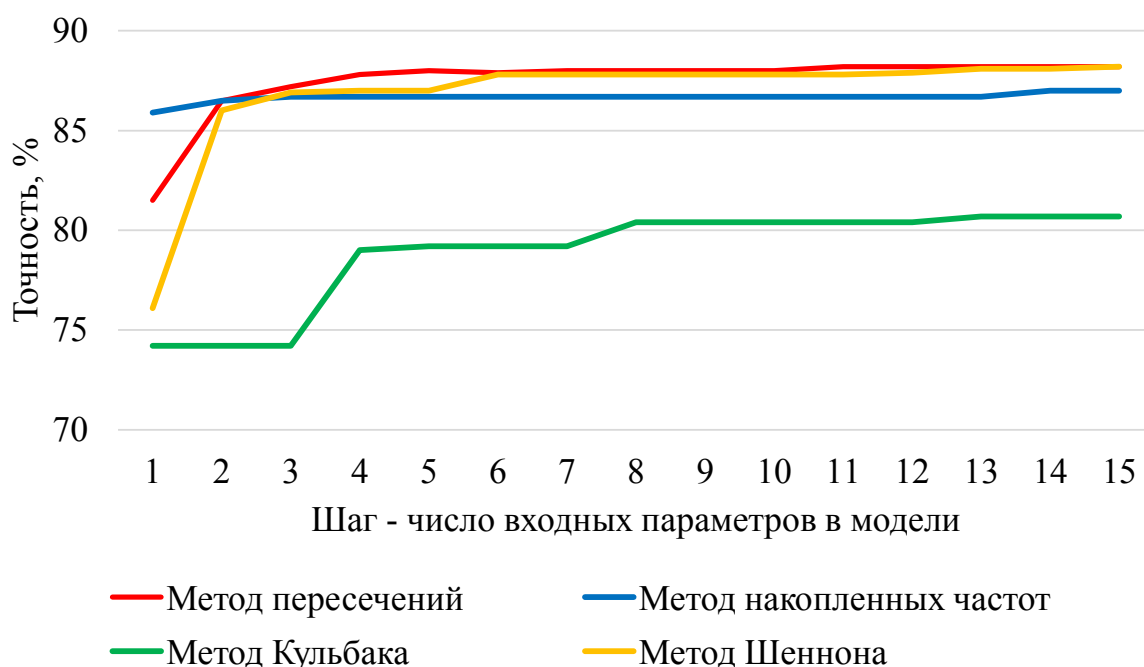


Рис. 2. – Результаты классификации с использованием дерева классификации

При использовании дискриминантного уравнения для идентификации объектов (рис. 3) максимальная точность также была получена при отборе параметров методом пересечений (87,4%). Данная точность была получена уже на 9 шаге. При этом точность классификации с использованием других методов отбора признаков на протяжении всех 15 шагов была меньше 87,4%.

