

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ПРИЗНАКОВ ДЛЯ ДИАГНОСТИКИ СОСУДИСТОЙ ПАТОЛОГИИ

© 2015 Н.Ю. Ильясова^{1,2}, Р.А. Парингер^{1,2}

¹ Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет)

² Институт систем обработки изображений РАН, г. Самара

Поступила в редакцию 30.07.2015

В статье предлагается метод фильтрации недостоверных данных с использованием алгоритма кластеризации для повышения качества работы разработанного алгоритма дискриминантного анализа векторов признаков. Дискриминантный анализ используется для оценки информативности диагностических признаков сосудов по критерию эффективности классификации и для формирования новых признаков с целью улучшения качества диагностики. В качестве интегральных показателей состояния сосудов глазного дна используется глобальный набор геометрических признаков, являющийся полной характеристикой диагностических изображений и позволяющий проводить диагностику сосудистой патологии.

Ключевые слова: классификация изображений сосудов глазного дна, фильтрация признаков, кластеризация методом k-средних, дискриминантный анализ, критерий разделимости.

Существует способ диагностики различных заболеваний, таких как сахарный диабет, гипертоническая болезнь и др., по состоянию сосудистой системы глазного дна на основе измерения различных параметров сосудов. При этом, так как нет уверенности в том, что существует прямая связь между геометрическими параметрами сосудов и заболеванием, одним из представляющих интерес подходом является использование статистических методов, которые позволяют найти закономерности и связи. При анализе признаков основной задачей является определение некоторого критерия, отображающего эффективность данного параметра при возможности формировании нового. Дискриминантный анализ является разделом многомерного статистического анализа, позволяющим изучать различия между двумя и более группами объектов по нескольким переменным одновременно. Основной целью дискриминации является нахождение такой линейной комбинации признаков, которая бы оптимально разделила рассматриваемые классы. Ниже представлено применение метода дискриминантного анализа для оценки диагностических признаков сосудов глазного дна по критерию эффективности классификации и формирования новых признаков для повышения разделимости степени патологии.

Совместно с врачами Медико-стоматологического университета г. Москвы кафедры глазных

болезней были проведены исследования на основе цифрового анализа изображений глазного дна [1] пациентов, имеющих различные стадии диабетической ретинопатии (ДР), которая является осложнением сахарного диабета (СД). Была разработана методика диагностирования глазных заболеваний на основе оценки глобальных сосудистых характеристик (признаков). В работе рассматриваются геометрические признаки, предложенные в [2-4]. Такими признаками являются: средний диаметр, прямолинейность, чёткообразность, амплитуда и частота колебаний радиуса сосуда, извилистость радиуса, амплитуда и частота колебаний трассы, извилистость трассы, которые соответствуют диагностическим признакам сосудов глазного дна.

При наличии двух или более классов (в нашем случае – 5 классов, включающих норму и 4 степени диабетической ретинопатии) задача выбора признаков состоит в отборе таких, которые являются наиболее эффективными с точки зрения разделимости классов [5, 6]. В дискриминантном анализе критерии разделимости классов формируются с использованием матриц рассеяния внутри классов и матриц рассеяния между классами [6, 7].

Матрица рассеяния внутри классов показывает разброс объектов относительно векторов математических ожиданий классов:

$$\mathbf{W} = \sum_{k=1}^g (\mathbf{X}_k - \bar{\mathbf{x}}_k)(\mathbf{X}_k - \bar{\mathbf{x}}_k)'$$
 где данным k – класса будут соответствовать вектора средних $\bar{\mathbf{x}}_k = [\bar{x}_{1k} \bar{x}_{21k} \dots \bar{x}_{pk}]$, g – общее количество классов.

Элементы матрицы рассеяния между классами \mathbf{B} рассчитывается по формуле:

Наталья Юрьевна Ильясова кандидат технических наук, доцент кафедры технической кибернетики СГАУ, старший научный сотрудник ИСОИ РАН.

E-mail: ilyasova@smr.ru,

Парингер Рустам Александрович, аспирант СГАУ, стажёр-исследователь ИСОИ РАН.

