

УДК 616.24-006.6+616.24-007.288]616-073.75+51(476)

В.А. Ковалев<sup>1</sup>, С.А. Хоружик<sup>2</sup>, А.Н. Михайлов<sup>3</sup>

## ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ЯРКОСТНОЙ ИНФОРМАЦИИ ДЛЯ СЕГМЕНТАЦИИ УЧАСТКОВ АТЕЛЕКТАЗА И РАКОВЫХ ОПУХОЛЕЙ НА ИЗОБРАЖЕНИЯХ ЛЕГКОГО, ПОЛУЧЕННЫХ С ПОМОЩЬЮ КОМПЬЮТЕРНОЙ ТОМОГРАФИИ\*

*Исследуется и количественно оценивается полезность яркостной информации КТ-изображений легкого для сегментации участков ателектаза и раковых опухолей, которые практически неразличимы при визуальном анализе. Проводится анализ статистической значимости межклассовых различий ателектаза и опухоли и осуществляется оценка точности классификации при использовании методов иерархической кластеризации, опорных векторов и метода случайного леса. Показано, что степень полезности яркостной информации изменяется от пациента к пациенту и зависит, главным образом, от размера выборки вокселей.*

### Введение

Ателектаз – это спадение всего легкого или его части вследствие нарушения вентиляции, обусловленного обтурацией бронха или сдавлением легкого. В настоящей работе рассматривается ателектаз, вызванный поражением бронха раковой опухолью.

Рентгеновская компьютерная томография (КТ) является основным способом получения изображения при раке легкого. Современные компьютерные томографы обеспечивают получение от нескольких десятков до нескольких сотен срезов (слоев). Слои расположены перпендикулярно оси туловища, непосредственно прилегают друг к другу и могут иметь толщину от 0,5 мм до 1 см. С точки зрения последующего анализа КТ-изображений проблема заключается в том, что участки ателектаза и раковой опухоли имеют очень близкие значения интенсивности (яркости) КТ-сигнала, что делает их практически неразличимыми при визуальном анализе КТ-изображений [1, 2]. Использование более редких и дорогих способов сканирования, таких, например, как позитронная эмиссионная томография (ПЭТ), позволяет различать ателектаз и опухоль [3, 4], но не обеспечивает приемлемого пространственного разрешения. Тем не менее, аккуратная и точная сегментация опухоли является крайне необходимой по следующим причинам. Во-первых, корректная локализация и сегментация опухоли, а также точное измерение ее размеров являются ключевыми моментами в определении стадии заболевания и выборе подходящей методики проведения операции. Во-вторых, в случае назначения лучевой терапии точное выделение границы опухоли необходимо для определения мишени, подлежащей облучению, и предотвращения облучения непораженных участков легкого [2, 4].

Данную работу можно рассматривать как начало более общего исследования, конечной целью которого является разработка методов, алгоритмов и программного обеспечения для интерактивной локализации и сегментации злокачественных опухолей легкого на фоне ателектаза. Статья посвящена исследованию и количественной оценке потенциальной полезности яркостной информации КТ-изображений легкого для дискриминации участков указанных двух типов. С этой целью в работе проводится оценка статистической значимости межклассовых различий ателектаза и опухоли и точности классификации с использованием наиболее передовых методов распознавания образов.

### 1. Материалы

В настоящем исследовании использовались КТ-изображения легких восьми пациентов с диагнозом «рак легкого» и ателектазом части легкого. Один из пациентов имел ателектаз двух

\* Работа выполнена при частичной финансовой поддержке гранта INTAS 04-77-7036.

долей легкого, три других – ателектаз одной доли и остальные четыре пациента – ателектаз сегмента (части доли) легкого. Группа пациентов состояла из семи женщин и одного мужчины в возрасте от 46 до 79 лет.

КТ-сканирование проводилось при стандартных значениях параметров напряжения и силы тока в рентгеновской трубке с задержкой дыхания на одном вдохе пациента. Размер воксела изображения составлял около 0,68 мм в аксиальной плоскости (т. е. в плоскости каждого слоя изображения) при толщине слоя и расстоянии между слоями 7 мм. Методика улучшения качества изображения путем внутривенного введения контрастного вещества не использовалась. Существенные отличия физических размеров вокселей в разных измерениях не позволили провести истинно трехмерный анализ КТ-изображений. Вместо этого каждое изображение трактовалось как совокупность (набор) двухмерных слоев. Типичные примеры слоев исходных КТ-изображений показаны на рис. 1.

