

## ОТБОР ПРИЗНАКОВ ДЛЯ ЗАДАЧИ ДИАГНОСТИКИ ОСТЕОПОРОЗА ПО РЕНТГЕНОВСКИМ ИЗОБРАЖЕНИЯМ ШЕЙКИ БЕДРА

А.В. Гайдель<sup>1,2</sup>, В.Р. Крашенинников<sup>3</sup>

<sup>1</sup> Самарский национальный исследовательский университет имени академика С.П. Королева, Самара, Россия,

<sup>2</sup> Институт систем обработки изображений РАН – филиал ФНИЦ «Кристаллография и фотоника» РАН, Самара, Россия,

<sup>3</sup> Ульяновский государственный технический университет, Ульяновск, Россия

### Аннотация

В работе анализируется информативность нескольких признаков текстуры рентгеновских изображений костной ткани для компьютерной диагностики остеопороза. Описываются четыре эвристических признака, также рассматривается тринадцать согласованных квадратичных признаков, описанных ранее. Решается задача выбора минимального набора из этих признаков, достаточного для линейной разделимости векторов признаков из обучающей выборки в соответствующем признаковом пространстве. В ходе вычислительных экспериментов установлено, что этому условию удовлетворяет набор из четырёх эвристических признаков, а также набор из трёх квадратичных признаков и одного из эвристических признаков.

**Ключевые слова:** текстурный анализ, отбор признаков, компьютерная диагностика, остеопороз, линейный классификатор, полиномиальные признаки.

**Цитирование:** Гайдель, А.В. Отбор признаков для задачи диагностики остеопороза по рентгеновским изображениям шейки бедра / А.В. Гайдель, В.Р. Крашенинников // Компьютерная оптика. – 2016. – Т. 40, № 6. – С. 939-946. – DOI: 10.18287/2412-6179-2016-40-6-939-946.

### Введение

Остеопороз – системное заболевание костной ткани, ранняя диагностика которого позволяет предотвратить переломы костей, ведущие к инвалидности. Традиционно для диагностики используются устройства, измеряющие минеральную плотность костной ткани, однако из-за их редкости предлагаются альтернативные методы диагностики. Один из них состоит в автоматизированном распознавании остеопороза на обычных рентгеновских снимках костной ткани, например, шейки бедра [1]. Поэтому актуальной является рассматриваемая в настоящей работе задача разработки алгоритмов такого распознавания.

Кости, поражённые остеопорозом, имеют меньшую минеральную плотность по сравнению со здоровыми костями, они более проницаемы для рентгеновских лучей. Поэтому рентгеновское изображение поражённой кости (рис. 1б) выглядит темнее, а трабекулы на нём более редкие и нитевидные по сравнению с изображением костной ткани здорового человека (рис. 1а).

Особенности задачи анализа цифровых рентгеновских изображений костной ткани подробно описаны в работе [2]. Основным результатом работы [2] является достижение вероятности ошибочной классификации 0,20 при использовании корреляционных признаков совместно с признаками Харалика. Эти результаты были улучшены за счёт использования согласованных полиномиальных признаков [3]. Оценка вероятности ошибочной классификации при таком подходе составила 0,06.

В обеих указанных выше работах производился отбор из достаточно большого количества признаков: в [2] для этого использовался алгоритм жадного добавления признаков в порядке возрастания индивидуальных значений функционала качества, а в [3] – подбор параметров у параметрического семейства признаков. Нужно отметить, что в случае небольшого конечного числа признаков можно использовать обычный перебор всех подмножеств признаков, как это сделано в [4].

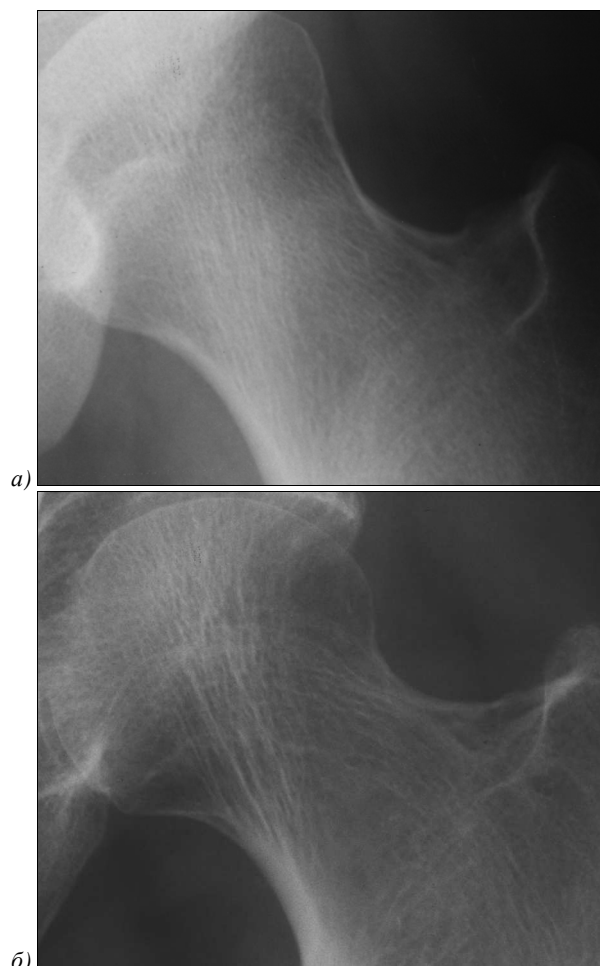


Рис. 1. Рентгеновские изображения шейки бедра: здорового человека (а), поражённой остеопорозом (б)

Некоторые морфометрические параметры изображений компьютерной томографии позвонков изучались в [5]. Полученные в этой работе коэффициенты корреляции между параметрами, оцененными по

изображениям, и параметрами гистоморфометрии трабекулярной части костной ткани составили от 0,71 до 0,76. В [6] рентгеновские изображения костной ткани анализируются с помощью фрактального параметра. Коэффициент корреляции этого параметра с радиальным количеством трабекул составил 0,47.