E-mail: rusparinger@gmail.com

$b_{ij} = \sum_{k=1}^g n_k (\bar{x}_{ik} - \bar{x}_i)(\bar{x}_{jk} - \bar{x}_j)$, $i, j = \overline{1, p}$, $\bar{x}_i = (1/n) \sum_{k=1}^g n_k \bar{x}_{ik}$ – среднее значение признака по всем классам, n_k – число объектов в k -м классе, $\bar{x}_{ik} = 1/n_k \sum_{m=1}^{n_k} \bar{x}_{ikm}$ – среднее значение признака в классе k , x_{ikm} – значение i -го признака для m -го объекта в k -м классе. Матрицы W и B содержат всю основную информацию о зависимости внутри классов и между классами. Для того чтобы получить критерий разделимости классов, нужно связать с этими матрицами некоторое число. Такие числа называют критериями разделимости. Существует несколько способов расчёта критериев разделимости [6].

ФОРМИРОВАНИЕ ПРИЗНАКОВ, МАКСИМИЗИРУЮЩИХ КРИТЕРИЙ РАЗДЕЛИМОСТИ

Рассмотрим способ формирования пространства эффективных признаков, максимизирующих критерий разделимости на основе методов дискриминантного анализа [6]. В качестве используемого критерия определим $J = tr(\mathbf{T}^{-1}\mathbf{B})$. Пусть $\mathbf{x} = [x_1 x_2 \dots x_p]^T$ – исходный вектор признаков, тогда $\mathbf{y} = [y_1 y_2 \dots y_m]^T$ – новый вектор признаков. Предположим, что выбираем m признаков, полученных умножением матрицы преобразования \mathbf{A} размерности $m \times p$ ($m < p$) на исходный p -мерный вектор \mathbf{x} : $\mathbf{y} = \mathbf{Ax}$. Задача выбора признаков, максимизирующих критерий J_1 , сводится к нахождению матрицы преобразования \mathbf{A} . Матрицы рассеяния в пространстве \mathbf{Y} , соответствующие матрицам \mathbf{B} и \mathbf{T} в пространстве \mathbf{X} , имеют вид: $\mathbf{B}_m = \mathbf{ABA}^T$, $\mathbf{T}_m = \mathbf{ATA}^T$. Пусть $\lambda_i, \mathbf{t}_i, i = 1, 2, \dots, p$ и $\mu_j, \mathbf{v}_j, j = 1, 2, \dots, m$ – соответственно собственные значения и собственные векторы матриц $\mathbf{T}^{-1}\mathbf{B}$ и $\mathbf{T}_m^{-1}\mathbf{B}_m$. В таком случае критерий J для p и m признаков принимает вид:

$$J(p) = tr\mathbf{T}^{-1}\mathbf{B} = \sum_{i=1}^p \lambda_i,$$

$$J(m) = tr\mathbf{T}_m^{-1}\mathbf{B}_m = \sum_{j=1}^m \mu_j.$$

Оптимальным является такое преобразование \mathbf{A} , при котором собственные значения матрицы $\mathbf{T}_m^{-1}\mathbf{B}_m$ в соответствующем m -мерном подпространстве будут равны: $\mu_i = \lambda_i, i = \overline{1, m}$, где λ_i упорядочены следующим образом: $\lambda_1 > \lambda_2 > \dots > \lambda_p$. Этого можно достигнуть, составив матрицу \mathbf{A}^T из первых m собственных векторов $\mathbf{v}_i, i = \overline{1, m}$: $\mathbf{A}^T = [\mathbf{v}_1 \mathbf{v}_2 \dots \mathbf{v}_m]$. На

практике рекомендуют для формирования признаков, максимизирующих критерий разделимости, вместо \mathbf{v}_i использовать вектор нормированных коэффициентов $\mathbf{\beta}_i$, где $\mathbf{\beta}_i = [\beta_0, \beta_1 \dots \beta_p]$, $i = \overline{1, m}$, который формируется следующим образом: $\beta_0 = -\sum_{i=1}^p \beta_i \bar{x}_i$, $\mathbf{\beta}_i = \mathbf{v}_i \sqrt{n-g}$. Тогда элементы нового вектора признаков $\mathbf{y} = [y_1 y_2 \dots y_m]^T$ для i -го объекта можно определить с использованием дискриминантной функции: $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$, $i = \overline{1, m}$ [6]. Поэтому можно представить следующий алгоритм проведения дискриминантного анализа признаков (рис.1) [8]:

