

ВОЗМОЖНОСТИ ТЕКСТУРНОГО АНАЛИЗА КОМПЬЮТЕРНЫХ ТОМОГРАММ В ДИАГНОСТИКЕ ХРОНИЧЕСКОЙ ОБСТРУКТИВНОЙ БОЛЕЗНИ

Гайдель А.В.^{1,2}, Зельтер П.М.³, Капишников А.В.³, Храмов А.Г.^{1,2}

¹ Институт систем обработки изображений РАН,

² Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет) (СГАУ),

³ Самарский государственный медицинский университет

Аннотация

Исследуется возможность использования разнообразных текстурных признаков для выявления патологических изменений лёгких по двумерным цифровым изображениям компьютерной томографии. В качестве информационных признаков используются гистограммные и корреляционные характеристики, а также признаки Харалика и признаки на основе длин серий. Отбор эффективных признаков проводится на основе критерия дискриминантного анализа. Применимость предложенного подхода для решения задачи автоматического выявления патологических изменений исследуется экспериментально на наборе из 160 изображений, полученных при обследовании больных хронической обструктивной болезнью лёгких. Полученная эффективная группа признаков состоит из двух признаков Харалика и трёх признаков на основе длин серий и обеспечивает долю неверных распознаваний 0,11, которая лучше аналогичных результатов, полученных без использования процедуры отбора признаков.

Ключевые слова: текстурный анализ, диагностика, признаки Харалика, длины серий, отбор признаков, дискриминантный анализ.

Введение

Хроническая обструктивная болезнь лёгких (ХОБЛ) – заболевание, связанное с повышенным хроническим воспалительным ответом лёгких на действие патогенных частиц или газов. ХОБЛ – это широко распространённое заболевание, в мире насчитывается более 200 миллионов больных. Одной из причин высокой смертности от заболевания является поздняя диагностика. Установлено, что только в 25 % случаев заболевание диагностируется своевременно. Компьютерная томография – наиболее точный метод диагностики морфологических изменений лёгочной ткани. Первоначально полученные изображения подвергаются визуальному исследованию квалифицированным врачом-рентгенологом, но, учитывая большой объём данных, активно разрабатываются компьютерные системы помощи врачу в выявлении и характеристике изменений. Использование компьютерного анализа томограмм лёгких в настоящее время уже помогает ускорить процедуру диагностики, объективизировать её, снизить её стоимость и повысить информативность диагностики [1]. Наиболее частыми признаками ХОБЛ при компьютерной томографии являются эмфизема (нарушение целостности альвеол с образованием воздушных полостей различного размера) и так называемые «воздушные ловушки», т.е. участки гипервоздушности, выявляемые на выдохе и свидетельствующие о наличии обструкции мелких бронхов (рис. 1) [2, 3].

Текстурный анализ уже использовался для диагностики подобных изображений в работе [4]. Доля правильно классифицированных изображений в этой работе составила от 74 % до 89 % в зависимости от выбранной области интереса. Большинство использованных признаков было основано на длинах серий.

Цель работы: исследовать возможности текстурного анализа при классификации компьютерных томограмм лёгких.

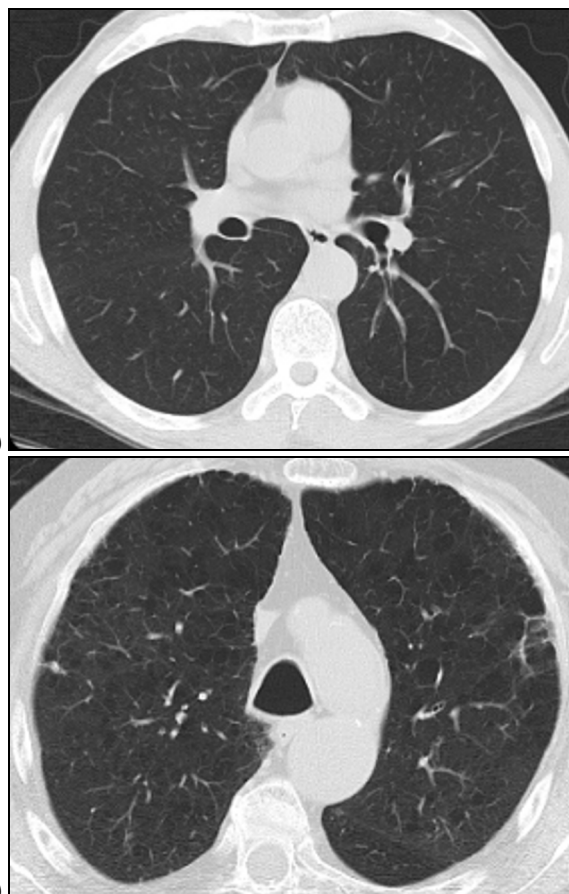


Рис. 1. Компьютерная томография лёгких: без патологии (а), с внутридольковой эмфиземой (б)

В настоящей работе также было решено использовать другие хорошо зарекомендовавшие себя для анализа биомедицинских изображений признаки. Так корреляционные характеристики хорошо показали себя в задаче диагностики остеопороза по рентгенов-

ским изображениям шейки бедра [5]. Признаки Харалика зарекомендовали себя для анализа нефрологических ультразвуковых изображений [6], а также для диагностики новообразований молочной железы [7] и скинтиграфических изображений лёгких [8]. Известны также результаты использования похожих методов в задачах анализа других биомедицинских изображений, таких как изображения сосудистой системы [9, 10, 11].