Существует множество известных признаков различного рода, подходящих для описания широкого класса изображений. Многие из них можно найти, например, в таких изданиях, как [7] и [8]. Однако часто для конкретных задач специально подобранные эвристические признаки могут оказаться более эффективными, чем широко распространённые универсальные признаки.

Например, в работе [9] эвристический признак, называемый анизотропией, был использован для обнаружения маркёров на изображениях фаций сыворотки крови. Другой признак, основанный на операторе Лапласа, был использован в работе [10] для совмещения изображений методом неподвижной точки. Имеются основания использовать эти признаки также и для анализа цифровых изображений костной ткани.

Таким образом, основная цель настоящей работы состоит в выборе набора из множества рассматриваемых признаков, обеспечивающего линейную разделимость выборки в соответствующем признаковом пространстве. Выбор производится из полиномиальных признаков работы [3] и четырёх эвристических признаков, описанных ниже.

### 1. Постановка задачи

Можно считать, что двумерные рентгеновские изображения представляют собой функции  $\hat{\omega}(x, y): \hat{D}_\omega \rightarrow \mathbf{R}_+$ , отображающие координату  $(x, y) \in \hat{D}_\omega \subseteq \mathbf{R}^2$  на значение интенсивности рентгеновского излучения в этой точке. Здесь под  $\mathbf{R}$  понимается множество вещественных чисел, а под  $\mathbf{R}_+ = \{x \in \mathbf{R} | x \geq 0\}$  – множество неотрицательных вещественных чисел. Однако в процессе оцифровки эти функции подвергаются процедурам дискретизации и квантования [11], так что компьютерной обработке подвергаются дискретные изображения

$$\omega(m, n) = \left\lfloor \frac{\hat{\omega}(mh_x, nh_y)}{h_\omega} + \frac{1}{2} \right\rfloor,$$

где  $h_x$  и  $h_y$  – шаги дискретизации по соответствующим координатам, а  $h_\omega$  – шаг квантования.

Дискретные изображения  $\omega(m, n): D_\omega \rightarrow \mathbf{Q}$  отображают некоторую область интереса  $D_\omega \subseteq \mathbf{Z}^2$ , включающую шейку бедра, на множество из  $Q$  уровней яркости  $\mathbf{Q} = [0; Q-1] \cap \mathbf{Z}$ . Здесь под  $\mathbf{Z}$  понимается множество целых чисел. Обозначим всё множество возможных изображений  $\Omega$ .

Эти изображения разбиты на два класса: изображения здоровой костной ткани  $\Omega_0$  и изображения костной ткани, поражённой остеопорозом,  $\Omega_1$ . Предполагается, что  $\Omega_0 \cap \Omega_1 = \emptyset$  и  $\Omega_0 \cup \Omega_1 = \Omega$ . Для опре-

делённости можно обозначить идеальный оператор распознавания  $\Phi(\omega): \Omega \rightarrow \{\Omega_0, \Omega_1\}$ . Этот оператор переводит изображение в его класс.

Решить задачу распознавания означает построить оператор  $\tilde{\Phi}(\omega): \Omega \rightarrow \{\Omega_0, \Omega_1\}$ , который также переводит изображение в его класс, но при этом не владеет информацией обо всех возможных изображениях. На практике обычно для построения этого оператора можно использовать лишь конечную обучающую выборку  $U \subseteq \Omega$ , для каждого изображения из которой известен его класс.

В действительности оператор  $\tilde{\Phi}(\omega)$  обычно строится как суперпозиция

$$\tilde{\Phi}(\omega) = C(\Psi(\omega)),$$

где  $\Psi(\omega): \Omega \rightarrow \mathbf{E}$  – оператор вычисления признаков, который переводит изображение  $\omega(m, n)$  в вектор признаков  $x \in \mathbf{E}$ , а  $C(x): \mathbf{E} \rightarrow \{\Omega_0, \Omega_1\}$  – классификатор, переводящий вектор признаков в его класс. Множество всех возможных векторов признаков  $\mathbf{E}$  называется признаковым пространством. На практике признаки обычно числовые, так что  $\mathbf{E} \subseteq \mathbf{R}^K$ , где  $K$  – количество признаков.

Правильный выбор оператора вычисления признаков  $\Psi(\omega)$ , а значит, и признакового пространства  $\mathbf{E}$  может существенно повысить качество распознавания. В настоящей работе критерием качества признакового пространства считается линейная разделимость векторов признаков  $\Psi(\omega)$  для изображений из обучающей выборки. Множество векторов  $\Psi(\omega)$  для  $\omega \in U$  называется линейно разделимым в пространстве  $\mathbf{E} \subseteq \mathbf{R}^K$ , если в этом пространстве существует гиперплоскость, такая что векторы признаков из разных классов лежат по разные стороны от неё, то есть

$$\begin{aligned} \exists w \in \mathbf{R}^K \exists w_0 \in \mathbf{R} \forall \omega_1, \omega_2 \in U : \Phi(\omega_1) = \Phi(\omega_2) &\Leftrightarrow \\ \Leftrightarrow (w^T \Psi(\omega_1) + w_0)(w^T \Psi(\omega_2) + w_0) > 0. \end{aligned}$$

В случае линейной разделимости вообще всех возможных векторов признаков даже простейшие линейные классификаторы могут безошибочно выполнять распознавание всех возможных изображений.