1. Разбиение полученных статистических данных на классы на основе формальных признаков, либо используя уже имеющуюся классификацию.
2. Расчёт признаков
3. Нахождение матриц рассеяния W и B для исходных признаков.
4. Расчёт критерия J .
5. Выбор признаков для формирования на их основе более эффективных.
6. Расчёт новых признаков.
7. Нахождение матриц рассеяния W и B для новых признаков.
8. Расчёт критерия J для новых признаков, сравнение со значением критерия для исходных признаков.
9. Проведение классификации с использованием сформированных признаков, определение ошибки классификации, сравнение со значением при использовании исходных признаков.

Отметим так же, что для оценки вклада каждого признака в значение нового признака будем использовать стандартизованные коэффициенты

$$[7]: c_i = \beta_i \sqrt{w_{ii} / (n-g)}, i = 1, \dots, p$$

Если абсолютная величина коэффициента для данной переменной мала, то эту переменную можно исключить. Если две переменных сильно коррелированы, их стандартизованные коэффициенты могут быть меньше по сравнению со случаями использования только одной из этих переменных. Для исключения влияния других переменных будем использовать матрицу внутригрупповых структурных коэффициентов: $\mathbf{S} = \mathbf{RC}$, где \mathbf{R} – корреляционная матрица элементов которой равны: $r_{ik} = \sum_{k=1}^p (w_{ik} / \sqrt{w_{ii} w_{kk}})$.

ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Был проведён ряд исследований на основе цифрового анализа изображений глазного дна, предназначенного для изучения особенностей формирования сосудистых нарушений при диа-