Для отбора наиболее эффективной группы из множества известных признаков целесообразно также использовать автоматическую процедуру, основанную на критерии дискриминантного анализа, пример одной из которых описан в [12]. Из приведённых работ только в [5] и [6] применяется такая процедура. Между тем в зарубежной литературе для отбора признаков используются и более сложные алгоритмы, такие как генетический алгоритм [13] и метод дифференциальной эволюции [14]. В отечественных работах дискриминантный анализ также уже использовался для формирования информационных признаков в задачах медицинской диагностики [15].

1. Текстурные признаки

Пусть имеется монохромное цифровое изображение $\omega(m, n)$, определённое на области интереса $D_\omega \subseteq \mathbb{Z}^2$ и отображающее во множество из Q оттенков яркости $[0, Q-1] \cap \mathbb{Z}$. Для описания этого изображения будем использовать несколько групп информационных признаков.

Гистограммные признаки

Простейшим описанием изображения являются его моментные характеристики. В предположении о том, что изображение $\omega(m, n)$ является реализацией стационарного эргодического случайного поля, моментные характеристики этого поля могут быть оценены статистически по отсчётам изображения:

$$v_k = \frac{1}{|D_\omega|} \sum_{(m,n) \in D_\omega} \omega^k(m, n),$$

$$\mu_k = \frac{1}{|D_\omega|} \sum_{(m,n) \in D_\omega} (\omega(m, n) - v_1)^k,$$

где оператор $|\cdot|$ для конечного множества возвращает количество элементов в нём.

Здесь v_k – оценки начальных моментов, μ_k – оценки центральных моментов. На их основе строятся следующие информационные признаки: средняя яркость v_1 , средняя энергия v_2 , среднеквадратическое отклонение $\sqrt{\mu_2}$, коэффициент асимметрии μ_3 / σ^3 и коэффициент эксцесса $(\mu_4 / \sigma^4 - 3)$. Нужно отметить, что эти характеристики не учитывают пространственное расположение отсчётов, так что, строго говоря, не являются текстурными признаками.

Корреляционные признаки

Простейшим способом статистически описать колебания яркости изображения в пространстве является оценка нормированной ковариационной функции:

$$R_{cov}(m, n) = \frac{\sum_{(i,j) \in D_\omega(m,n)} \omega(i, j)\omega(i+m, j+n)}{v_2 |D_\omega(m, n)|}, \quad (1)$$

где $D_\omega(m, n) = \{(i, j) \in D_\omega \mid (i+m, j+n) \in D_\omega\}$.

В [16] рекомендуется использовать для текстурного анализа именно нормированную ковариационную функцию вместо центрированной корреляционной функции, поскольку она лучше описывает пространственные изменения яркости. В настоящей работе в качестве признаков были использованы 8 отсчётов функции (1): для четырёх направлений (горизонтального, вертикального и двух диагональных) и двух расстояний (ближнего и дальнего).

Признаки Харалика

Признаки Харалика впервые были описаны в [17] и с тех пор получили широкое распространение в задачах текстурного анализа биомедицинских изображений. Они основаны на матрице вхождений

$$p_{m,n}(i, j) = \frac{|D_{m,n}^\#(i, j)|}{|D_\omega(m, n)|}, \quad (2)$$

определённой на $([0, Q-1] \cap \mathbb{Z})^2$, где

$$D_{m,n}^\#(i, j) = \{(k_1, k_2) \in D_\omega(m, n) \mid \{\omega(k_1, k_2), \omega(k_1+m, k_2+n)\} = \{i, j\}\}.$$

То есть это частота встречаемости заданной пары цветов на изображении в заданной конфигурации.

По одной такой матрице могут быть вычислены 14 признаков Харалика: второй угловой момент, контраст, корреляция, дисперсия, обратный разностный момент, средняя сумма цветов, дисперсия суммы цветов, энтропия суммы цветов, энтропия, дисперсия разности цветов, энтропия разности цветов, две информационных меры корреляции и максимальный коэффициент корреляции.

Так же как и в [5], в работе использовались 8 различных матриц (2) для четырёх направлений и двух расстояний. Таким образом, общее количество признаков Харалика, участвовавших в отборе, составило 112 штук.

Признаки на основе длин серий

Эта группа признаков основана на матрицах длин серий

$$r_\theta(i, j) = \left| \{(m, n) \in D_\omega \mid \{\omega(k, l) \mid (k, l) \in S_\theta(m, n, j)\} = \{i\}\} \right|, \quad (3)$$

определённых на $([0, Q-1] \cap \mathbb{Z}) \times \mathbb{N}_0$, где

$$S_{\theta}(m, n, j) = \{ \lfloor [m + t \cos \theta] \rfloor, \lfloor [n + t \sin \theta] \rfloor \mid \exists T > 0 :$$

$$t \in [0, T] \wedge |S_{\theta}(m, n, j)| = j \wedge S_{\theta}(m, n, j) \subseteq D_{\omega} \} -$$

серия длины j в направлении θ , начинающаяся с отсчёта (m, n) . Таким образом, $r_{\theta}(i, j)$ – это количество серий длины j в направлении θ , целиком входящих в область интереса D_{ω} и состоящих из отсчётов одинаковой яркости i .