В настоящей работе рассматривается множество из нескольких признаков. Задачей является определение небольшого набора из них, достаточного для эффективной классификации имеющихся рентгеновских изображений.

### 2. Эвристические признаки

#### Анизотропия

На изображении поражённой кости трабекулы располагаются реже, чем на изображении здоровой кости (рис. 1). Около них окружающий фон темнее. Поэтому изображение больной кости более анизотропно: яркость в перпендикулярном к трабекуле направлении изменяется значительно быстрее, чем у здоровой кости.

В качестве показателя анизотропии по направлению орта  $e_1 = (\cos \alpha, \sin \alpha)$  принимается максимальное из от-

ношений средних модулей производных по этому орту и перпендикулярному ему орту  $e_2 = (-\sin \alpha, \cos \alpha)$ :

$$A(\alpha) = \max(G_1(\alpha)/G_2(\alpha), G_2(\alpha)/G_1(\alpha)), \quad (1)$$

где

$$G_k(\alpha) = (1/|D_\omega|) \sum_{(m,n) \in D_\omega} |\partial \omega(m,n) / \partial e_k| = \\ = (1/|D_\omega|) \sum_{(m,n) \in D_\omega} |\nabla \omega(m,n) \cdot e_k|.$$

Здесь  $|D_\omega|$  – количество элементов в конечном множестве  $D_\omega$ , « $\cdot$ » – знак скалярного произведения. Заменяя производные по направлениям конечными разностями и вводя угол  $\alpha$  между ортом  $e_1$  и осью абсцисс, получаем

$$G_1(\alpha) = \frac{1}{|D_\omega|} \sum_{(m,n) \in D_\omega} |(\omega(m+1,n) - \omega(m,n)) \cos \alpha + \\ + (\omega(m,n+1) - \omega(m,n)) \sin \alpha|, \\ G_2(\alpha) = \frac{1}{|D_\omega|} \sum_{(m,n) \in D_\omega} |-(\omega(m+1,n) - \omega(m,n)) \sin \alpha + \\ + (\omega(m,n+1) - \omega(m,n)) \cos \alpha|.$$

Анизотропия (1) зависит от угла  $\alpha$ , поэтому показателем анизотропии изображения выбирается наибольшее её значение

$$A = \sup_{\alpha \in [0; \pi]} A(\alpha). \quad (2)$$

Можно ожидать, что у больных костей анизотропия (2) больше, чем у здоровых костей. В работе [9] этот признак был использован для обнаружения маркёров на изображениях фаций сыворотки крови.

#### Области постоянной яркости

На больной кости имеется много поражённых участков, поэтому яркость изображения быстро изменяется, а области постоянства яркости значительно мельче, чем на изображении здоровой кости (рис. 1). В качестве характеристики этой особенности изображений принимается средняя площадь (количество отсчётов) кластеров постоянной яркости:

$$P = |D_\omega| / \kappa(\omega), \quad (3)$$

где  $\kappa(\omega)$  – количество связанных областей постоянной яркости на изображении  $\omega(m,n)$ . Можно предположить, что значение признака (3) у поражённых остеопорозом костей меньше.

#### Вариация

Визуально заметно, что яркость на изображении поражённой остеопорозом кости варьируется в целом сильнее, чем на изображении здоровой кости (рис. 1). Это качество можно численно оценить средней вариацией

$$V_T = (1/|D_\omega|) \sum_{(m,n) \in D_\omega} |\nabla \omega(m,n)| = \\ = (1/|D_\omega|) \sum_{(m,n) \in D_\omega} \sqrt{\nabla \omega(m,n) \cdot \nabla \omega(m,n)}, \quad (4)$$

где

$$\nabla \omega(m,n) \cdot \nabla \omega(m,n) = (\omega(m+1,n) - \omega(m,n))^2 + \\ + (\omega(m,n+1) - \omega(m,n))^2.$$

Величина (4) пропорциональна средней яркости изображения, что нежелательно, так как эта средняя яркость зависит от длительности экспозиции и других условий получения рентгенограмм. Поэтому предлагается нормировать (4) величиной средней яркости:

$$V = V_T / \bar{\omega}, \quad (5)$$

где

$$\bar{\omega} = (1/|D_\omega|) \sum_{(m,n) \in D_\omega} \omega(m,n) - \text{средняя яркость.}$$

Следует ожидать, что на изображениях поражённых остеопорозом костей значение признака (5) больше, чем на изображениях здоровых костей.

#### Лапласиан

Оператор Лапласа

$$\Delta \hat{\omega}(x,y) = \nabla^2 \hat{\omega}(x,y) = \frac{\partial^2 \hat{\omega}(x,y)}{\partial x^2} + \frac{\partial^2 \hat{\omega}(x,y)}{\partial y^2}$$

равен нулю для гармонических функций, то есть он отражает меру отличия свойств функции от свойств гармонических функций. Уравнению Лапласа  $\Delta \hat{\omega}(x,y) = 0$  соответствует, например, стационарное распределение температур пластины. Отличие лапласиана от нуля отражает наличие динамики. Можно предположить, что развитие патологического процесса в костях приводит к отличию лапласиана рентгенограммы от нуля.