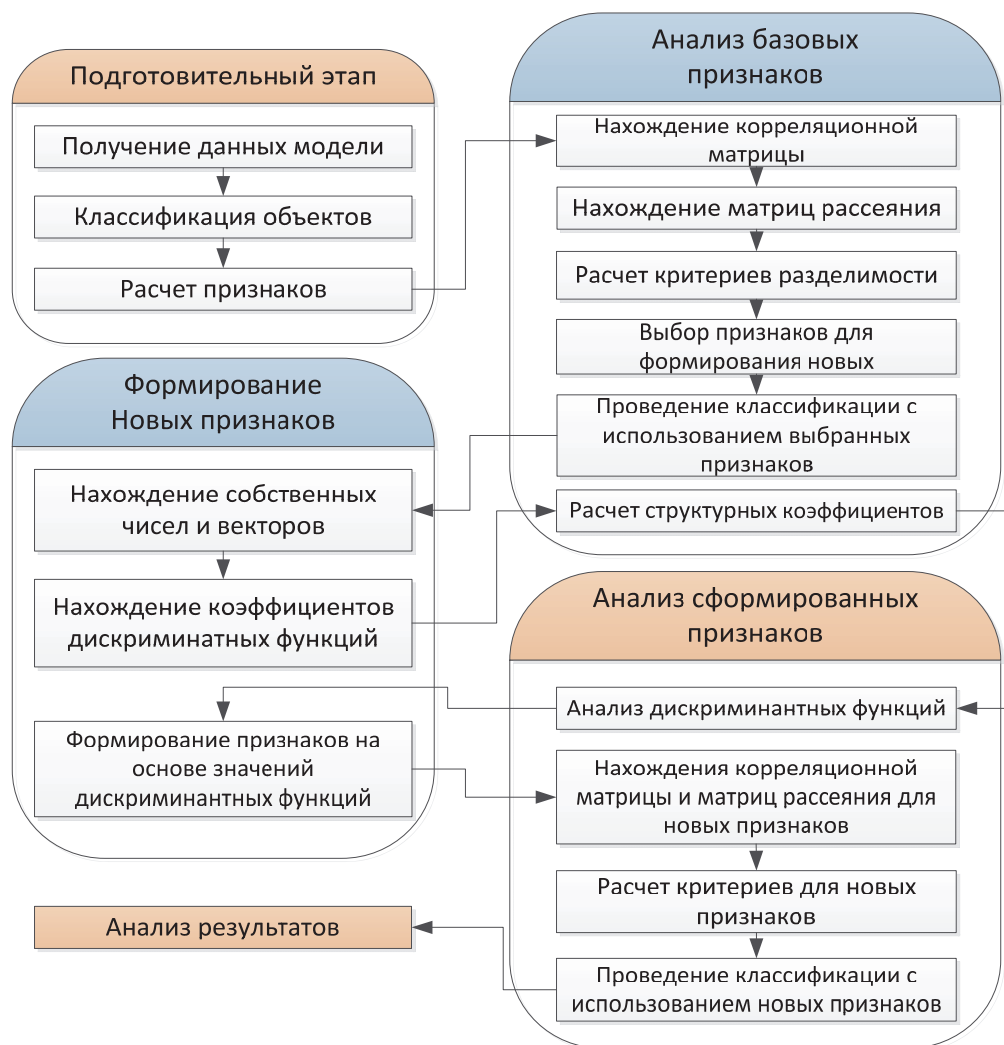


Рис. 1. Алгоритм проведения дискриминантного анализа признаков

бетической ретинопатии у 151 пациента с сахарным диабетом (СД). После обработки изображений выборка составила 8175 измерений, из них артериолы первого порядка – 1490, артериолы второго порядка – 2345, вены первого порядка – 1960, вены второго порядка – 2380. Врачи рассматривают вены и артериолы отдельно, так как в этих классах наблюдаются различные тенденции изменения сосудов при различных стадиях патологии. Поэтому анализировались 4 ГРУППЫ сосудов. На рис. 2 представлены примеры диагностических изображений глазного дна различных стадий сахарного диабета (1СД, 4СД) и схема измерений сосудистой системы, где указан принятый офтальмологами порядок исследования сосудов. Для диагностики используются только две основные группы сосудов – это сосуды первого и второго порядков.

При исследовании признаков был сделан вывод о наличии двух сильно коррелированных групп признаков. В первую группу входят признаки, описывающие параметры трасс (например, прямолинейность и извилистость трассы), а вторую группу составляют признаки, характеризу-

ющие функцию радиуса сосуда (например, извилистость радиуса и чёткообразность). На рисунке 3 представлены значения критериев разделимости одиночных признаков для каждой из четырёх ГРУПП (вены, артериолы 1 и 2 порядков).

Из рисунка видно, что наибольшими значениями критериев обладают признаки сосудов, которые не являются особо значимыми для врачей при диагностике. Кроме того, при определённой патологии для каждой группы сосудов может присутствовать сосуд, несоответствующий данной патологии (например, норма). Поэтому можно сделать вывод о том, что выборка содержит зашумлённые данные.

Для устранения различных шумов в выборке предлагается отфильтровать исходную выборку, используя метод кластеризации k -средних. Каждая ГРУППА была разделена на 5 кластеров, по количеству исходных классов внутри каждой из ГРУПП. Векторы признаков, которые не попали в нужный кластер, удалялись из выборки. Значения индивидуальных критериев разделимости для признаков внутри ГРУПП после фильтрации представлены на рис. 4.



Рис. 2. Примеры диагностических изображений глазного дна различных стадий сахарного диабета (1 СД, 4 СД), схема измерений сосудистой системы