Участки ателектаза и опухоли выделялись на КТ-изображениях органов грудной клетки каждого пациента путем ручного сегментирования квалифицированным специалистом-рентгенологом. У разных пациентов опухоль и ателектаз присутствовали на 5–26 (в среднем на 12) слоях. Во время ручной сегментации использовалось так называемое «тканевое окно» представления яркости изображений, при котором видны мягкие ткани, включая ателектаз и опухоль, но собственно легочная ткань не видна (см. черные участки изображений на рис. 1). При этом учитывалось, что опухоль, как правило, располагалась в легком ближе к центру (т. е. ближе к средостению), в то время как области ателектаза чаще наблюдались ближе к периферии легкого. Кроме того, контуры опухоли имеют более выпуклую форму, в то время как контуры области ателектаза нередко прямолинейные вследствие ограничения междолевой плеврой.

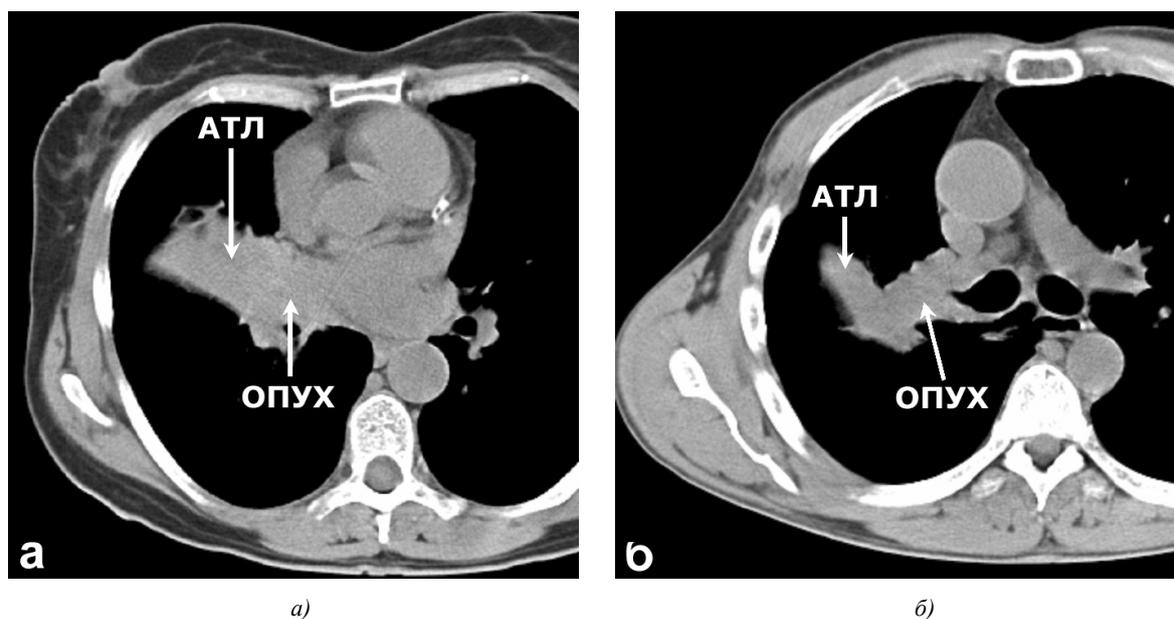


Рис. 1. КТ-изображения легких пациентов с ателектазом (АТЛ) и раковой опухолью (ОПУХ):  
а) пациента 1 с диагнозом «рак промежуточного и среднедолевого бронхов правого легкого», ателектаз средней доли; б) пациента 2 с диагнозом «рак правого верхнедолевого бронха», ателектаз заднего сегмента верхней доли

Прежде чем переходить к количественному анализу яркостных свойств ателектаза и злокачественных опухолей, необходимо подчеркнуть, что ручная сегментация проводилась рентгенологом интуитивно на основе клинического опыта. Поэтому в некоторых случаях выделение истинной границы не могло быть гарантировано в силу чрезвычайно близких параметров яркостей опухоли и ателектаза.