Рассмотрим характеристику изображения

$$L = (1/\bar{\omega}|D_\omega|) \sum_{(m,n) \in D_\omega} |\Delta \omega(m,n)| = \\ = 1/\bar{\omega}|D_\omega| \sum_{(m,n) \in D_\omega} |\omega_m^{(2)}(m,n) + \omega_n^{(2)}(m,n)|, \quad (6)$$

то есть средний модуль значений оператора Лапласа, нормированный на среднюю яркость, где вторые производные заменены конечными разностями:

$$\omega_m^{(2)}(m,n) = \omega(m-1,n) - 2\omega(m,n) + \omega(m+1,n),$$

$$\omega_n^{(2)}(m,n) = \omega(m,n-1) - 2\omega(m,n) + \omega(m,n+1).$$

Можно предположить, что значение признака (6) выше для изображений поражённых остеопорозом костей, чем для изображений здоровых костей. Подробное описание результатов использования похожего признака для совмещения изображений методом неподвижной точки можно найти в работах [10, 12].

### **3. Экспериментальные исследования**

#### Материалы исследования

Материалом для исследования послужила выборка из 30 реальных рентгеновских изображений шейки бедра, полученных в ходе клинического обследования

ния пациентов в клиниках Самарского государственного медицинского университета. Из них на 19 изображениях представлены кости, поражённые остеопорозом. Изображения были оцифрованы с помощью транспарентного сканера в разрешении 400 точек на дюйм, то есть примерно 16 линий на миллиметр. Количество уровней яркости  $Q = 256$ . Для каждого изображения из выборки квалифицированными медицинскими работниками на основании множества факторов было принято решение о наличии либо отсутствии остеопороза. Средний размер изображений –  $1040 \times 860$  отсчётов.

На каждом изображении вручную была выделена область интереса, включающая только шейку бедра. Для каждого изображения были вычислены значения четырёх признаков, описанных в предыдущем разделе, а также тринадцати квадратичных признаков, описанных в работе [3]. Квадратичные признаки были согласованы с обучающей выборкой из работы [3], а не с выборкой, на которой проводится исследование в этой работе. В дальнейшем для различных подмножеств признаков определялось, существует ли в соответствующем признаковом пространстве гиперплоскость, безошибочно разделяющая векторы признаков из различных классов. Наличие такой гиперплоскости само по себе свидетельствует о качестве соответствующего признакового пространства.

**Индивидуальная разделяющая способность признаков**

В первую очередь была изучена индивидуальная разделяющая способность каждого эвристического признака. Линейный классификатор в одномерном случае представляет собой обычный порог, такой что решения об отнесении объекта к тому или иному классу принимаются в зависимости от того, по какую сторону от этого порога находится значение признака. Значение порога для каждого признака было определено простым подбором с целью минимизировать количество ошибок при классификации исходной выборки соответствующим пороговым классификатором.

В табл. 1 приведены средние значения и среднеквадратические отклонения четырёх приведённых в настоящей статье признаков для каждого класса изображений из исходной выборки. Видно, что соотношения о связи между классами и значениями этих признаков подтверждаются.

Табл. 1. Средние значения и среднеквадратические отклонения признаков для каждого класса

Признак	Без патологий		Остеопороз	
	Среднее	СКО	Среднее	СКО
A	1,06	0,06	1,15	0,09
P	6,59	1,23	4,63	1,23
V	0,04	0,01	0,06	0,04
L	0,06	0,01	0,11	0,06

В табл. 2 приведены результаты исследования индивидуальной разделяющей способности четырёх описанных во втором параграфе эвристических признаков. Для каждого признака приведено значение порога, обеспечивающего наилучшую разделяемость векторов из обучающей выборки, наименьшее число

объектов, оказавшихся не по ту сторону от порога, а также доля от общего числа объектов обучающей выборки, классифицированных правильно. Видно, что наилучшие индивидуальные результаты показывает признак, основанный на операторе Лапласа, но ни один из признаков сам по себе не обеспечивает линейную разделяемость обучающей выборки в соответствующем одномерном признаковом пространстве. Этого и следовало ожидать, учитывая соотношения между средними значениями и СКО из табл. 1.

Табл. 2. Индивидуальная разделяющая способность эвристических признаков

Признак	Порог	Число ошибок	Достоверность
A	1,100	7	0,77
P	6,000	5	0,83
V	0,034	6	0,80
L	0,064	4	0,87

**Исследование пар признаков**

Далее исследовались признаковые пространства, образованные различными парами из этих четырёх эвристических признаков. Например, интерес представляет распределение векторов признаков из обучающей выборки в двумерном признаковом пространстве, образованном признаками A и P. Как можно судить из рис. 2, в таком пространстве существует прямая, разделяющая векторы признаков из обучающей выборки, так что векторы из разных классов лежат по разные стороны от этой прямой, и при этом совершается лишь 4 ошибки: по две в каждую сторону. Эта прямая имеет уравнение  $P = 10F - 4,5$  и была выбрана эвристически.

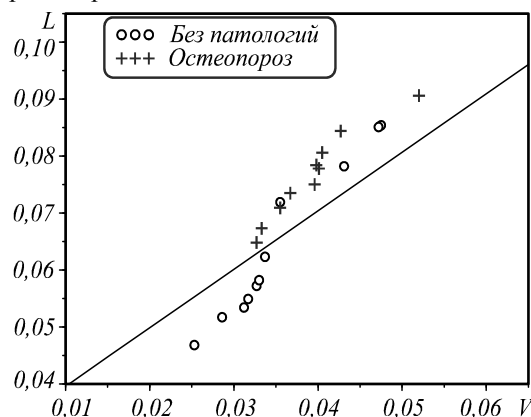


Рис. 2. Выборка в пространстве признаков, основанных на анизотропии и областях постоянной яркости

На рис. 3 показано распределение векторов признаков из обучающей выборки в двумерном признаковом пространстве, образованном признаками V и L. Как видно, в этом пространстве можно провести прямую  $L = 1,027A + 0,293$ , отличающуюся тем, что она разделяет векторы признаков из разных классов так, что всего четыре изображения здоровой костной ткани оказываются по ту же сторону от этой прямой, что и изображения костной ткани, поражённой остеопорозом. Нужно отметить, что с таким же успехом можно использовать один только признак L (это видно и из рис. 3, и из табл. 2).

