

ФОРМИРОВАНИЕ ПРИЗНАКОВ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА МЕДИЦИНСКОЙ ДИАГНОСТИКИ НА ОСНОВЕ МЕТОДОВ ДИСКРИМИНАНТНОГО АНАЛИЗА

Ильёсова Н.Ю., Куприянов А.В., Парингер Р.А.

Институт систем обработки изображений РАН,

Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет) (СГАУ)

Аннотация

Рассматривается компьютерная система диагностики глазных заболеваний, которую для повышения качества диагностики предлагается дополнить алгоритмом формирования информативных признаков, с использованием методов дискриминантного анализа. Описан способ получения оценки информативности. При помощи классификации методом опорных векторов было проведено исследование, подтверждающее эффективность сформированных признаков для классификации изображений глазного дна. Алгоритм обладает достаточной степенью универсальности и может быть применён для повышения информативности любого набора признаков.

Ключевые слова: глазное дно, классификация изображений сосудов, дискриминантный анализ, метод опорных векторов.

Введение

В статье рассматривается методология выделения диагностически значимой информации на изображениях кровеносных сосудов глазного дна (рис. 1). Доступность к неинвазивному осмотру и визуализации делает сосуды глазного дна весьма информативными для анализа и прогностически значимыми в оценке состояния всего организма.

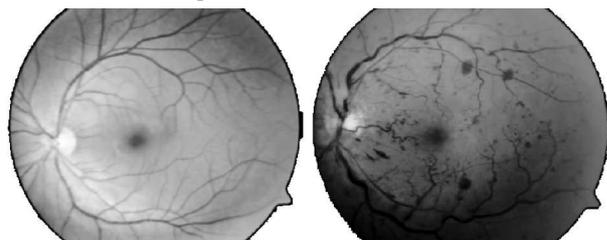


Рис. 1. Примеры диагностических изображений глазного дна. Слева – норма, справа – патология

Совместно с врачами кафедры офтальмологии Медико-стоматологического университета г. Москвы были проведены исследования для определения диагностических возможностей методов цифрового анализа изображений глазного дна [1]. Была разработана методика диагностирования глазных заболеваний, основанная на построении количественных оценок элементов патоморфологической картины, используемой для формирования оценки степени патологии [2].

Был предложен глобальный набор геометрических признаков [3–8]: средний диаметр D_{cp} , прямолинейность P , чёткообразность S , амплитуда колебаний толщины A_0 , частота колебаний толщины ω_0 , извилистость толщины I_0 , амплитуда колебаний трассы A_1 , частота колебаний трассы ω_1 , извилистость трассы I_1 , – достаточно точно соответствующий медико-диагностическим признакам картины глазного дна, рекомендованным в медицинскую практику.

Дополнительной информацией для решения задачи повышения качества диагностики по изображениям кровеносных сосудов глазного дна может служить уровень взаимодействия между признаками, который выражается величиной коэффициентов корреляции. Если связь между какими-то признаками является достаточно

сильной, то возникает вопрос о целесообразности использования этих признаков и появляется возможность уменьшить размерность пространства признаков либо путём отказа от каких-то из них и использования только одного из связанных признаков, либо формированием на их базе некоторого совокупного признака. В первом случае мы уменьшаем количество вычислений, оставляя признаки, характеризующиеся наилучшим значением какого-либо критерия. Во втором случае появляется возможность получить новый, более эффективный признак для классификации.

Существует несколько подходов, применяемых для повышения информативности признаков: корреляционный анализ, регрессионный анализ, факторный анализ, кластерный анализ, дискриминантный анализ [9].

При анализе признаков по критерию эффективности классификации наибольший интерес представляют методы дискриминантного анализа, а точнее, методы интерпретации межгрупповых различий. Дискриминантный анализ – раздел многомерного статистического анализа, который позволяет изучать различия между двумя и более классами объектов по нескольким переменным одновременно [10]. При этом определяется возможность различить классы с использованием заданного набора признаков, а также решается задача выбора признаков, обеспечивающих наилучшее разделение классов.