На основе матриц (3) традиционно строится 11 информационных признаков [18].

Резкость коротких серий:

$$s_1 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} j^{-2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость длинных серий:

$$s_2 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} j^{+2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Неоднородность яркости:

$$s_3 = \frac{\sum_{i=0}^{Q-1} \left(\sum_{j=1}^{\infty} r_{\theta}(i, j) \right)^2}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Неоднородность длины:

$$s_4 = \frac{\sum_{j=1}^{\infty} \left(\sum_{i=0}^{Q-1} r_{\theta}(i, j) \right)^2}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Доля серий:

$$s_5 = \frac{1}{|D_{\omega}|} \sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j).$$

Резкость низкой яркости:

$$s_6 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{-2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость высокой яркости:

$$s_7 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{+2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость коротких серий и низкой яркости:

$$s_8 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{-2} j^{-2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость коротких серий и высокой яркости:

$$s_9 = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{+2} j^{-2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость длинных серий и низкой яркости:

$$s_{10} = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{-2} j^{+2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

Резкость длинных серий и высокой яркости:

$$s_{11} = \frac{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} i^{+2} j^{+2} r_{\theta}(i, j)}{\sum_{i=0}^{Q-1} \sum_{j=1}^{\infty} r_{\theta}(i, j)}.$$

К сожалению, в отечественной литературе эти признаки пока что описаны мало. Также существует не слишком много публикаций, в которых исследуется их эффективность для анализа биомедицинских изображений и для автоматической диагностики. В этой работе матрицы (3) вычислялись для четырёх традиционных направлений, так что в общей сложности использовались 44 признака на основе длин серий.

2. Отбор признаков

Пусть каждому допустимому изображению $\omega(m, n)$ уже поставлен в соответствие вектор его признаков $x(k)$ и множество векторов признаков Ω разбито на два подмножества: Ω_0 – вектора признаков здоровых лёгких, Ω_1 – вектора признаков лёгких с патологией, так что $\Omega_0 \cup \Omega_1 = \Omega$ и $\Omega_0 \cap \Omega_1 = \emptyset$. На этом множестве определён идеальный оператор классификации $\Phi(x) : \Omega \rightarrow \{\Omega_1, \Omega_2\}$, ставящий в соответствие вектору его класс. Обозначим $U \subseteq \Omega$ конечную обучающую выборку признаков изображений, класс которых заранее известен.

Понятно, что в силу гетерогенной природы все признаки имеют разный масштаб, то есть различные признаки могут вносить сильно отличающийся вклад в расстояние между векторами признаков. Для решения этой проблемы воспользуемся *процедурой стандартизации*: полагая вектора признаков случайными, для отдельных признаков оценим их математические ожидания

$$M(k) = \frac{1}{|U|} \sum_{x \in U} x(k)$$

и корреляции между признаками

$$R(m, n) = \frac{1}{|U|} \sum_{x \in U} (x(m) - M(m))(x(n) - M(n)).$$

Будем пересчитывать все имеющиеся и последующие признаки, центрируя и нормируя их:

$$\Lambda[x] = \left(\frac{x(k) - M(k)}{\sqrt{R(k, k)}} \right)_k.$$

Получим новое пространство признаков $\Omega' = \{\Lambda[x] | x \in \Omega\} = \Omega'_1 \cup \Omega'_2$, на котором также определён свой идеальный классификатор $\Phi'(x)$, такой что $\Phi'(\Lambda[x]) = \Phi(x)$, и новую обучающую выборку $U' = U'_1 \cup U'_2$. Имеется в виду, что $\Omega'_i = \{\Lambda[x] | \Phi(x) = \Omega_i\}$ и $U'_i = \{\Lambda[x] | x \in U \wedge \Phi(x) = \Omega_i\}$.

Считаем, что признаки в этом пространстве имеют нулевые математические ожидания и единичные дисперсии. Для оценки эффективности отдельных признаков вычислим внутриклассовые математические ожидания

$$M_i(k) = \frac{1}{|U'_i|} \sum_{x \in U'_i} x(k)$$

и внутриклассовые корреляционные матрицы

$$R_i(m, n) = \frac{1}{|U'_i|} \sum_{x \in U'_i} (x(m) - M_i(m))(x(n) - M_i(n)).$$

Положим, что группа признаков тем лучше разделяет изображения из разных классов, чем больше критерий

$$J = (\text{tr}(R_0 + R_1))^{-1}, \tag{4}$$

основанный на одном из критериев дискриминантного анализа [19]. Его значение тем больше, чем меньше разброс значений векторов признаков внутри классов по сравнению с общим разбросом этих векторов вне зависимости от класса.