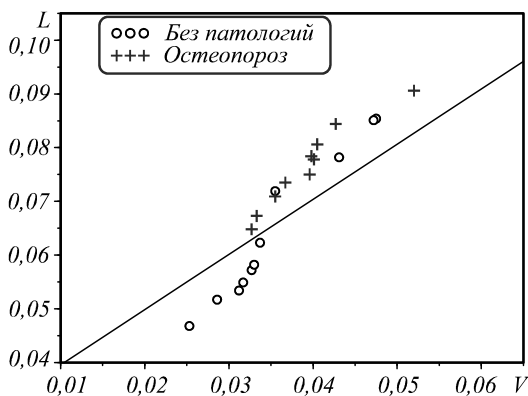


Рис. 3. Выборка в пространстве признаков, основанных на вариации и операторе Лапласа

Линейный классификатор по комбинациям признаков с обучающей выборкой

Не слишком корректно оценивать количество ошибок классификации на той же выборке, по которой строилась разделяющая поверхность, поэтому для дальнейшего исследования было решено разбить исходную выборку на обучающую и контрольную. В обучающую выборку были случайным образом включены шесть изображений здоровых костей и девять изображений костей, поражённых остеопорозом, остальные изображения составили контрольную выборку.

Для диагностики был использован линейный классификатор, построенный следующим образом. По обучающей выборке определялись выборочные средние значения векторов каждого класса, а гиперплоскость проводилась перпендикулярно прямой, соединяющей эти средние, через точку на этой линии, обеспечивающую классификацию с наименьшим количеством ошибок на обучающей выборке. Далее обученный классификатор применялся к контрольной выборке. Перебирались все 15 непустых комбинаций из четырёх эвристических признаков.

В табл. 3 приведены результаты исследования эффективности всех различных непустых комбинаций эвристических признаков с использованием обучающей и контрольной выборок. Наилучшей комбинацией признаков при таком подходе можно считать пару признаков A и V, которые обеспечивают наименьшее количество ошибок на контрольной выборке. Однако оценивать вероятность ошибочной классификации по такой небольшой выборке было бы не вполне корректно. Кроме того, изменение состава обучающей выборки и контрольной выборки существенно влияло на построенный классификатор и на количество ошибок. Так или иначе, средняя доля неверно распознанных изображений по всем комбинациям признаков составила 0,15.

Исследование признакового пространства, включающего все четыре признака

Далее в ходе исследования предпринимались попытки выяснить, существует ли в четырёхмерном пространстве из четырёх эвристических признаков, описанных во втором параграфе, гиперплоскость, безошибочно разделяющая векторы признаков из разных классов для всей исходной выборки.

Табл. 3. Результаты исследования эффективности комбинаций эвристических признаков

Признаки	Число ошибок на обучающей выборке	Число ошибок на контрольной выборке
A	2	3
P	2	2
V	4	2
L	2	3
A, P	0	2
A, V	1	1
A, L	2	3
P, V	3	2
P, L	2	3
V, L	2	2
A, P, V	0	2
A, P, L	1	3
A, V, L	2	3
P, V, L	2	2
A, P, V, L	2	2
<b>В среднем</b>	<b>1,8</b>	<b>2,3</b>

Искомая гиперплоскость имеет вид

$$T = c_1A + c_2P + c_3 V + c_4 L = c \cdot x, \tag{7}$$

где  $T \in \mathbf{R}$  – искомое пороговое значение,  $c \in \mathbf{R}^4$  – вектор искомых коэффициентов перед соответствующими значениями признаков,  $x \in \mathbf{E}$  – вектор признаков. В качестве векторов весов  $c$  предлагается перебирать орты

$$\begin{aligned} c_1 &= \cos \alpha, \\ c_2 &= \sin \alpha \cos \beta, \\ c_3 &= \sin \alpha \sin \beta \cos \gamma, \\ c_4 &= \sin \alpha \sin \beta \sin \gamma. \end{aligned}$$

Углы  $\alpha, \beta$  и  $\gamma$  определяют направление орта, который является нормалью к разделяющей гиперплоскости в пространстве признаков. В этом случае значение  $T$  из формулы (7) является проекцией вектора признаков  $x$  на нормаль  $c$ .

Углы  $\alpha, \beta$  и  $\gamma$  перебирались с шагом  $h = \pi / 20$ , причём для  $\alpha$  перебирались значения от  $-10h$  до  $9h$ , а  $\beta$  и  $\gamma$  – от  $0h$  до  $39h$ . Таким образом, общее количество комбинаций составило 3200. Для каждого вектора из обучающей выборки вычислялось значение  $T$  по формуле (7) и перебиралось значение порога, находящегося между очередными значениями  $T$ , как в одномерном случае.

Оптимальным оказался весовой вектор

$$\hat{c} = (0,309; -0,149; -0,760; 0,552),$$

который обеспечивает безошибочное разделение всех векторов признаков из исходной выборки, такое что векторы признаков из разных классов находятся по разные стороны от соответствующей гиперплоскости.