## 2. Методы

Очевидно, что в общем случае количественная оценка степени различия двух произвольных классов объектов зависит как от объективно присутствующих межклассовых различий, так и от метода сравнения. Для того чтобы исключить предвзятость одного конкретного метода, яркостные свойства ателектаза и раковой опухоли сравнивались четырьмя различными способами, а именно путем статистической оценки и путем применения трех различных методов классификации. Методика сравнения заключалась в многократном повторении процедуры, которая включала формирование двух выборок одинакового размера, состоящих из случайно взятых вокселей ателектаза и опухоли (безотносительно к слою трехмерного изображения, на котором они находятся), и последующую оценку степени их различия в зависимости от размера выборки. Для удобства интерпретации результатов размер выборок задавался так, чтобы он был равен числу вокселей квадратного окна слоя КТ-изображения со сторонами 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 и 30 вокселей. Это соответствует одиннадцати типам выборок размером 9, 16, 25, 36, 49, 64, 81, 100, 225, 400 и 900 вокселей.

### 2.1. Статистический анализ

При оценке статистической значимости различий в яркостных свойствах ателектаза и опухоли процедура тестирования включала:

- формирование для каждого пациента двух выборок заданного размера, состоящих из вокселей, взятых случайным образом из участков ателектаза и раковой опухоли безотносительно к положению слоя трехмерного КТ-изображения;
- оценку статистической значимости различий яркостей вокселей двух выборок с использованием стандартного критерия значимости различия средних Стьюдента (так называемый  $t$ -тест [5–7]).

Известно, что значения  $t$ -статистик Стьюдента, выражающие степень значимости различий, зависят от степеней свободы (в данном случае от размера выборок вокселей). С целью упрощения сравнения результатов оценки значимости, полученных при различных размерах выборок, результирующие величины значимости  $t$  преобразовывались в  $z$ -значимость по формуле

$$z = \sqrt{d \ln\left(1 + \frac{t^2}{d}\right)\left(1 - \frac{0,5}{d}\right)}, \quad d = 2N - 2,$$

где  $d$  – степень свободы;  $N$  – размерность выборок вокселей ателектаза и опухоли. Различия в средней яркости рассматривались как незначимые при стандартном для биомедицинских исследований пороге вероятности  $p < 0,05$ , что, как известно из стандартных статистических таблиц, соответствует  $z < 1,96$ . Для каждого пациента и каждого из 11 размеров выборок вокселей описанная выше процедура тестирования повторялась 200 раз с целью получения устойчивых оценок процента пар выборок  $P_S$ , в которых различия были статистически значимыми, а также среднего значения  $z$ -значимости  $M_Z$ . Заметим, что операция вычисления среднего арифметического величины  $z$  корректна, в то время как для величины значимости  $t$  данная операция корректной не является.

### 2.2. Кластерный анализ

Кластерный анализ выборок вокселей ателектаза и опухоли проводился в двух целях. Во-первых, значение точности классификации, получаемое в результате решения задачи, может служить мерой сходства или различия изучаемых классов. Во-вторых, при использовании алгоритма прямой сегментации ателектаза и опухоли путем классификации вокселей скользящего окна (т. е. без выделения специальных вторичных признаков) значения точности классификации могут непосредственно интерпретироваться как возможная точность сегментации.

В данной работе использовались три различных метода классификации, наиболее часто встречающиеся в современной англоязычной литературе: иерархическое рекурсивное разбиение

ние (recursive partitioning), метод опорных векторов (support vector machines) и так называемый «случайный лес» (random forests).

Иерархическое рекурсивное разбиение, или иерархическая кластеризация [8], заключается в последовательном разбиении исходного множества объектов на основе значений их параметров. Значения относительной важности параметров, полученные в результате тренировки классификатора на обучающей выборке, используются на этапе распознавания объектов как весовые коэффициенты, обеспечивая тем самым очень высокую скорость классификации. Другими преимуществами данного метода являются простота содержательной интерпретации его работы, естественная возможность разбиения исходного множества объектов на более чем два класса и простота встраивания априорных экспертных знаний о предметной области. К общепризнанным недостаткам относится невысокая устойчивость и большая зависимость от состава обучающей выборки. В настоящей работе использовалась программная реализация метода, описанная в работе [9].