Очевидно, что выбрать группу признаков заданного размера, обеспечивающую максимум критерия (4), можно с помощью обычного жадного алгоритма, поскольку каждый признак вносит в изменение следа матрицы вклад, равный обратному значению этого критерия для этого отдельного признака. С этой целью для каждого признака вычислим индивидуальное значение критерия (4) по обучающей выборке и отсортируем признаки в порядке убывания этого критерия. На k -м шаге рассмотрим группу из первых k признаков и произведём с её помощью классификацию векторов из конечной контрольной выборки $\tilde{U}' \subseteq \Omega'$, такой что $\tilde{U}' \cap U' = \emptyset$. На каждом шаге вычислим значение ошибки классификации

$$\varepsilon = \frac{1}{|\tilde{U}'|} \left| \left\{ x \in \tilde{U}' \mid \tilde{\Phi}'(x) \neq \Phi'(x) \right\} \right|, \tag{5}$$

где $\tilde{\Phi}'(x)$ – это некоторый построенный классификатор, обученный на векторах из U' . Окончательно объявим эффективной группу признаков, обеспечивающую приемлемую ошибку классификации ε . Аналогичные процедуры и критерии использовались в работе [5].

3. Экспериментальные исследования

Для исследования использовались 160 реальных диагностических изображений компьютерной томо-

графии лёгких, полученных в результате обследования пациентов клиник Самарского государственного медицинского университета. Изображения были получены на компьютерном томографе Aquilion 32 (Toshiba). Эти изображения были вручную ориентированы, и квалифицированным специалистом на них были выделены области интереса, включающие только лёгочную ткань. Изображения были классифицированы на основании визуальной оценки на норму (44 изображения), по типам эмфиземы и по наличию симптома «воздушной ловушки».

Фильтрация шума на изображении не проводилась для сохранения высокочастотной информации, которая может иметь диагностический потенциал. Набор компьютерных томограмм случайным образом поровну распределён на обучающую и контрольную выборки.

Для классификации во всех случаях использовался метод ближайшего соседа ввиду высокой скорости его работы, простоты реализации, возможности обучения с учителем и известной связи ошибки классификации с ошибкой классификации при помощи байесовского классификатора [19]. Так, ошибка байесовского классификатора не более чем в два раза меньше, чем ошибка классификатора по методу ближайшего соседа. В качестве расстояния между векторами признаков использовалось обычное евклидово расстояние, что вполне допустимо, поскольку признаки стандартизованы.

В табл. 1 представлены результаты индивидуального анализа десяти лучших признаков по критерию (4). В третьем столбце указаны сами значения критерия, а в четвёртом – значения ошибки классификации (5), проведённой с использованием соответствующего единственного признака. Видно, что даже поодиночке лучшие признаки обеспечивают ошибку классификации от 0,20 до 0,28. Между тем явная зависимость между значениями критерия и ошибкой классификации отсутствует. Кроме того, нужно заметить, что, хотя первые два места занимают признаки Харалика, остальные признаки в десятке лучших основаны на длинах серий.

Таблица 1. Качество отдельных признаков

№	Признак	$J(k)$	ε
1	Ближняя горизонтальная первая информационная мера корреляции	7,06	0,27
2	Ближняя вертикальная первая информационная мера корреляции	6,57	0,23
3	Вертикальная резкость длинных серий	6,29	0,24
4	Горизонтальная резкость длинных серий	6,18	0,21
5	Вертикальная доля серий	6,04	0,28
6	Побочнодиагональная резкость длинных серий	5,96	0,26
7	Вертикальная резкость коротких серий	5,95	0,27
8	Горизонтальная доля серий	5,92	0,20
9	Главная диагональная резкость длинных серий	5,88	0,24
10	Побочнодиагональная доля серий	5,86	0,23

Общее количество отдельных признаков, среди которых проводился отбор, составило 169 штук. В табл. 2 приведены результаты выбора лучшей группы признаков, произведённого с помощью процедуры, описанной в предыдущем разделе. Номера признаков соответствуют табл. 1. Полученные результаты свидетельствуют о том, что наилучшая в рамках выбранной процедуры ошибка классификации в 0,11 достигается уже при использовании группы из пяти лучших признаков, в которую входят два признака Харалика и три признака на основе длин серий. В дальнейшем из-за переобучения оценка ошибки классификации только увеличивается при добавлении новых признаков.

Таблица 2. Результаты отбора группы признаков

Группа признаков	ϵ
1	0,27
1, 2	0,22
1, 2, 3	0,13
1, 2, 3, 4	0,13
1, 2, 3, 4, 5	0,11
1, 2, 3, 4, 5, 6	0,11
1, 2, 3, 4, 5, 6, 7	0,11
1, 2, 3, 4, 5, 6, 7, 8	0,12
1, 2, 3, 4, 5, 6, 7, 8, 9	0,12
1, 2, 3, 4, 5, 6, 7, 8, 9, 10	0,12

Однако на рис. 2 видно, как ошибка классификации достигает ещё одного локального минимума в 0,07 позже для группы из 29 признаков. Хотя это меньше, чем 0,11, 29 разнородных признаков будут вычисляться значительно дольше, и нет оснований считать, что такое неожиданное уменьшение ошибки вызвано именно качеством группы признаков, а не произошло случайно.