Исследование эвристических признаков совместно с согласованными квадратичными признаками

Наконец, была исследована эффективность совместного использования четырёх эвристических признаков и 13 квадратичных признаков, согласованных

с обучающей выборкой из работы [3], то есть всего 17 признаков. В процессе исследования перебирались все  $(2^{17} - 1)$  непустых комбинаций из этих признаков, и для каждого признакового пространства определялось, существует ли в нём гиперплоскость, безошибочно разделяющая векторы признаков из исходной выборки, состоящей из всех 30 изображений. Для построения разделяющей гиперплоскости использовался классический метод опорных векторов, описанный ещё в [13].

На рис. 4 для примера показано распределение векторов признаков из обучающей выборки в двумерном признаковом пространстве, образованном согласованным квадратичным признаком, обозначенным R, и признаком P, основанным на областях постоянной яркости. Видно, что в этом пространстве можно провести прямую, разделяющую векторы признаков из исходной выборки, так что только четыре из них оказываются не по ту сторону от этой прямой: по два вектора из каждого класса. По качеству это пространство похоже на пространство из признаков A и P, для которого распределение исходной выборки показано на рис. 2. Конечно, в этом пространстве исходная выборка не является линейно разделяемой, но при использовании ещё двух других согласованных квадратичных признаков совместно с этими двумя можно получить четырёхмерное признаковое пространство, в котором векторы признаков из исходной выборки являются линейно разделяемыми.

В результате исследования были установлены следующие факты.

1. Существуют четырёхмерные признаковые пространства, в которых векторы признаков из исходной выборки являются линейно разделяемыми.

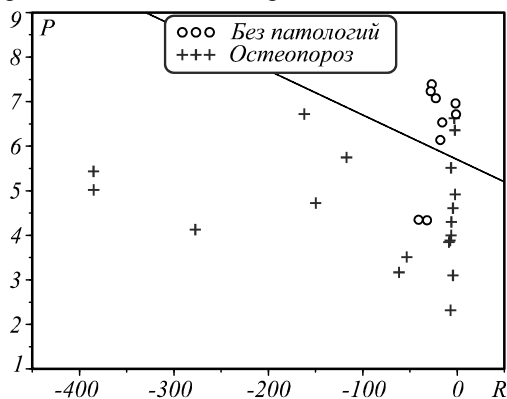


Рис. 4. Выборка в пространстве из согласованного квадратичного признака и признака, основанного на областях постоянной яркости

2. Не существует никаких признаков пространств меньшей размерности, составленных из рассмотренных признаков, в которых векторы признаков из исходной выборки являются линейно разделяемыми.

3. Векторы признаков из исходной выборки являются линейно разделяемыми в пространстве из четырёх эвристических признаков, описанных в первом параграфе настоящей работы.

4. Векторы признаков из исходной выборки являются линейно разделяемыми в пространстве из трёх согласованных квадратичных признаков и признака P, основанного на областях постоянной яркости.

5. Ни в каком признаковом пространстве, состоящем только из рассмотренных согласованных квадратичных признаков, векторы признаков из обучающей выборки не являются линейно разделяемыми.

### Заключение

В работе исследовались эвристические признаки, из которых два уже использовались ранее для других задач, а два других предложены впервые. Рассматривались различные признаковые пространства, образованные предложенными эвристическими признаками совместно с согласованными квадратичными признаками, и для каждого признакового пространства выяснялось, существует ли гиперплоскость, разделяющая векторы из обучающей выборки. Для построения гиперплоскости использовались как эвристические приёмы в простых случаях, так и метод опорных векторов в более общих случаях.

В результате было установлено, что в признаковом пространстве из четырёх эвристических признаков исходная выборка является линейно разделяемой, то есть векторы признаков разных классов из исходной выборки находятся по разные стороны от разделяющей гиперплоскости. Этим же свойством отличается признаковое пространство, состоящее из трёх согласованных квадратичных признаков и эвристического признака, основанного на доле областей постоянной яркости. При этом ни в каком пространстве из трёх рассмотренных признаков исходная выборка уже не является линейно разделяемой. То же относится и к любому пространству, образованному только согласованными квадратичными признаками. Если считать разделяющую гиперплоскость линейным классификатором и предположить, что безошибочная разделяемость масштабируется на выборки больших объёмов, то можно заключить, что полученные результаты качества классификации превосходят соответствующие результаты, полученные в работах [2] и [3].

В целом проведённое исследование показывает, что рассмотренные признаки являются достаточно информативными, и из них можно выбрать небольшой набор, обеспечивающий достаточно эффективную диагностику остеопороза по рентгеновским изображениям костных тканей.

### Благодарности

Работа выполнена при поддержке грантов РФФИ 14-07-97040-р\_поволжье\_a, 16-41-630761 р\_a и 16-41-732041, а также Министерства образования и науки РФ в рамках мероприятий Программы повышения конкурентоспособности Самарского университета среди ведущих мировых научно-образовательных центров на 2013-2020 годы и Программы фундаментальных исследований ОНИТ РАН «Биоинформатика, современные информационные технологии и математические методы в медицине».