Метод опорных векторов был предложен в работе [10] и в настоящее время приобрел такую же популярность, как в свое время нейросетевые алгоритмы. Некоторые исследователи даже полагают, что в силу универсальности используемого подхода к решению общей (нелинейной) задачи классификации, регрессии и обнаружения резко выделяющихся объектов (детектирования «новизны») метод опорных векторов может рассматриваться как универсальное средство решения большинства существующих задач кластерного анализа [11]. Суть метода заключается в поиске оптимальной разделяющей гиперплоскости между двумя классами путем максимизации промежутка между крайними (пограничными) точками классов. При этом точки, лежащие на границах, называются опорными векторами, а центральная плоскость промежутка – оптимальной разделяющей гиперплоскостью. Объекты классов, лежащие не на «своей» стороне гиперплоскости, дискриминируются весовыми коэффициентами так, чтобы уменьшить их влияние (так называемый «мягкий промежуток»). Если линейное разделение найти не удастся, то точки проецируются в пространство более высокой размерности, где они становятся линейно разделимыми. Затем задача кластеризации формулируется как задача квадратичной оптимизации и решается с помощью одного из известных методов. В данной работе была использована программная реализация метода, доступная в виде свободно распространяемой C++ библиотеки LIBSVM, интегрированной в программный пакет-оболочку e1071 [12].

В последние годы наблюдается значительный интерес к разработке и использованию наборов классификаторов, т. е. методов генерации множества классификаторов и агрегирования результатов их работы при решении конкретной задачи. Среди наиболее известных подходов следует отметить приемы так называемого бустинга [13] и баггинга [14] деревьев классификации на этапе обучения. В случае бустинга последовательные поддеревья назначают меньшие веса объектам, которые были некорректно классифицированы предыдущими деревьями, и заключительное решение принимается путем взвешенного голосования деревьев. В случае же баггинга каждое дерево работает независимо на основе случайно выбранного подмножества объектов обучающей выборки (так называемый метод бутстрапа, широко применяющийся в современной статистике [15]). По завершении процесса вопрос отнесения объекта к тому или иному классу решается простым голосованием. В 2001 г. профессор Л. Брейман предложил вариант дальнейшего развития техники баггинга, получивший название «метод случайного леса» [16]. В данном методе добавляется еще один уровень случайности, касающийся конструирования классификационных деревьев. Известно, что в случае стандартного подхода каждая вершина дерева расщепляется с использованием того параметра объекта, который обеспечивает лучшее разделение. В случае же применения метода случайного леса каждая вершина расщепляется с использованием лучшего параметра среди подмножества параметров, выбираемого случайным образом в каждой отдельной вершине. Вычислительные эксперименты, проведенные автором метода и другими исследователями, показали, что метод случайного леса во многих случаях обеспечивает лучшие результаты классификации по сравнению с дискриминантным анализом, нейронными сетями и методом опорных векторов, а также обладает рядом других преимуществ. В настоящей работе использована свободно распространяемая авторская про-

граммная реализация метода, выполненная на языке Фортран и вызываемая в данном случае через пакет-оболочку randomForest [17].

Общая процедура решения задачи классификации выборок вокселей ателектаза и опухоли с использованием описанных выше методов включала:

- формирование обучающего набора данных, состоящего из 100 выборок вокселей заданного размера, взятых случайным образом с участков ателектаза, и 100 выборок вокселей, взятых с участков опухоли трехмерного КТ-изображения одного и того же пациента;
- тренировку классификаторов всех трех описанных выше типов на сформированном обучающем наборе данных;
- формирование тестового набора данных тех же размеров, состоящего из вокселей, взятых с КТ-изображения того же пациента;
- независимое решение задачи классификации 200 выборок текущего тестового набора данных всеми тремя методами и подсчет точности классификации для каждого из них в процентах.