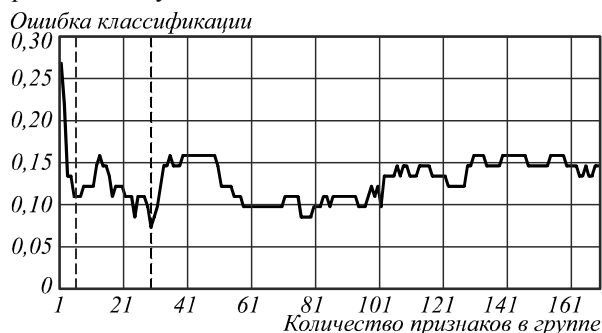


Рис. 2. Зависимость ошибки классификации от размера группы признаков

На рис. 3 изображены некоторые векторы признаков из контрольной выборки в пространстве из двух лучших признаков Харалика, указанных в табл. 1 под номерами 1 и 2. Для наглядности указаны не стандартизованные значения признаков. Все векторы не приведены, чтобы не загромождать рисунок. Окружностями обозначены векторы, соответствующие изображениям здоровых лёгких, а звёздами – изображениям лёгких с патологией. Заметна явная тенденция к

разделимости. Также наблюдается довольно сильная корреляция между признаками, что не очень хорошо, поскольку второй признак несёт не слишком много новой информации, но, к сожалению, использованная процедура никак не учитывает эту корреляцию при отборе признаков.

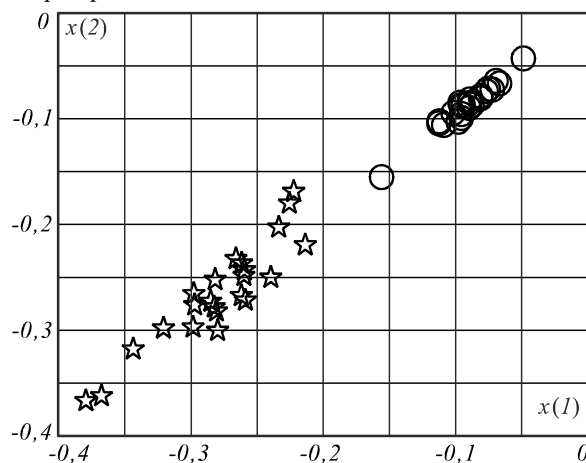


Рис. 3. Векторы признаков из контрольной выборки в пространстве из двух лучших признаков

Напоследок отметим, что получили оценку вероятности ошибочной классификации $\epsilon = 0,11$ в результате $n = 80$ экспериментов, каждый из которых мог закончиться успехом или провалом. Вычислим для нашей оценки доверительный интервал Агрести–Коула [20] с уровнем доверия $\alpha = 0,05$. Для этого пересчитаем саму вероятность как

$$\hat{\epsilon} = \frac{\epsilon n + 0,5z^2}{n + z^2},$$

тогда доверительный интервал составит

$$\hat{\epsilon} \pm z \sqrt{\frac{\hat{\epsilon}(1-\hat{\epsilon})}{n + z^2}},$$

где z – это $(1 - 0,5\alpha)$ -квантиль стандартного нормального распределения. Таким образом, не учитывая количество степеней свободы при выборе признаков, вероятность ошибочной классификации $\hat{\epsilon} = 0,13$ лежит вне интервала $(0,06; 0,20)$ лишь с вероятностью $\alpha = 0,05$. Был использован именно доверительный интервал Агрести–Коула, поскольку в работе [21] показывается его превосходство над аналогами и стандартным интервалом Вальда при числе испытаний $n > 40$.

Заключение

В работе экспериментально проверена применимость текстурного анализа для задачи автоматического выявления патологических изменений и классификации изображений при компьютерной томографии лёгких. Для этого использована процедура отбора наиболее эффективной группы признаков, основанная на критерии дискриминантного анализа. Отбор производился среди 169 различных признаков, включая уже использовавшиеся раньше гистограммные,

корреляционные признаки и признаки Харалика, а также реже встречающиеся в отечественных работах признаки, основанные на длинах серий.

Получена наиболее эффективная группа признаков, состоящая из двух информационных мер Харалика и трёх признаков на основе длин серий. Признаки на основе длин серий также заняли оставшиеся места в десятке лучших и эффективно себя показали при выявлении патологических изменений. Это может свидетельствовать об их общей эффективности для текстурного анализа биомедицинских изображений такого типа.

Лучшая группа признаков обеспечивает долю неверно классифицированных изображений, равную 0,11, то есть из 80 изображений контрольной выборки 9 были классифицированы с ошибкой. Согласно доверительному интервалу Агрести–Коула, это означает, что вероятность ошибочной классификации $\hat{\epsilon} = 0,13$ лежит в интервале с уровнем доверия $\alpha = 0,05$ (стандартный показатель для медицинской практики [3]). Например, он превосходит показатель, полученный в аналогичной работе [4].

Кроме прочего, в ходе экспериментов лучшее качество распознавания было продемонстрировано при выявлении «воздушных ловушек». Это является надёживающим показателем, так как «воздушные ловушки» проявляются у пациентов на ранней стадии ХОБЛ, часто являясь единственным отклонением от нормы [3]. Менее точные результаты показали изображения лёгких с внутридольковой эмфиземой, следовательно, для её выявления необходима разработка других методик.

Благодарности

Работа выполнена при финансовой поддержке гранта РФФИ 14-07-97040-р_поволжье_а и Министерства образования и науки РФ в рамках мероприятий Программы повышения конкурентоспособности СГАУ среди ведущих мировых научно-образовательных центров на 2013–2020 годы.