**Литература**

1. **Рейнберг, С.А.** Рентгенодиагностика заболеваний костей и суставов / С.А. Рейнберг. – М.: Медицина, 1964. – Т. 1. – 530 с.
2. **Гайдель, А.В.** Исследование текстурных признаков для диагностики заболеваний костной ткани по рентгеновским изображениям / А.В. Гайдель, С.С. Первушкин // Компьютерная оптика. – 2013. – Т. 37, № 1. – С. 113-119.
3. **Гайдель, А.В.** Согласованные полиномиальные признаки для анализа полутоновых биомедицинских изображений / А.В. Гайдель // Компьютерная оптика. – 2016. – Т. 40, № 2. – С. 232-240. – DOI: 10.18287/2412-6179-2016-40-2-232-239.
4. **Ильяслова, Н.Ю.** Формирование признаков для повышения качества медицинской диагностики на основе методов дискриминантного анализа / Н.Ю. Ильяслова, А.В. Куприянов, Р.А. Парингер // Компьютерная оптика. – 2014. – Т. 38, № 4. – С. 851-855.
5. **Килина, О.Ю.** Оценка микроархитектоники костной ткани путем цифрового анализа компьютерных томограмм для диагностики остеопороза / О.Ю. Килина, В.Д. Завадовская, Р.В. Данильчук, Е.М. Третьяков, О.В. Родионова, О.В. Баранова // Бюллетень сибирской медицины. – 2003. – № 2. – С. 94-100.
6. **Bacchetta, J.** Assessment of bone microarchitecture in chronic kidney disease: A comparison of 2D bone texture analysis and high-resolution peripheral quantitative computed tomography at the radius and tibia / J. Bacchetta, S. Boutroy, N. Vilayphiou, A. Fouque-Aubert, P.D. Delmas, E. Lespessailles, D. Fouque, R. Chapurlat // Calcified Tissue International. – 2010. – Vol. 87(5). – P. 385-391. – DOI: 10.1007/s00223-010-9402-z.
7. **Визильтер, Ю.В.** Обработка и анализ изображений в задачах машинного зрения / Ю.В. Визильтер, С.Ю. Желтов, А.В. Бондаренко, М.В. Ососков, А.В. Моржин. – М.: Физматкнига, 2010. – 672 с. – ISBN: 978-5-89155-201-2.
8. **Ильяслова, Н.Ю.** Информационные технологии анализа изображений в задачах медицинской диагностики / Н.Ю. Ильяслова, А.В. Куприянов, А.Г. Храмов. – М.: Радио и связь, 2012. – 424 с. – ISBN: 5-89776-014-4.
9. **Krasheninnikov, V.R.** Identification of pectinate structures in images of blood serum facia / V.R. Krasheninnikov, A.S. Kopylova // Pattern Recognition and Image Analysis. – 2011. – Vol. 21(3). – P. 508-510. – DOI: 10.1134/S1054661811020623.
10. **Krasheninnikov, V.R.** Estimating parameters of interframe geometric transformation of an image sequence by the fixed point method / V.R. Krasheninnikov, M.A. Potapov // Pattern Recognition and Image Analysis. – 2010. – Vol. 20(3). – P. 316-323. – DOI: 10.1134/S1054661810030077.
11. Методы компьютерной обработки изображений / М.В. Гашников, Н.И. Глумов, Н.Ю. Ильяслова, В.В. Мясников, С.Б. Попов, В.В. Сергеев, В.А. Соيفер, А.Г. Храмов, А.В. Чернов, В.М. Чернов, М.А. Чичёва, В.А. Фурсов; под ред. В.А. Соифера. – 2-е изд., испр. – М.: Физматлит, 2003. – 784 с. – ISBN: 5-9221-0270-2.
12. **Васильев, К.К.** Статистический анализ изображений / К.К. Васильев, В.Р. Крашенинников. – Ульяновск: УлГТУ, 2014. – 214 с. – ISBN: 5-8946-234-6.
13. **Вапник, В.Н.** Теория распознавания образов (статистические проблемы обучения) / В.Н. Вапник, А.Я. Червоненкис. – М.: Наука, 1974. – 416 с.

**Сведения об авторах**

**Гайдель Андрей Викторович**, 1989 года рождения, кандидат технических наук, в 2012 году окончил Самарский государственный аэрокосмический университет имени академика С.П. Королёва по направлению «Прикладная математика и информатика», в 2015 году защитил кандидатскую диссертацию. Работает ассистентом кафедры технической кибернетики Самарского университета и стажёром-исследователем в Институте систем обработки изображений РАН – филиале ФНИЦ «Кристаллография и фотоника» РАН. Область научных интересов: компьютерная обработка изображений, теория распознавания образов, интеллектуальный анализ данных, теория алгоритмов. E-mail: [andrey.gaidel@gmail.com](mailto:andrey.gaidel@gmail.com).

**Крашенинников Виктор Ростиславович**, 1945 года рождения, в 1967 году окончил Казанский государственный университет по специальности «Вычислительная математика», доктор технических наук (1996), профессор, заведующий кафедрой прикладной математики и информатики Ульяновского государственного технического университета, имеет более 300 научных публикаций по анализу случайных процессов и изображений. E-mail: [kvrulstu@mail.ru](mailto:kvrulstu@mail.ru).

ГРПТИ 28.23.15

Поступила в редакцию 29 августа 2016 г. Окончательный вариант – 1 ноября 2016 г.

**FEATURE SELECTION FOR DIAGNOZING THE OSTEOPOROSIS BY FEMORAL NECK X-RAY IMAGES**A.V. Gaidel<sup>1,2</sup>, V.R. Krasheninnikov<sup>3</sup><sup>1</sup> Samara National Research University, Samara, Russia,<sup>2</sup> Image Processing Systems Institute of RAS, – Branch of the FSRC “Crystallography and Photonics” RAS, Samara, Russia,<sup>3</sup> Ulyanovsk State Technical University, Ulyanovsk, Russia**Abstract**

We analyzed the quality of a number of features describing the texture of digital X-ray images of the bone tissue for the computer-aided diagnosis of the osteoporosis. We introduced four heuristic features, also considering thirteen adjusted quadratic features described in a previous paper. We solved a problem of selecting the smallest feature subset in order to provide the linear separability

of the feature vectors from the learning sample in the corresponding feature space. During the experimental studies we found that the subset of four heuristic features fulfils the separability condition as well as the subset including three quadratic features and one heuristic feature does.