Точность классификации корректировалась с учетом случайного правильного отнесения выборки к одному из двух рассматриваемых классов. Таким образом, значение точности менялось не в пределах от 50 (статистический минимум при совершенно случайном отнесении объекта к одному из двух классов) до 100 % (абсолютно точное распознавание), а в пределах от 0 до 100 %. Описанная выше процедура тестирования повторялась 100 раз для каждого пациента и каждого размера выборки с целью получения устойчивых оценок точности распознавания. Таким образом, общее количество решенных задач классификации составило: 8 пациентов  $\times$  11 типов размеров выборок  $\times$  3 метода  $\times$  100 независимых повторений = 26 400. На персональном компьютере типа Pentium4 1,8 ГГц, 1 Гб RAM для этого потребовалось 7 ч 42 мин.

### 2.3. Визуализация межкластерных различий

Следуя широко распространенной в теории и практике распознавания образов интерпретации, случайные выборки вокселей размерности  $N$ , взятые с участков ателектаза и опухоли, могут рассматриваться как отдельные точки в  $N$ -мерном пространстве признаков. Одним из эффективных способов визуализации распределения объектов в  $N$ -мерном пространстве является метод многомерного шкалирования [18], суть которого заключается в проецировании точек  $N$ -мерного пространства на двумерную плоскость с условными координатами таким образом, чтобы расстояние между произвольной парой точек в  $N$ -мерном пространстве и на плоскости было приблизительно одним и тем же. В данной работе метод многомерного шкалирования (версия, предложенная в работе [19] и реализованная программно в [5, 6]) использовался для визуального анализа различий в яркостных свойствах двух изучаемых классов изображений легкого.

В заключение следует отметить, что наряду с некоторыми частными методами общий подход к исследованию эффективности яркостной информации для разделения ателектаза и опухолей, использованный в настоящей работе, также построен по принципу бутстрапа [15], т. е. многократного повторения процедуры формирования случайной подвыборки и последующего ее анализа.

## 3. Результаты

Результаты статистического оценивания значимости различий в средней яркости между участками ателектаза и раковых опухолей на КТ-изображениях легких восьми пациентов показали, что процент пар выборок  $P_s$ , в которых различия являются статистически значимыми (рис. 2, а), и средняя значимость  $Mz$  (рис. 2, б) варьируют в широких пределах в зависимости от размера выборки и от конкретного пациента. Так, для пациента 5 процент выборок, в которых средние величины яркости ателектаза и опухоли значимо различаются, превышает величину  $P_s = 50\%$  и средняя значимость превышает статистический порог  $Mz = 1,96$  уже при размере выборки  $N=16$  (участок изображения  $4 \times 4$  вокселя). В то же время для пациента 1 аналогичный уровень значимости достигается лишь при размере выборки  $N=225$  вокселей, что соответствует участку слоя КТ-изображения размером  $15 \times 15$  вокселей. Судя по данным, представленным на рис. 2, статистики пациента 7 значительно отличаются от остальных, что говорит о не-

обходимости отдельного изучения возможных причин как технического, так и биомедицинского характера, объясняющих подобное отклонение. В качестве общего итога проведенного статистического оценивания можно отметить тот факт, что размер выборок вокселей, при котором доля пар со статистически значимым различием средней яркости достигает 100 %, составляет около 900 вокселей. Это соответствует достаточно большому участку слоя изображения размером порядка  $30 \times 30$  вокселей.