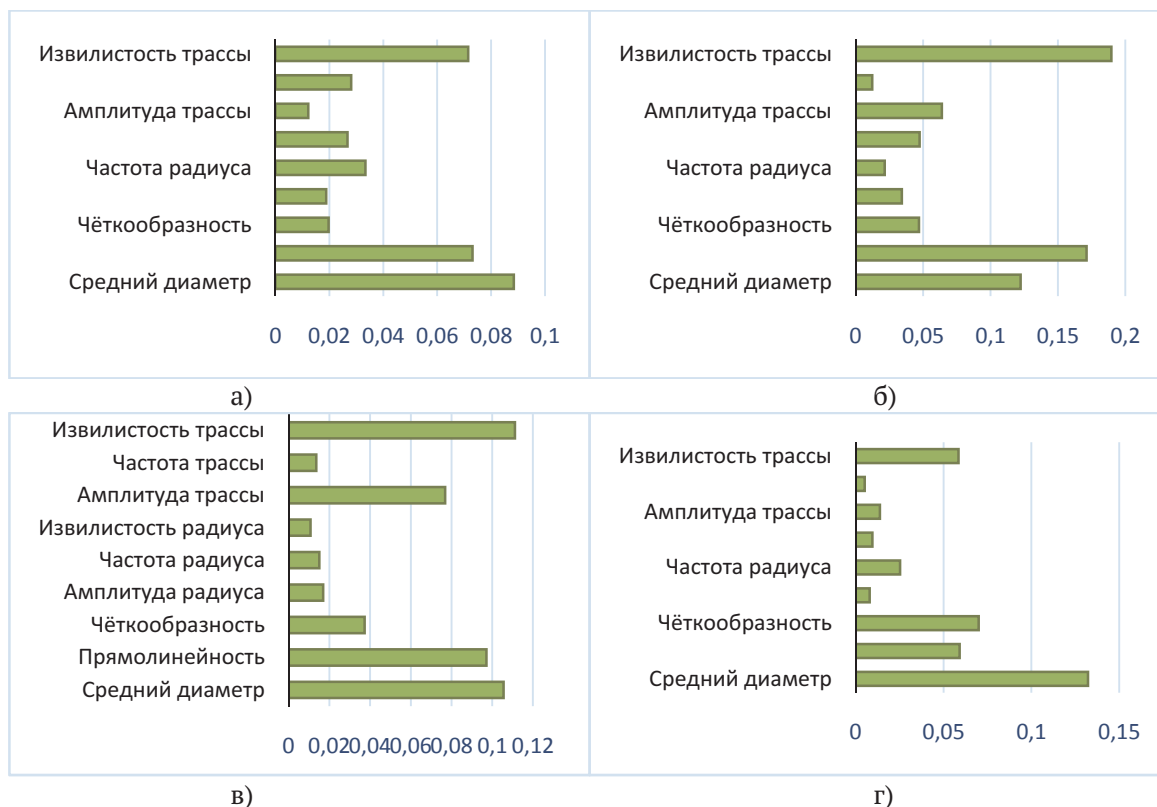


Рис. 3. Значения критерия разделимости признаков для четырёх ГРУПП сосудов до фильтрации: а), б) артериолы 1-го и 2-го порядка, в), г) вены 1-го и 2-го порядка

Анализируя рис. 3, можно сделать вывод, что в выборке, полученной после фильтрации, признаки, обладающие наибольшим критерием разделимости, являются признаками, которые врачам несут особую диагностическую значимость при визуальной диагностике патологии, что соответствует информационному письму [9].

Алгоритм дискриминантного анализа применялся для исходной и фильтрованной выборки. Для формирования новых признаков был произведён полный перебор исходных признаков для поиска комбинации новых признаков, которая максимизировала критерий разделимости. В результате был получен набор из четырёх признаков для всех ГРУПП обеих выборок.

Для полученных таким образом ГРУПП была проведена оценка ошибки классификации. Ошибка оценивалась U-методом [6]. Было сфор-

мировано две выборки: обучающая и тестовая. С использованием обучающей выборки был настроен классификатор, основанный на методе опорных векторов, с помощью которого классифицировалась тестовая выборка. Для синтеза классификатора используются только объекты обучающей выборки, которые не содержатся в тестовой выборке. Для реализации U-метода существует много возможностей, при проведении исследования для оценивания вероятности ошибки классификации использовался метод исключения одного объекта. Результаты представлено в табл. 1, где значения критерия до дискриминантного анализа приведены для комбинации четвёрок исходных признаков с наилучшими значениями критериев разделимости внутри ГРУПП.

Анализируя полученные результаты, можно сделать вывод, что фильтрация позволяет не