Таким образом, включение в компьютерную систему диагностики глазных заболеваний алгоритма формирования пространства эффективных признаков на основе дискриминантного анализа позволит максимизировать разделимость классов «норма» и различных степеней патологии.

1. Алгоритм дискриминантного анализа

Основным предположением дискриминантного анализа является то, что классификация будет осуществляться тем лучше, чем меньше рассеяние точек признакового пространства относительно центра внутри класса и чем больше расстояние между центрами классов [10].

Пусть дана выборка, состоящая из n элементов, разделённых на g классов, и содержащая p признаков. В дискриминантном анализе мерой эффективности выборки являются критерии разделимости, которые рассчитываются по формулам:

$$J_1 = tr((\mathbf{T})^{-1}\mathbf{B}), \tag{1}$$

$$J_2 = \ln\{|\mathbf{T}|/|\mathbf{W}|\}, \tag{2}$$

где $\mathbf{T} = \mathbf{B} + \mathbf{W}$.

\mathbf{B} – матрица межгруппового рассеяния, элементы которой рассчитываются по формуле:

$$b_{ij} = \sum_{k=1}^g n_k (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j), i, j = \overline{1, p}, \tag{3}$$

\mathbf{W} – матрица внутригруппового рассеяния, элементы которой рассчитываются по формуле:

$$w_{ij} = \sum_{k=1}^g \sum_{m=1}^{n_k} (x_{ikm} - \bar{x}_{ik})(x_{jkm} - \bar{x}_{jk}), i, j = \overline{1, p}, \tag{4}$$

x_{ikm} – значение i -го признака для m -го элемента в классе k , $\bar{x}_{ik} = 1/n_k \sum_{m=1}^{n_k} x_{ikm}$ – среднее значение i -го признака в классе k , $\bar{x}_i = (1/n) \sum_{k=1}^g n_k \bar{x}_{ik}$ – среднее значение i -го признака по всем классам, n_k – число элементов в классе k .

Чем больше значение критерия, тем больше разделимость классов.

Пусть $\mathbf{x} = [x_1, x_2, \dots, x_p]^T$ – исходный вектор признаков. Рассмотрим алгоритм формирования новых признаков $\mathbf{y} = [y_1, y_2, \dots, y_m]^T$.

1. Для матрицы $\mathbf{T}^{-1}\mathbf{B}$ определим значения собственных векторов $\mathbf{v}_i, i = \overline{1, p}$.

2. Определим вектора нормированных коэффициентов $\beta_i = [\beta_0, \beta_1, \dots, \beta_p], i = \overline{1, m}$, где элементы $\beta_i, i = \overline{0, p}$ вычисляются следующим образом:

$$\beta_0 = -\sum_{i=1}^p \beta_i \bar{x}_i, \beta_i = v_i \sqrt{n-g}, i = \overline{1, p}. \tag{5}$$

3. Вычислим элементы вектора новых признаков по формуле:

$$y_i = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p, i = \overline{1, m}. \tag{6}$$

Число сформированных признаков не может превышать числа исходных.

Отметим, что для оценки вклада каждого исходного признака в значение каждого нового признака можно использовать *стандартизованные коэффициенты*:

$$c_i = \beta_i \sqrt{w_{ii}/n-g}, i = \overline{1, p}. \tag{7}$$

Если абсолютная величина коэффициента для данной переменной мала, то эту переменную можно исключить, тем самым сократив число переменных [11].

2. Описание эксперимента

Для решения поставленной задачи было реализовано программное обеспечение. Исследование работоспособности алгоритма проводилось на данных, полученных по результатам диагностики 151 пациента с различными стадиями заболевания. Было обработано 302

изображения: по одному для левого и правого глаза каждого пациента. Для количественного выражения медико-диагностических характеристик сосудов были сформированы наборы глобальных признаков. Для получения признаков использовалась разработанная компьютерная система диагностики глазных заболеваний.