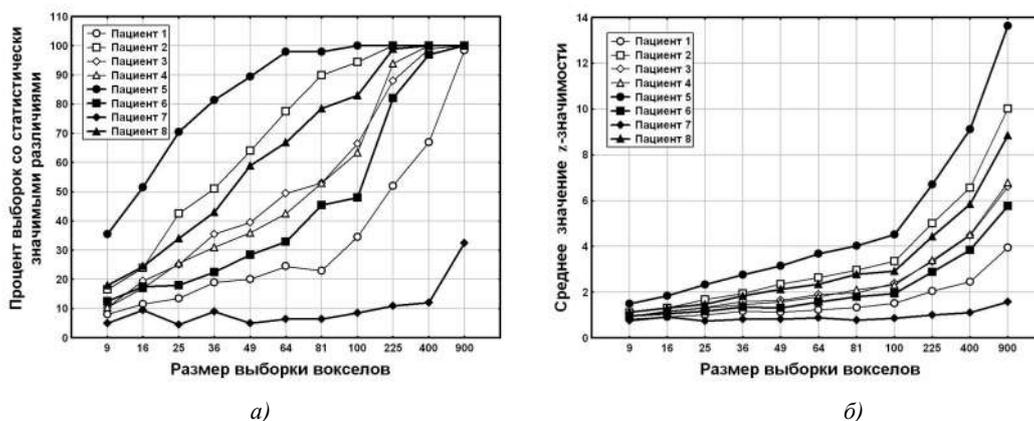


Рис. 2. Статистическая значимость различий яркостных свойств участков ателектаза и опухоли на КТ-изображениях легких восьми пациентов как функция размера выборки вокселей: а) процент пар выборок, в которых различия в средней яркости статистически значимы; б) средняя величина статистической значимости наблюдаемых различий

Результаты классификации (рис. 3) демонстрируют более высокую степень точности классификации и регулярности ее поведения как в зависимости от размера выборки вокселей, так и в зависимости от пациента (за исключением пациента 7). Что касается преимуществ того или иного метода классификации, то в контексте данной работы наиболее перспективным представляется метод случайного леса. Действительно, можно предположить, что при использовании алгоритма прямой сегментации ателектаза и опухоли путем классификации вокселей скользящего окна и назначения соответствующей метки класса его центральному вокселу минимально необходимый процент точности распознавания может быть установлен на уровне порядка 90 % (в допущении, что при неизбежной пространственной коррелированности меток класса оставшиеся 10 % «ошибочных» меток могут быть скорректированы на этапе постобработки любым алгоритмом сглаживания типа «наиболее часто встречающаяся метка» в скользящем окне того же размера). Тогда, как легко видеть на рис. 3, в случае использования классификатора случайного леса минимально необходимый размер окна изображения будет равен  $8 \times 8 = 64$  вокселям, в то время как для оставшихся двух классификаторов заданная точность достигается только при 400 и 225 вокселях (см. рис. 3, а и рис. 3, б соответственно). Учитывая, что аксиальный размер вокселя КТ-изображений в данном исследовании составляет 0,68 мм и точность сегментации не может превышать половины размера скользящего окна, нижняя оценка потенциальной ошибки прямой сегментации внутри слоя составляет 2,72 мм.

Пример использования метода многомерного шкалирования для визуализации распределения выборок вокселей исследуемых классов в  $N$ -мерном пространстве показан на рис. 4. Следует отметить, что этот рисунок иллюстрирует лишь один вариант распределения выборок вокселей для каждого из трех использованных размеров, выбранных случайным образом на изображении пациента 6. Поэтому данные результаты, в отличие от результатов, представленных на рис. 2 и 3, статистически репрезентативными не являются. Тем не менее, метод многомерного шкалирования представляет собой незаменимый инструмент визуализации большого количества многомерных наблюдений.