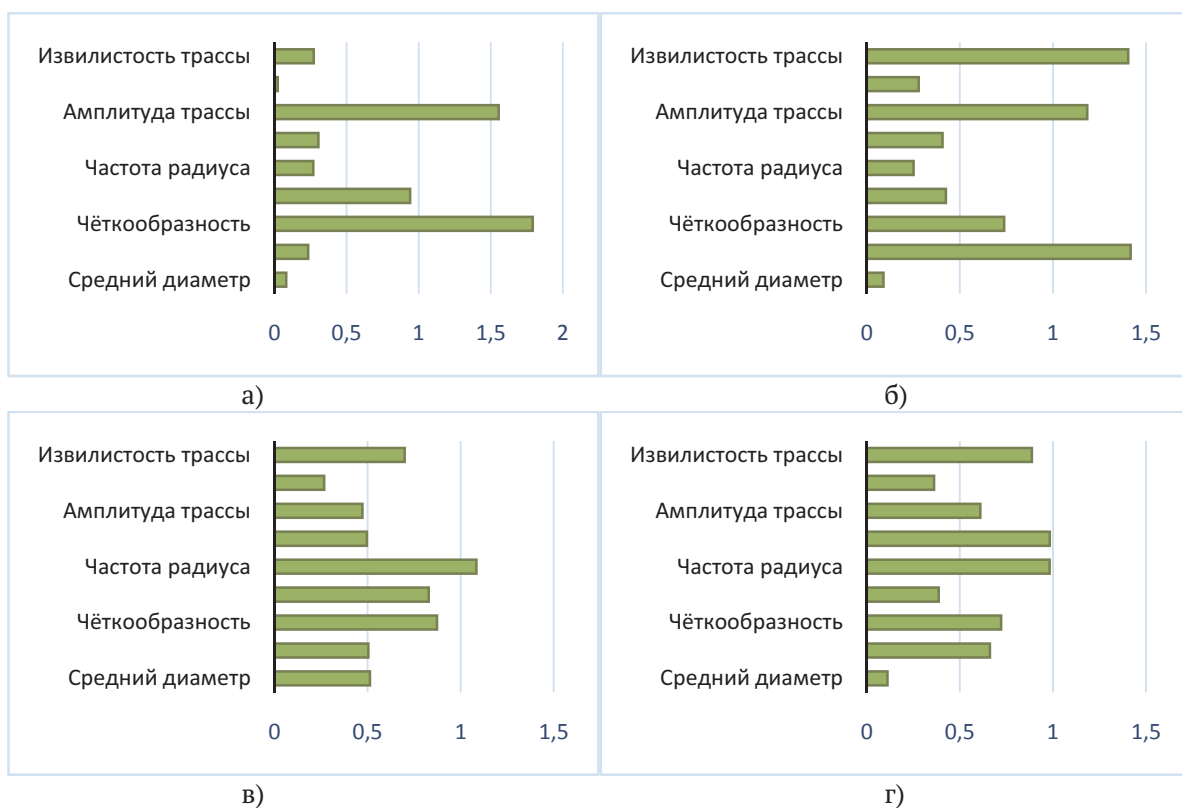


Рис. 4. Значения критерия разделимости признаков для четырёх ГРУПП сосудов после фильтрации: а), б) артериолы 1-го и 2-го порядка, в), г) вены 1-го и 2-го порядка

Таблица 1. Результаты дискриминантного анализа признакового пространства

Группа		Исходная выборка			Отфильтрованная выборка		
		J	Повышение критерия	Ошибка	J	Повышение критерия	Ошибка
Артериолы 1 порядка	до	0,21	40%	0,185	1,69	9%	0,071
	после	0,31		0,104	1,85		0,024
Артериолы 2 порядка	до	0,35	19%	0,144	2,48	6%	0,083
	после	0,42		0,090	2,62		0,028
Вены 1 порядка	до	0,29	50%	0,128	2,07	17%	0,045
	после	0,44		0,096	2,41		0,035
Вены 2 порядка	до	0,28	27%	0,162	2,21	12%	0,072
	после	0,36		0,113	2,47		0,023

только увеличить критерии разделимости признаков и уменьшить ошибку классификации, но и выявить диагностически значимые признаки. Исследования на четырёх ГРУППАХ сосудов показали, что для каждой ГРУППЫ важен свой набор диагностических признаков, что подтверждается клиническими исследованиями врачей. Результаты исследований показали, что применение предложенной методики формирования признаков позволило не только исключить недостоверные данные, но и привело к уменьшению ошибки классификации. В результате было получено значительное увеличение критерия разделимости для всех ГРУПП сосудов. Также была получена дополнительная информация по

используемым признакам, такая как их информативность, выделены связи между некоторыми признаками.