Так как врачи рассматривают вены и артерии первого и второго порядков отдельно, было сформировано четыре группы сосудов, как представлено в табл. 1.

Табл. 1. Группы сосудов

	Артериолы	Венулы
Первый порядок	Группа 1	Группа 3
Второй порядок	Группа 2	Группа 4

Каждая группа делится на пять различных классов, в зависимости от степени развития заболевания. «Класс 0» соответствует отсутствию патологии. «Класс 1» – «Класс 4» соответствуют степеням развития диабетической ретинопатии. Распределение векторов признаков на группы и классы представлено в табл. 2. Классификация проводилась по степеням патологии внутри групп.

Табл. 2. Распределение векторов признаков по группам и классам

Классы	0	1	2	3	4
Группа 1	90	340	540	530	160
Группа 2	710	400	885	530	150
Группа 3	250	420	630	530	410
Группа 4	520	500	720	490	450

Рассмотрим основные этапы вычислительного эксперимента:

1. Формируем выборку. Получаем данные с использованием уже имеющихся трасс натуральных изображений.

2. Разбиваем полученную выборку сосудов на классы на основе формальных признаков.

3. Оцениваем глобальные признаки сосудов $D_{cp}, S, P, A_0, \omega_0, I_0, A_1, \omega_1, I_1$.

4. Выбираем комбинацию исходных признаков.

5. Рассчитываем критерии J_1, J_2 для выбранных признаков.

6. Формируем новые признаки по представленному ранее алгоритму.

7. Рассчитываем критерии J_1, J_2 для сформированных признаков.

При проведении эксперимента учитывались все возможные комбинации исходных признаков. В результате наилучшие значения критериев были получены при формировании четырёх новых признаков с использованием всех девяти исходных. Значения полученных стандартизованных коэффициентов для «Группа 1» и «Группа 4» приведены в табл. 3 и 4 соответственно.

Проанализировав результаты эксперимента, можно сделать вывод о том, что для каждой группы характерен свой набор сформированных признаков. Критерием в данном случае является величина абсолютного значения стандартизованного коэффициен-

та: чем больше значение, тем более весомый вклад внёс признак. На основании полученных результатов можно сделать вывод о том, что признак частота колебаний толщины ω_0 является весомым для обеих групп, но для «Группа 1» также весомым является признак чёткообразность S , а для «Группа 4» – извилистость толщины I_0 .

Табл. 3. Значения стандартизованных коэффициентов.
Группа 1

	p_0	p_1	p_2	p_3
D_{cp}	0,291	0,015	-0,814	0,017
S	-30,142	0,768	116,796	9,595
P	0,042	0,004	-1,160	0,004
A_0	-0,236	-0,004	7,944	0,006
ω_0	-0,578	0,788	-135,584	-0,765
I_0	-1,400	-0,568	63,797	-0,275
A_1	0,042	-0,004	-0,098	0,010
ω_1	-3,026	-0,088	-3,326	0,282
I_1	7,598	-0,117	-28,633	-2,533

Табл. 4. Значения стандартизованных коэффициентов.
Группа 4

	p_0	p_1	p_2	p_3
D_{cp}	-0,319	10,583	2,415	-0,136
S	-1,082	4,256	3,201	0,318
P	-0,158	0,184	1,146	-0,080
A_0	-0,126	28,249	-5,741	1,572
ω_0	-28,731	383,399	143,014	9,390
I_0	34,427	-304,961	-90,478	-5,145
A_1	-0,066	1,284	-0,370	-0,002
ω_1	-9,422	-49,036	-5,399	0,250
I_1	1,217	0,019	-1,727	-0,214

По значениям критериев разделимости производился выбор нового набора признаков. Для каждого набора новых признаков проводился расчёт ошибки классификации, при этом использовался метод опорных векторов [12]. При проведении эксперимента для оценивания вероятности ошибки классификации использовался метод исключения одного объекта [10].