**Keywords:** texture analysis, feature selection, computer-aided diagnosis, osteoporosis, linear classifier, polynomial features.

**Citation:** Gaidel AV, Krasheninnikov VR. Feature selection for diagnosing the osteoporosis by femoral neck X-ray images. *Computer Optics* 2016; 40(6): 939-946. DOI: 10.18287/2412-6179-2016-40-6-939-946.

**Acknowledgements:** The work was partially funded by the Russian Foundation of Basic Research (grants 14-07-97040-p\_поволжье\_a,16-41-630761 p\_a and 16-41-732041), the Russian Federation Ministry of Education and Science as part of Samara University's competitiveness enhancement program in 2013-2020 and the RAS basic research program "Bio-informatics, modern information technologies and mathematical methods in medicine".

### References

- [1] Reinberg SA. X-ray diagnosis of diseases of bones and joints [In Russian]. Moscow: "Medicine" Publisher; 1964.
- [2] Gaidel AV, Pervushkin SS. Research of the textural features for the bony tissue diseases diagnostics using the roentgenograms. *Computer Optics* 2013; 37(1): 113-119.
- [3] Gaidel AV. Matched polynomial features for the analysis of grayscale biomedical images. *Computer Optics* 2016; 40(2): 232-240. DOI: 10.18287/2412-6179-2016-40-2-232-239.
- [4] Ilyasova NYu, Kupriyanov AV, Paringer RA. Formation of features for improving the quality of medical diagnosis based on discriminant analysis methods. *Computer Optics* 2014; 38(4): 851-855.
- [5] Kilina OYu, Zavadovskaya VD, Danilchuk RV, Tretyakov YeM, Rodionova OV, Baranova OV. Assessment of bone tissue architectonics with digital analysis of computed tomograms for osteoporosis diagnostics [In Russian]. *Bulletin of Siberian Medicine* 2003; 2: 94-100.
- [6] Bacchetta J, Boutroy S, Vilayphiou N, Fouque-Aubert A, Delmas PD, Lespessailles E, Fouque D, Chapurlat K. Assessment of bone microarchitecture in chronic kidney disease: A comparison of 2D bone texture analysis and high-resolution peripheral quantitative computed tomography at the radius and tibia. *Calcif Tissue Int* 2010; 87(5): 385-391. – DOI: 10.1007/s00223-010-9402-z.
- [7] Vizilter YuV, Zheltov SYu, Bondarenko AV, Ososkov MV, Morzhin AV. Image processing and analysis in machine vision problems [In Russian]. Moscow: "Fizmatkniga" Publisher; 2010. ISBN: 978-5-89155-201-2.
- [8] Ilyasova NYu, Kupriyanov AV, Khramov AG. Information technologies of the image analysis in the medical diagnosis problems [In Russian]. Moscow: "Radio and Svyaz" Publisher; 2012. ISBN: 5-89776-014-4.
- [9] Krasheninnikov VR, Kopylova AS. Identification of pectinate structures in images of blood serum facia. *Pattern Recognition and Image Analysis* 2011; 21(3): 508-510. DOI: 10.1134/S1054661811020623.
- [10] Krasheninnikov VR, Potapov MA. Estimating parameters of interframe geometric transformation of an image sequence by the fixed point method. *Pattern Recognition and Image Analysis* 2010; 20(3): 316-23. DOI: 10.1134/S1054661810030077.
- [11] Soifer VA, ed, Gashnikov MV, Glumov NI, Ilyasova NYu, Myasnikov VV, Popov SB, Sergeev VV, Khramov AG, Chernov AV, Chernov VM, Chicheva MA, Fursov VA. Methods for computer image processing [In Russian]. Moscow: "Fizmatlit"; 2003. ISBN: 5-9221-0270-2.
- [12] Vassiliev KK, Krasheninnikov VR. Statistical analysis of images [In Russian]. Ulyanovsk: "UISTU" Publisher; 2014. ISBN: 5-8946-234-6.
- [13] Vapnik VN, Chervonenkis AY. Pattern recognition theory (statistical problems of learning) [In Russian]. Moscow: "Nauka" Publisher; 1974.

### Authors' information

**Andrey Viktorovich Gaidel** (b. 1989) graduated from Samara State Aerospace University in 2012, majoring in Applied Mathematics and Informatics. He received his Candidate of Science degree in Physics and Math in 2015 from the SSAU. Currently he is a teaching assistant of the Technical Cybernetics sub-department and Engineer at SSAU's laboratory SRL-35 of Samara State Aerospace University, also working as an intern researcher of the Image Processing Systems Institute of the Russian Academy of Sciences, – Branch of the FSRC "Crystallography and Photonics" RAS, Samara, Russia. His research interests currently focus on computer image processing, pattern recognition, data mining and theory of computation. E-mail: [andrey.gaidel@gmail.com](mailto:andrey.gaidel@gmail.com).

**Victor Rostislavovich Krasheninnikov** (b. 1945), graduated from Kazan State University in 1967, Doctor of Technical Sciences (1996), Professor, the Head of Applied Mathematics and Informatics department of Ulyanovsk State Technical University (Russia). He has over 300 scientific papers on random process and image analysis. E-mail: [kvrulstu@mail.ru](mailto:kvrulstu@mail.ru).

*Received August 29, 2016. The final version – November 1, 2016.*