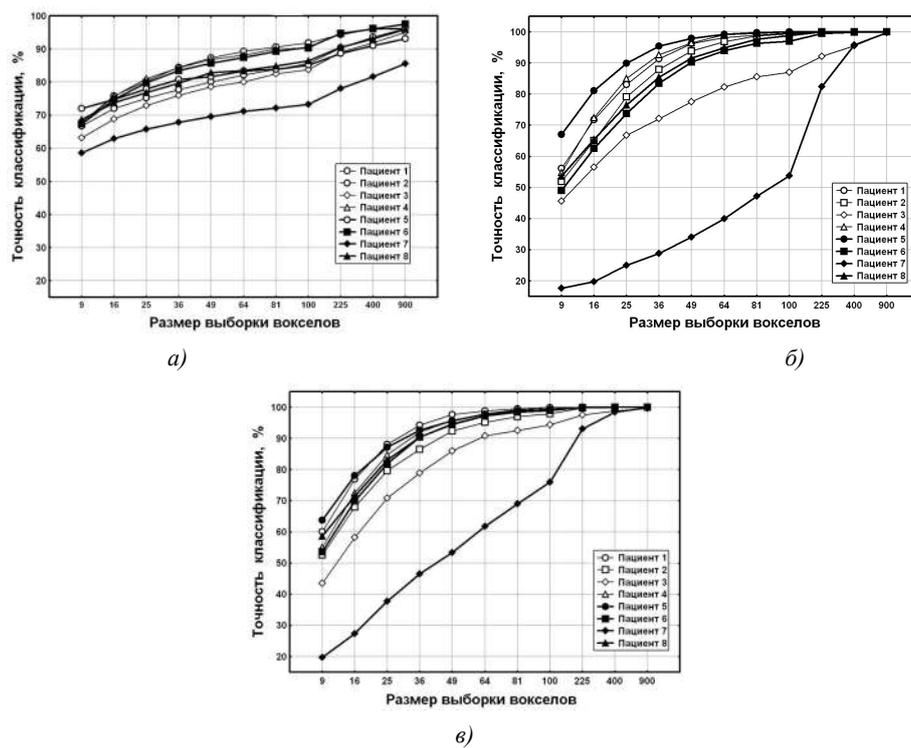


Рис. 3. Точность классификации выборок вокселей, взятых случайным образом из участков ателектаза и опухоли на КТ-изображениях легких восьми пациентов, как функция размера выборки вокселей при использовании методов: а) иерархической кластеризации; б) опорных векторов; в) случайного леса

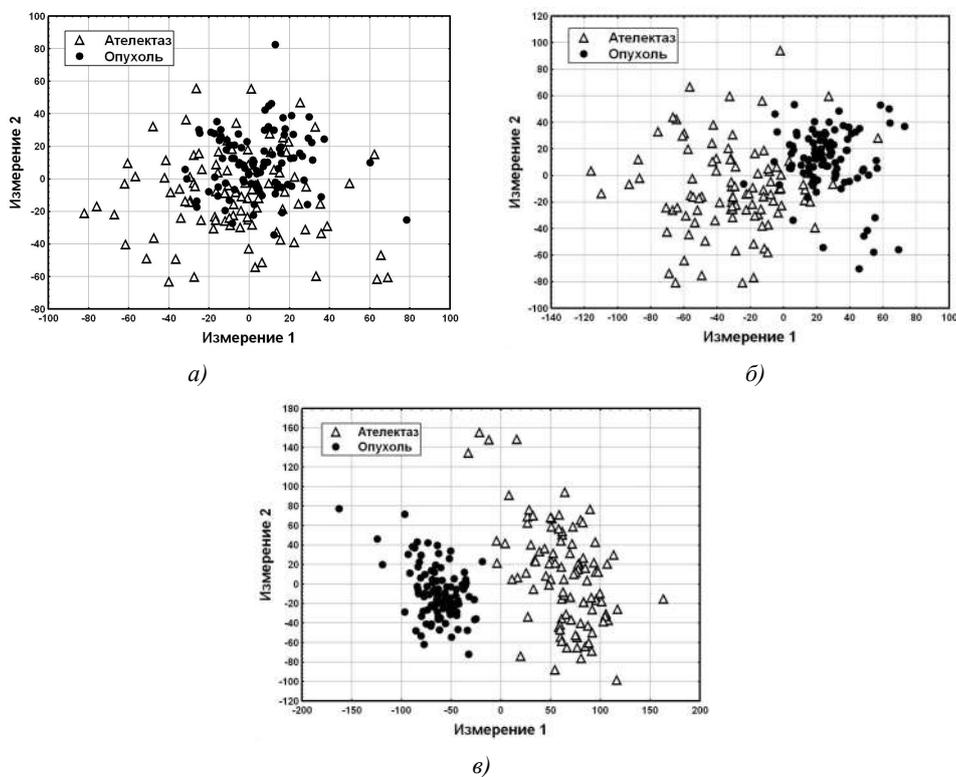


Рис. 4. Пример визуализации разделимости выборок вокселей ателектаза и опухоли в многомерном пространстве признаков, взятых из КТ-изображения легкого пациента 6 при размере выборок: а)  $N=9$  вокселей; б)  $N=36$  вокселей; в)  $N=100$  вокселей. Во всех случаях положения выборок вокселей (точек) в многомерном пространстве (размерности пространств 9, 36 и 100 соответственно) проецировались на плоскость с двумя условными координатами при помощи метода многомерного шкалирования

### Заключение

Результаты исследования, представленные в данной работе, позволяют сделать следующие выводы:

1. Вопреки данным, полученным в работе [1], яркостная информация, предоставляемая КТ-изображениями легкого, несомненно, является полезной для разделения участков ателектаза и раковых опухолей. При условии использования наиболее современных методов классификации процедура распознавания выборок вокселей может быть интегрирована в более общий алгоритм локализации и сегментации участков КТ-изображений указанных двух типов.