Полученные значения критериев разделимости и ошибки классификации для всех групп сосудов представлены в табл. 5.

Табл. 5. Значения критериев разделимости классов и ошибка классификации

Группа		J_1	J_2	Ошибка, %
1	до	1,98	3,52	10,5
	после	2,63	5,06	2,4
2	до	2,27	4,07	7,3
	после	2,59	5,28	1,8
3	до	2,06	3,61	6,1
	после	2,59	4,86	2
4	до	2,36	4,01	14,3
	после	2,61	4,86	1,8

Заключение

В компьютерную систему диагностики глазных заболеваний [1] был включён алгоритм формирования пространства эффективных признаков на основе дискриминантного анализа. Алгоритм позволяет повысить эффективность классификации сосудов глазного дна по классам нормы и различным степеням патологии, при этом ошибка классификации сосудов составляет не более 2,4 %.

Показано, что для каждой группы сосудов глазного дна важен свой базовый набор диагностических признаков, что подтверждается клиническими исследованиями. Применение алгоритма для каждой группы сосудов приводит к улучшению критерия разделимости изображений сосудистых систем на классы патологий, включающие 4 стадии диабетической ретинопатии: «Группа 1» – на 32 %, «Группа 2» – на 14,1 %, «Группа 3» – на 25,9 %, «Группа 4» – на 15,8 %.

Предложенный алгоритм обладает достаточной степенью универсальности и применим для повышения информативности любого набора признаков.

Внедрение проблемно-ориентированного программного комплекса позволит эффективно решать задачи анализа изображений глазного дна для ранней диагностики заболеваний.

Благодарности

Работа выполнена при государственной поддержке Министерства образования и науки РФ в рамках реализации мероприятий Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013-2020 годы; грантов РФФИ 12-01-00237-а, 14-01-00369-а, 14-07-97040-р_поволжье_а; программы № 6 фундаментальных исследований ОНИТ РАН «Биоинформатика, современные информационные технологии и математические методы в медицине» 2014 г.

Литература

1. **Ильясова, Н.Ю.** Диагностический комплекс анализа изображений сосудов глазного дна // Биотехносфера. – 2014. – № 3. – С. 132-138.
2. **Ильясова, Н.Ю.** Методы цифрового анализа сосудистой системы человека. Обзор литературы // Компьютерная оптика. – 2013. – Т. 37, № 4. – С. 511-535. – ISSN 0134-2452.
3. **Ильясова, Н.Ю.** Биомеханические характеристики сосудов для цифрового анализа изображений глазного дна / Н.Ю. Ильясова, А.В. Куприянов, Н.А. Гаврилова, Г.А. Шилкин, Н.И. Ланевская // Биомеханика глаза. III семинар: сборник трудов. – М., 2002. – С. 18-30.
4. **Сойфер, В.А.** Методы компьютерного анализа диагностических изображений глазного дна / В.А. Сойфер, Н.Ю. Ильясова, А.В. Куприянов, А.Г. Храмов, М.А. Ананьин // Технология живых систем. – 2008. – Т. 5, № 5-6. – С. 61-71.
5. **Ильясова, Н.Ю.** Информационные технологии анализа изображений в задачах медицинской диагностики / Н.Ю. Ильясова, А.В. Куприянов, А.Г. Храмов. – М.: Радио и связь, 2012. – 424 с.
6. **Ильясова, Н.Ю.** Измерение биомеханических характеристик сосудов для ранней диагностики сосудистой патологии глазного дна / Н.Ю. Ильясова, А.В. Куприянов,