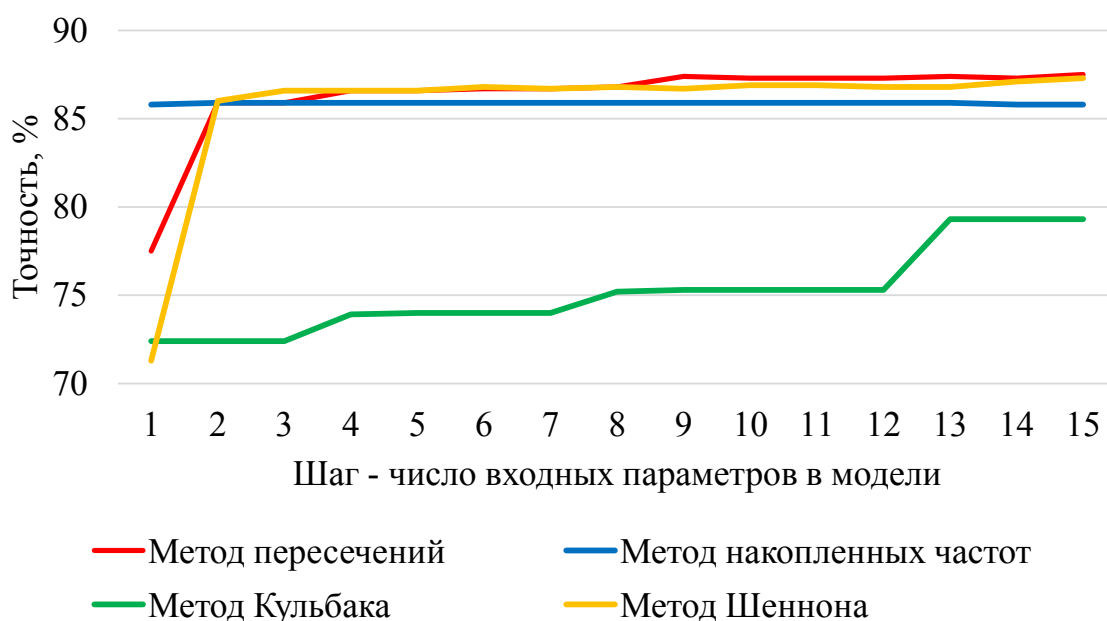


Рис. 3. – Результаты классификации с использованием дискриминантного анализа

### Заключение

Таким образом, наилучшим методом отбора признаков для распознавания объектов на цифровых изображениях микроскопических препаратов мокроты, окрашенной по методу Циля-Нильсена, является метод пересечений. Данный метод позволяет осуществлять отбор признаков из всего признакового пространства так, что классификационные модели дают максимальную точность классификации при наименьшем числе входных параметров.

## Литература

1. Ананченко В.Н., Ананченко А.И., Литвин А.В., Литвин А.А. Компьютерная обработка медицинских изображений в системе MatLab // Известия ЮФУ. Технические науки. 2002. №5. С. 68-71.
  2. Косых Н.Э., Смагин С.И., Гостюшкин В.В., Савин С.З. Система автоматизированного компьютерного анализа медицинских изображений // Информационные технологии и вычислительные системы. 2011. №3. С. 51-56.
  3. Филимонова Е.С., Тарасенко С.Л., Дыхно Ю.А., Хлебникова Ф.Б. Оценка эффективности цитологической диагностики злокачественных новообразований легких // Сибирское медицинское обозрение. 2014. №3. С. 65-69.
  4. Rodehorst V., Koschan A. Comparison and evaluation of feature point detectors. 5th International Symposium Turkish-German Joint Geodetic Days, 2006. URL: [pdfs.semanticscholar.org/1d8a/4cff206e229f480b8c92fc04f67f231cc788.pdf](https://pdfs.semanticscholar.org/1d8a/4cff206e229f480b8c92fc04f67f231cc788.pdf).
  5. Liu Q. A. location-to-segmentation strategy for automatic exudate segmentation in color retinal fundus images // Computerized medical imaging and graphics. 2016. №55. pp. 78-86.
  6. Порев В.Н. Компьютерная графика. СПб, 2002. 432 с.
  7. Воронин В.В., Сизякин Р.А., Гапон Н.В., Франц В.А., Колосов А.Ю. Алгоритм реконструкции изображений на основе анализа локальных бинарных окрестностей // Инженерный вестник Дона. 2013. №3. URL: [ivdon.ru/ru/magazine/archive/n3y2013/1857](http://ivdon.ru/ru/magazine/archive/n3y2013/1857).
  8. Самородов А.В. Автоматизированный морфологический анализ цитологических препаратов // Биомедицинская радиоэлектроника. 2009. №10. С. 35-40.
-