Литература

1. Зельтер, П.М. Структурно-функциональное фенотипирование хронической обструктивной болезни лёгких / П.М. Зельтер, М.С. Устинов, А.В. Капишников // Терапевт. – 2014. – № 7. – С. 18-23.
2. Engeler, C.E. Ground-glass opacity of the lung parenchyma: A guide to analysis with high-resolution CT / C.E. Engeler, J.H. Tashjian, S.W. Trenkner, J.W. Walsh // American Journal of Roentgenology / American Roentgen Ray Society. – 1993. – Vol. 160(2). – P. 249-251. – ISSN 0361-803X.
3. Зельтер, П.М. Роль компьютерной томографии в раннем выявлении признаков хронической обструктивной болезни лёгких / П.М. Зельтер // Врач-аспирант. – 2014. – № 4.2(65). – С. 228-233.
4. Ginsburg, S.B. Automated Texture-based Quantification of Centrilobular Nodularity and Centrilobular Emphysema in Chest CT Images / S.B. Ginsburg, D.A. Lynch, R.P. Bowler, J.D. Schroeder // Academic Radiology. – 2012. – Vol. 19(10). – P. 1241-1251. – ISSN 1076-6332.
5. 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.
6. Гайдель, А.В. Исследование текстурных признаков для диагностики нефрологических заболеваний по ультразвуковым изображениям / А.В. Гайдель, С.Н. Ларионова, А.Г. Храмов // Вестник Самарского государственного аэрокосмического университета имени академика С.П. Королёва (национального исследовательского университета). – 2014. – № 1(43). – С. 229-237.
7. Садыков, С.С. Компьютерная диагностика новообразований на маммографических снимках / С.С. Садыков, Ю.А. Буланова, Е.А. Захарова // Компьютерная оптика. – 2014. – Т. 38, № 1. – С. 131-138.
8. Глумов, Н.И. Компьютерная обработка скинтиграфических изображений лёгких / Н.И. Глумов, А.В. Капишников // Компьютерная оптика. – 2003. – № 25. – С. 158-164.
9. Ильясова, Н.Ю. Методы цифрового анализа сосудистой системы человека. Обзор литературы // Компьютерная оптика. – 2013. – Т. 37, № 4. – С. 511-535. – ISSN 0134-2452.
10. 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.
11. Ильясова, Н.Ю. Оценивание геометрических признаков пространственной структуры кровеносных сосудов // Компьютерная оптика. – 2014. – Т. 38, № 3. – С. 529-538. – ISSN 0134-2452.
12. Глумов, Н.И. Метод отбора информативных признаков на цифровых изображениях / Н.И. Глумов, Е.В. Мясников // Компьютерная оптика. – 2007. – Т. 31, № 3. – С. 73-76.
13. Tsai, C.-F. Genetic algorithms in feature and instance selection / C.-F. Tsai, W. Eberle, C.-Y. Chu // Knowledge-Based Systems. – 2013. – Vol. 39. – P. 240-247. – ISSN 0950-7051.
14. Khushaba, R.N. Feature subset selection using differential evolution and a statistical repair mechanism / R.N. Khushaba, A. Al-Ani, A. Al-Jumaily // Expert Systems with Applications. – 2011. – Vol. 38(9). – P. 11515-11526. – ISSN 0957-4174.
15. Ильясова, Н.Ю. Формирование признаков для повышения качества медицинской диагностики на основе методов дискриминантного анализа / Н.Ю. Ильясова, А.В. Куприянов, Р.А. Парингер // Компьютерная оптика. – 2014. – Т. 38, № 4. – С. 851-856.
16. Petrou, M. Image Processing: Dealing with Texture / M. Petrou, P. Garcia Sevilla. – Chichester, UK: John Wiley & Sons, Ltd., 2006. – 618 p.
17. Haralick, R.M. Textural features for image classification / R.M. Haralick, K. Shanmugam, I. Dinstein // IEEE Transactions on Systems, Man, and Cybernetics. – 1973. – Vol. SMC-3(6). – P. 610-621. – ISSN 0018-9472.
18. Mollazade, K. Analysis of texture-based features for predicting mechanical properties of horticultural products by laser light backscattering imaging / K. Mollazade, M. Omid, F. Akhlaghian Tab, Y.R. Kalaj, S.S. Mohtasebi, M. Zude // Computers and Electronics in Agriculture. – 2013. – Vol. 98. – P. 34-45. – ISSN 0168-1699.
19. Fukunaga, K. Introduction to statistical pattern recognition / K. Fukunaga. – San Diego: Academic Press, 1990. – 592 p.
20. Agresti, A. Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions / A. Agresti, B.A. Coull // American Statistician. American Statistical Association. – 1998. – Vol. 52(2). – P. 119-126. – ISSN 0003-1305.
21. Brown, L.D. Interval Estimation for a Binomial Proportion / L.D. Brown, T.T. Cai, A. DasGupta // Statistical Science. – 2001. – Vol. 16(2). – P. 101-133. – ISSN 0883-4237.