- М.А. Ананьин // Компьютерная оптика. – 2005. – № 27. – С. 165-169. – ISSN 0134-2452.
7. **Куприянов, А.В.** Разработка информационной технологии оценивания геометрических параметров изображений глазного дна / А.В. Куприянов, Н.Ю. Ильясова // Вестник Самарского государственного аэрокосмического университета им. академика С.П. Королёва. – 2008. – № 2(15). – С. 221-235.
 8. **Ильясова, Н.Ю.** Оценивание геометрических признаков пространственной структуры кровеносных сосудов // Компьютерная оптика. – 2014. – Т. 38, № 3. – С. 529-538. – ISSN 0134-2452.
 9. **Симчера, В.М.** Методы многомерного анализа статистических данных / В.М. Симчера. – М.: Финансы и статистика, 2008. – 400 с. – ISBN: 978-5-279-03184-9
 10. **Фукунага, К.** Введение в статистическую теорию распознавания образов / К. Фукунага. – М.: Наука, 1979. – 270 с.
 11. **Ким, Дж.-О.** Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка. – М.: Финансы и статистика, 1989. – 215 с.
 12. **Мookiahа, M.R.K.** Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features / M.R.K. Mookiahа, U.R. Acharyaa, C.M. Lima, A. Petznickb, S. Jasjit // Knowledge-Based Systems. – 2012. – Vol. 33. – P. 73-82.
4. **Soifer, V.A.** Methods for Computer Diagnostics using Eye's Fundus Images / V.A. Soifer, N.Yu. Ilyasova, A.V. Kupriyanov, A.G. Khramov, M.A. Ananin // Technologies of the Living Systems. – 2008. – Vol. 5, N 5-6. – P. 61-71. – (In Russian).
 5. **Ilyasova, N.Yu.** Information technologies of image analysis in medical diagnostics / N.Yu. Ilyasova, A.V. Kupriyanov, A.G. Khramov. – Moscow: "Radio I Svyazj" Publisher, 2012. – 424 p. – ISBN 5-89776-014-4. – (In Russian).
 6. **Ilyasova, N.Yu.** Measurement of the biomechanical vessels parameter for the diagnostics of the early stages of the retina vascular pathology / N.Yu. Ilyasova, A.V. Kupriyanov, M.A. Ananin // Computer Optics. – 2005. – Vol. 27. – P. 165-169. – (In Russian).
 7. **Kupriyanov, A.V.** Development of the information technology for estimation of fundus image geometric parameters / A.V. Kupriyanov, N.Yu. Ilyasova // Herald of Samara State Aerospace University named by S.P. Korolyev. – 2008. – Vol. 2(15). – P. 221-235. – (In Russian).
 8. **Ilyasova, N.Yu.** Estimating the geometric features of a 3D vascular structure / N.Yu. Ilyasova // Computer Optics. – 2014. – Vol. 38(3). – P. 529-538. – ISSN 0134-2452.
 9. **Simchera, V.M.** Methods of multivariate statistical analysis / V.M. Simchera. – Moscow: "Financy I Statistika" Publisher, 2008. – 400 p. – ISBN: 978-5-279-03184-9. – (In Russian).
 10. **Fukunaga, K.** Introduction to statistical pattern recognition / K. Fukunaga. – New York and London: Academic Press, 1972. – P. 369.
 11. **Kim, J.-A.** Factor, discriminant and cluster analysis / J.-A. Kim, Ch.U. Myuller, WR Klekka. – Moscow: "Financy I Statistika" Publisher, 1989. – 215 p. – (In Russian).
 12. **Mookiahа, M.R.K.** Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features / M.R.K. Mookiahа, U.R. Acharyaa, C.M. Lima, A. Petznickb, S. Jasjit // Knowledge-Based Systems. – 2012. – Vol. 33. – P. 73-82.

References

FORMATION OF FEATURES FOR IMPROVING THE QUALITY OF MEDICAL DIAGNOSIS BASED ON DISCRIMINANT ANALYSIS METHODS

*N.Yu. Ilyasova, A.V. Kupriyanov, R.A. Paringer
Image Processing Systems Institute, Russian Academy of Sciences,
Samara State Aerospace University*

Abstract

A computer aided diagnosis system for ocular disease detection is considered. To improve the quality of diagnostics we propose an algorithm for the informative features formation, using methods of discriminant analysis. A method for obtaining an informative value estimate is described. A research confirming the efficiency of the feature-based classification of fundus images was conducted by means of classification by support vector machine. The algorithm possesses a sufficient level of universality and may be applied to increase the informative value of any feature set.