9. Наркевич А.Н. Алгоритмы сегментации цифровых микроскопических изображений мокроты, окрашенной по методу Циля-Нильсена // World Science: Proceedings of articles the international scientific conference. Киров, 2017. С. 431-436.
10. Наркевич А.Н., Виноградов К.А., Корецкая Н.М., Соболева В.О. Сегментация микроскопических изображений мокроты, окрашенной по методу Циля-Нильсена, с использованием вейвлет-преобразования Mexican Hat // Acta Biomedica Scientifica. 2017. №5. С.141-146.
11. Наркевич А.Н., Шеломенцева И.Г., Виноградов К.А., Сысоев С.А. Сравнение методов сегментации цифровых микроскопических изображений мокроты, окрашенной по методу Циля-Нильсена // Инженерный вестник Дона. 2017. №4. URL: [ivdon.ru/ru/magazine/archive/n4y2017/4430](http://ivdon.ru/ru/magazine/archive/n4y2017/4430).
12. Наркевич А.Н., Виноградов К.А., Корецкая Н.М. Параметризация объектов на цифровых микроскопических изображениях мокроты, окрашенной по методу Циля-Нильсена // Сибирское медицинское обозрение. 2017. №5. С. 53-59.
13. Наркевич А.Н., Виноградов К.А., Корецкая Н.М., Катаева А.В., Журбенко Е.О. Оценка информативности и отбор признаков при идентификации объектов на цифровых изображениях микроскопических препаратов, окрашенных по методу Циля-Нильсена // В мире научных открытий. 2017. №4. С. 106-121.

### References

1. Ananchenko V.N., Ananchensko A.I., Litvin A.V., Litvin A.A. Izvestiya YuFU. Tekhnicheskie nauki. 2002. №5. pp. 68-71.
  2. Kosykh N.E., Smagin S.I., Gostyushkin V.V., Savin S.Z. Informatsionnye tekhnologii i vychislitel'nye sistemy. 2011. №3. pp. 51-56.
  3. Filimonova E.S., Tarasenko S.L., Dykhno Yu.A., Khlebnikova F.B. Sibirskoe meditsinskoe obozrenie. 2014. №3. pp. 65-69.
-



4. Rodehorst V., Koschan A. 5th International Symposium Turkish-German Joint Geodetic Days, 2006. URL: [pdfs.semanticscholar.org/1d8a/4cff206e229f480b8c92fc04f67f231cc788.pdf](https://pdfs.semanticscholar.org/1d8a/4cff206e229f480b8c92fc04f67f231cc788.pdf).
5. Liu Q. A. Computerized medical imaging and graphics. 2016. №55. pp. 78-86.
6. Porev V.N. Komp'yuternaya grafika [Computer graphics]. SPb, 2002. 432 s.
7. Voronin V.V., Sizyakin R.A., Gapon N.V., Frants V.A., Kolosov A.Yu. Inzhenernyj vestnik Dona (Rus), 2013. №3. URL: [ivdon.ru/ru/magazine/archive/n3y2013/1857](http://ivdon.ru/ru/magazine/archive/n3y2013/1857).
8. Samorodov A.V. Biomeditsinskaya radioelektronika. 2009. №10. pp. 35-40.
9. Narkevich A.N. World Science: Proceedings of articles the international scientific conference. Kirov, 2017. pp. 431-436.
10. Narkevich A.N., Vinogradov K.A., Koretskaya N.M., Soboleva V.O. Acta Biomedica Scientifica. 2017. №5. pp.141-146.
11. Narkevich A.N., Shelomentseva I.G., Vinogradov K.A., Sysoev S.A. Inzhenernyj vestnik Dona (Rus), 2017. №4. URL: [ivdon.ru/ru/magazine/archive/n4y2017/4430](http://ivdon.ru/ru/magazine/archive/n4y2017/4430).
12. Narkevich A.N., Vinogradov K.A., Koretskaya N.M. Sibirskoe meditsinskoe obozrenie. 2017. №5. pp. 53-59.
13. Narkevich A.N., Vinogradov K.A., Koretskaya N.M., Kataeva A.V., Zhurbenko E.O. V mire nauchnykh otkrytiy. 2017. №4. pp. 106-121.