References

1. **Zelter, P.M.** Structure and functional phenotyping of COPD / P.M. Zelter, M.S. Ustinov, A.V. Kapishnikov // *Physician*. – 2014. – Vol. 7(65). – P. 18-23. – ISSN 20726-0277. – (In Russian).
2. **Engeler, C.E.** Ground-glass opacity of the lung parenchyma: A guide to analysis with high-resolution CT / C.E. Engeler, J.H. Tashjian, S.W. Trenkner, J.W. Walsh // *American Journal of Roentgenology*. – 1993. – Vol. 160(2). – P. 249-251. – ISSN 0361-803X.
3. **Zelter, P.M.** Role of CT in early evaluating of COPD signs / P.M. Zelter // *Postgraduate Doctor*. – 2014. – Vol. 4.2(65). – P. 228-233. – ISSN 1816-5214. – (In Russian).
4. **Ginsburg, S.B.** Automated Texture-based Quantification of Centrilobular Nodularity and Centrilobular Emphysema in Chest CT Images / S.B. Ginsburg, D.A. Lynch, R.P. Bowler, J.D. Schroeder // *Academic Radiology*. – 2012. – Vol. 19(10). – P. 1241-1251. – ISSN 1076-6332.
5. **Gaidel, A.V.** Research of the textural features for the bony tissue diseases diagnostics using the roentgenograms / A.V. Gaidel, S.S. Pervushkin // *Computer Optics*. – 2013. – Vol. 37(1). – P. 113-119. – ISSN 0134-2452.
6. **Gaidel, A.V.** Research of the textural features for the nephrological diseases diagnostics using the ultrasound images / A.V. Gaidel, S.N. Larionova, A.G. Khramov // *Herald of the Samara State Aerospace University*. – 2014. – Vol. 43(1). – P. 229-237. – ISSN 1998-6629. – (In Russian).
7. **Sadykov, S.S.** Computer diagnosis of tumors in mammograms / S.S. Sadykov, Yu.A. Bulanov, E.A. Zakharova // *Computer Optics*. – 2014. – Vol. 38(1). – P. 131-138. – ISSN 0134-2452.
8. **Glumov, N.I.** Computer processing of lung scintigraphic images / N.I. Glumov, A.V. Kapishnikov // *Computer Optics*. – 2003. – Vol. 25. – P. 158-164. – ISSN 0134-2452. – (In Russian).
9. **Ilyasova, N.Yu.** Methods for digital analysis of human vascular system. Literature review / N.Yu. Ilyasova // *Computer Optics*. – 2013. – Vol. 37(4). – P. 511-535. – ISSN 0134-2452.
10. **Ilyasova, N.Yu.** Information technologies of image analysis in medical diagnostics / N.Yu. Ilyasova, A.V. Kupriyanov, A.G. Khramov. – Moscow: "Radio I Svyaz" Publisher, 2012. – 424 p. – ISBN 5-89776-014-4. – (In Russian).
11. **Ilyasova, N.Yu.** Estimating the geometric features of a 3D vascular structure // *Computer Optics*. – 2014. – Vol. 38(3). – P. 529-538. – ISSN 0134-2452.
12. **Glumov, N.I.** Method of the informative features selection on the digital images / N.I. Glumov, E.V. Myasnikov // *Computer Optics*. – 2007. – Vol. 31(3). – P. 73-76. – ISSN 0134-2452. – (In Russian).
13. **Tsai, C.-F.** Genetic algorithms in feature and instance selection / C.-F. Tsai, W. Eberle, C.-Y. Chu // *Knowledge-Based Systems*. – 2013. – Vol. 39. – P. 240-247. – ISSN 0950-7051.
14. **Khushaba, R.N.** Feature subset selection using differential evolution and a statistical repair mechanism / R.N. Khushaba, A. Al-Ani, A. Al-Jumaily // *Expert Systems with Applications*. – 2011. – Vol. 38(9). – P. 11515-11526. – ISSN 0957-4174.
15. **Ilyasova, N.Yu.** Formation features for improving the quality of medical diagnosis based on the discriminant analysis methods / N.Yu. Ilyasova, A.V. Kupriyanov, R.A. Paringer // *Computer Optics*. – 2014. – Vol. 38(4) – P. 851-856. – ISSN 0134-2452. – (In Russian).
16. **Petrou, M.** *Image Processing: Dealing with Texture* / M. Petrou, P. Garcia Sevilla. – Chichester, UK: John Wiley & Sons, Ltd., 2006. – 618 p.
17. **Haralick, R.M.** Textural features for image classification / R.M. Haralick, K. Shanmugam, I. Dinstein // *IEEE Transactions on Systems, Man, and Cybernetics*. – 1973. – Vol. SMC-3(6). – P. 610-621. – ISSN 0018-9472.
18. **Mollazade, K.** Analysis of texture-based features for predicting mechanical properties of horticultural products by laser light backscattering imaging / K. Mollazade, M. Omid, F. Akhlaghian Tab, Y.R. Kalaj, S.S. Mohtasebi, M. Zude // *Computers and Electronics in Agriculture*. – 2013. – Vol. 98. – P. 34-45. – ISSN 0168-1699.
19. **Fukunaga, K.** *Introduction to statistical pattern recognition* / K. Fukunaga. – San Diego: Academic Press, 1990. – 592 p.
20. **Agresti, A.** Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions / A. Agresti, B.A. Coull // *American Statistician*. – 1998. – Vol. 52(2). – P. 119-126. – ISSN 0003-1305.
21. **Brown, L.D.** Interval Estimation for a Binomial Proportion / L.D. Brown, T.T. Cai, A. DasGupta // *Statistical Science / Institute of Mathematical Statistics*. – 2001. – Vol. 16(2). – P. 101-133. – ISSN 0883-4237.