Key words: ocular fundus, classification of the vessels images, linear discriminant analysis.

Сведения об авторах

Ильясова Наталья Юрьевна, 1966 года рождения. В 1991 году окончила с отличием Самарский государственный аэрокосмический университет имени С.П. Королёва (СГАУ). В 1997 году защитила диссертацию на соискание степени кандидата технических наук. В настоящее время работает старшим научным сотрудником в Учреждении Российской академии наук Институте систем обработки изображений РАН и одновременно доцентом кафедры технической кибернетики СГАУ. Круг научных интересов включает цифровую обработку сигналов и изображений, анализ и интерпретацию биомедицинских изображений. Имеет более 100 публикаций, в том числе 35 статей и три монографии (в соавторстве).

E-mail: ilyasova@smr.ru.



Nataly Yu. Pyasova, (b. 1966), graduated (1991) from the S.P. Korolyov Samara State Aerospace University (SSAU). She received her PhD in Technical Sciences (1997). At present, she is a senior researcher at the Image Processing Systems Institute of the Russian Academy of Sciences, and holding a part-time position of Associate Professor at SSAU's Technical Cybernetics sub-department. The area of interests includes digital signals and image processing, pattern recognition and artificial intelligence, biomedical imaging and analysis. She's list of publications contains more than 100 scientific papers, including 35 articles and 3 monographs published with coauthors.



Куприянов Александр Викторович, 1978 года рождения, доктор технических наук, старший научный сотрудник ИСОИ РАН, по совместительству доцент кафедры технической кибернетики СГАУ. В 2001 году окончил с отличием факультет информатики СГАУ по специальности «Прикладная математика». В 2004 году защитил кандидатскую диссертацию, в 2013 году – докторскую диссертацию. Круг научных интересов включает цифровую обработку сигналов и изображений, распознавание образов и искусственный интеллект, анализ и интерпретация биомедицинских сигналов и изображений. Имеет более 90 публикаций, в том числе 35 статей, две монографии (в соавторстве), одна из которых издана на английском языке. Член Российской ассоциации распознавания образов и анализа изображений.

E-mail: akupr@smr.ru .

Alexander Victorovich Kupriyanov, (b. 1978) senior researcher at the Image Processing Systems Institute of the Russian Academy of Sciences. At present he is holding a part-time position of Associate Professor at SSAU's Technical Cybernetics sub-department. Graduated (2001) from Samara State Aerospace University (SSAU). PhD in Technical sciences in 2004, DrSC in 2013. The area of interests includes digital signals and image processing, pattern recognition and artificial intelligence, biomedical imaging and analysis. He's list of publications contains more than 90 scientific papers, including 35 articles and 2 monographs published with coauthors, one is published in English. The member of Russian Association of Pattern Recognition and Image Analysis.



Парингер Рустам Александрович, 1990 года рождения, аспирант Самарского государственного аэрокосмического университета имени академика С.П. Королёва (национальный исследовательский университет), стажёр-исследователь ИСОИ РАН. В 2013 году окончил факультет информатики СГАУ. Круг научных интересов включает цифровую обработку изображений, распознавание образов, а также анализ и интерпретацию биомедицинских сигналов и изображений.

E-mail: rusparinger@gmail.com .

Rustam Aleksandrovich Paringer, (b. 1990) post-graduate student of Samara State Aerospace University named after S.P. Korolyov (National Research University), trainee-researcher at the Image Processing Systems Institute of the Russian Academy of Sciences. The area of interests includes image processing, pattern recognition, biomedical imaging and analysis.

Поступила 8 сентября 2014г.