ЗАКЛЮЧЕНИЕ

Для анализа информативности и формирования более эффективных диагностических признаков изображений кровеносных сосудов была разработана технология, включающая процедуру фильтрации недостоверных данных, а также процедуру дискриминантного анализа, основанную на максимизации критерия разделимости. Был разработан алгоритм, основанный на отборе признаков, имеющих наибольшее значение критерия

разделимости, а также на полном переборе с последующим формированием новых признаков, максимизирующих данный критерий. Критерии делимости формируются с использованием матриц рассеяния между классами и внутри классов. В результате дискриминантного анализа для каждой группы сосудов были определены лучшие признаки по критерию делимости. Было показано, что в каждой из 4 ГРУПП сосудов эффективен свой набор глобальных геометрических признаков, что подтверждается клиническими исследованиями. Подсчитана ошибка классификации для каждой группы сосудов до и после работы алгоритма. Показано, что предложенная технология анализа признакового пространства по группам, включающая алгоритм фильтрации выборочных данных, алгоритм формирования пространства эффективных признаков, позволила повысить эффективность классификации сосудов по классам «норма» и различным степеням «патологии». При этом ошибка классификации была снижена до 2,3% - 3,5% для различных ГРУПП патологий.

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 14-01-00369-а, 14-07-97040-р_поволжье_a; программы № 6 фундаментальных исследований ОНИТ РАН «Биоинформатика, современные информационные технологии и математические методы в медицине» 2015 г.

СПИСОК ЛИТЕРАТУРЫ

1. *Ильясова Н.Ю.* Диагностический комплекс анализа изображений сосудов глазного дна // Биотехносфера. 2014. №3. С. 132-138.
2. Информационные технологии анализа изображений в задачах медицинской диагностики / *Н.Ю. Ильясова, А.В. Курпьянов, А.Г. Храмов*. М.: Радио и связь, 2012. 424 с.
3. *Ильясова Н.Ю.* Оценивание геометрических признаков пространственной структуры кровеносных сосудов // Компьютерная оптика. 2014. Т. 38, № 3. С. 529- 538.
4. *Ilyasova N.* Computer Systems for Geometrical Analysis of Blood Vessels Diagnostic Images // *Optical Memory and Neural Networks (Information Optics)*. 2014. Vol.23, Issue 4. P. 278-286.
5. Измерение биомеханических характеристик сосудов для ранней диагностики сосудистой патологии глазного дна / *Н.Ю. Ильясова, А.В. Курпьянов, М.А. Ананьин, Н.А. Гаврилова* // Компьютерная оптика. 2005. № 27. С. 165-170.
6. *Фукунага К.* Введение в статистическую теорию распознавания образов. М.: Наука, 1979. 270 с.
7. Факторный, дискриминантный и кластерный анализ / *Дж.-О. Ким, Ч.У. Мьюллер, У.Р. Клекка и др.*; [под ред. И.С. Енюкова]; пер. с англ. М.: Финансы и статистика, 1989. 215 с.
8. Формирование признаков для повышения качества медицинской диагностики на основе методов дискриминантного анализа / *Н.Ю. Ильясова, А.В. Курпьянов, Р.А. Парингер* // Компьютерная оптика. 2014. Т. 38, № 4. С. 751-756.
9. Информационное письмо №1 – Метод цифровой обработки изображений глазного дна / *Л.К. Мошетьова, Н.Д. Ющук, Д.И. Цыганов, В.Г. Систер, С.Л. Бранчевский, Н.Ю. Ильясова, Ю.А. Павлова*.

RESEARCH EFFECTIVENESS OF FEATURES FOR THE VASCULAR PATHOLOGIES DIAGNOSIS

© 2015 N.Yu. Ilyasova^{1,2}, R.A. Paringer^{1,2}

²Samara State Aerospace University named after Academician S.P. Korolyov
(National Research University)

²Image Processing Systems Institute, Samara

A method of filtering the source data using clustering algorithm, to improve the quality of discriminant analysis of sample data, was proposed. To improve the quality of diagnostics we propose an algorithm for the informative features formation, using methods of discriminant analysis. As integral indicators of fundus vessels, a global set of geometric features was used. It is a complete characterization for vascular disease diagnosis.

Keywords: classification of the vessels images, filtering features, k-means clustering, linear discriminant analysis, the separability criterion.

Natalya Ilyasova Candidate of Technics, Associate Professor at the Technical Cybernetics Department of SSAU, Senior Research Fellow of IPSI RAS. E-mail: ilyasova@smr.ru
Rustam Paringer, Graduate Student of SSAU, Trainee Researcher of IPSI RAS. E-mail: rusparinger@gmail.com