COMPUTED TOMOGRAPHY TEXTURE ANALYSIS CAPABILITIES IN DIAGNOSING A CHRONIC OBSTRUCTIVE PULMONARY DISEASE

A.V. Gaidel^{1,2}, P.M. Zelter³, A.V. Kapishnikov³, A.G. Khramov^{1,2}

¹Image Processing Systems Institute, Russian Academy of Sciences,

²Samara State Aerospace University,

³Samara State Medical University

Abstract

The possibility of application of different textural features for the lung disease automatic diagnosis on the basis of the 2D digital computed tomography (CT) images was studied. Histogram features, covariance features, Haralick's features and run length features were used. A procedure based on the discriminant analysis criterion was used for the selection of the best features group. We experimentally showed that the approach offered is convenient to use for solving the problem of automatic diagnosis on a 160-image set received during examination of patients with a chronic obstructive pulmonary disease. The resulting group of effective features includes two Haralick's features and three run length features, providing the error rate of 0.11, which is better than similar results obtained without a feature selection procedure.

Key words: textural analysis, diagnosis, Haralick's features, run length, feature selection, discriminant analysis.

Сведения об авторах

Гайдель Андрей Викторович, 1989 года рождения, в 2012 году с отличием окончил Самарский государственный аэрокосмический университет имени академика С. П. Королёва по направлению «Прикладная математика и информатика», работает инженером НИЛ-35 СГАУ и стажёром-исследователем в Институте систем обработки изображений РАН. Область научных интересов: компьютерная обработка изображений, теория распознавания образов, математическое моделирование, теория алгоритмов.

E-mail: andrey.gaidel@gmail.com.

Andrey Viktorovich Gaidel (b. 1989) graduated with honors from Samara State Aerospace University in 2012, majoring in Applied Mathematics and Informatics. Engineer of SRL-35 laboratory of Samara State Aerospace University, intern researcher of the Image Processing Systems Institute of the Russian Academy of Sciences. Research interests are currently focused on computer image processing, pattern recognition, mathematical modeling and theory of computation.



Зельтер Павел Михайлович, 1986 года рождения, в 2010 году с отличием окончил Самарский государственный медицинский университет, в 2012 году – ординатуру по специальности «Рентгенология», работает ассистентом кафедры лучевой диагностики и лучевой терапии с курсом медицинской информатики СамГМУ. Область научных интересов: компьютерная томография, ХОБЛ, компьютерный анализ медицинских изображений.

E-mail: pzelter@mail.ru.

Pavel Mikhailovich Zelter (b. 1986) graduated with honors from Samara State Medical University in 2010, residency in radiology in 2012. Assistant lecturer of Radiology department with course of Medical informatics of Samara State Medical University. Research interests are currently focused on computerized tomography, COPD, digital analysis of medical images.



Капишников Александр Викторович, 1955 года рождения. В 1972 году окончил лечебный факультет Куйбышевского медицинского института. Докторскую диссертацию защитил в 2004 году. Заведует кафедрой лучевой диагностики и лучевой терапии с курсом медицинской информатики СамГМУ Минздрава России. Область научных интересов: лучевая диагностика и ядерная медицина в пульмонологии, остеологии, нефрологии и онкологии, информационные технологии медицинской радиологии, CADe и CADx – системы для компьютерной поддержки диагностики.

E-mail: a.kapishnikov@gmail.com.

Aleksandr Viktorovich Kapishnikov (b.1955). In 1972 he graduated from the Medical faculty of Kuibyshev Medical Institute. Doctor of Science in Medicine (2004). Head of Diagnostic Radiology and Medical Informatics department of Samara State Medical University. Research interests: diagnostic radiology and nuclear medicine in pulmonology, osteology, nephrology and oncology, information technology in medical imaging, computer-based diagnostic support system (CADe and CADx).



Храмов Александр Григорьевич, 1954 года рождения, д.т.н., доцент. В 1977 году окончил Куйбышевский авиационный институт им. академика С.П. Королёва по специальности «Автоматизированные системы обработки информации и управления». В 2006 году защитил докторскую диссертацию. В настоящее время профессор кафедры технической кибернетики Самарского государственного аэрокосмического университета имени академика С.П. Королёва (национальный исследовательский университет), а также старший научный сотрудник в Институте систем обработки изображений РАН. Область научных интересов: обработка и анализ биомедицинских изображений, интеллектуальный анализ данных, теория случайных процессов.

E-mail: khramov@smr.ru.

Alexander Grigorievich Khramov (b. 1954) is a Doctor of Engineering, Associate Professor. In 1977 he graduated from Kuibyshev Aviation Institute named after academician S.P. Korolyov, majoring in Automated Systems of Information Processing and Management. He received his D.Sc. degree in 2006. Currently he is a professor at the Technical Cybernetics department of Samara State Aerospace University and a Senior Researcher at the Image Processing Systems Institute of the RAS. Research interests: biomedical image processing and analysis, data mining, theory of random processes.

Поступила в редакцию 16 октября 2014 г.