2. Степень полезности яркостной информации варьирует от пациента к пациенту и зависит, главным образом, от размера выборки. В работе приводятся детальные количественные данные по статистической значимости и точности распознавания как функции размера выборки.

3. Результаты классификации ателектаза и опухоли значительно меньше зависят от индивидуальных особенностей пациента, чем результаты оценки статистической значимости различия яркостей. Среди протестированных трех классификаторов метод случайного леса является наилучшим.

4. Показано, что на исходных данных, использованных в настоящей работе, нижняя оценка точности сегментации путем прямой классификации вокселей скользящего окна составляет около 2,7 мм при условии использования метода случайного леса и принятом уровне точности классификации 90 %. Однако следует подчеркнуть, что прямая сегментация возможна лишь при условии доступности обучающих выборок вокселей с теми же свойствами, что и участки изображений, подлежащие классификации.

5. Пространственная точность сегментации может быть значительно повышена как за счет увеличения разрешения КТ-изображений в плоскости аксиальных слоев (уменьшения поля сканирования), так и за счет перехода к трехмерному скользящему окну, что возможно при уменьшении толщины слоя до величины, близкой к аксиальному размеру вокселя. Однако непременным предварительным условием является повторение всех проведенных измерений и оценок на представительной тестовой выборке КТ-изображений увеличенного разрешения.

### Список литературы

1. Mira J.G., Fullerton G.D., Ezekiel J. Evaluation of computed tomography numbers for treatment planning of lung cancer // *International Journal of Radiation Oncology, Biology, Physics*. – V. 8. – № 9. – 1982. – P. 1625–1628.

2. Bowden P., Fisher R., Mac Manus M. Measurement of lung tumor volumes using three-dimensional computer planning software // *International Journal of Radiation Oncology, Biology, Physics*. – V. 53 – № 3. – 2002. – P. 566–573.

3. Bradley J., Thorstad W.L., Mutic S. Impact of FDG–PET on radiation therapy volume delineation in non–small–cell lung cancer // *International Journal of Radiation Oncology, Biology, Physics*. – V. 59. – № 1. – 2004. – P. 78–86.

4. Balogh J.B., Caldwell C.B., Ung Y.C. Interobserver variation in contouring gross tumour volume in carcinoma of the lung associated with pneumonitis and atelectasis: The impact of 18FDG–hybrid pet fusion // *International Journal of Radiation Oncology, Biology, Physics*. – V. 48. – № 3. – 2000. – P. 128–129.

5. Venables W.N., Ripley B.D. *Modern applied statistics with S*, 4th edition. – Berlin: Springer, 2002. – 495 p.

6. Maindonald J., Braun J. *Data analysis and graphics using R: an example-based approach*. – Cambridge: Cambridge University Press, 2003. – 400 p.

7. Press W.H., Teukolsky S.A., Vetterling W.T. *Numerical recipes in C: the art of scientific computing*, 2nd edition. – Cambridge: Cambridge University Press, 2002. – 994 p.

8. Breiman L.J., Friedman H., Olshen R.A. *Classification and regression trees*. – Belmont: Chapman & Hall (reprinted by First CRC Press), 1984. – 368 p.

9. Therneau T.M., Atkinson B. *Rpart: recursive partitioning in R, package version 3.1-22*. – Rochester: Mayo Clinic, College of Medicine, Division of Biostatistics, 2005. – 28 p.

10. Cortes C., Vapnik V. Support-vector network // *Machine Learning*. – V. 20. – № 3. – 1995. – P. 273–297.

11. Bennett K.P., Campbell C. Support vector machines: Hype or Hallelujah? // ACM Special Interest Group on Knowledge Discovery and Data Mining, Explorations. – V. 2. – № 2. – 2000. – P. 1–13.
12. Dimitriadou E., Hornik K., Leisch F. E1071: Miscellaneous functions of the department of statistics, R package version 1.5-7. – TU Wien, Austria, 2005. – 58 p.
13. Shapire R., Freund Y., Bartlett P. Boosting the margin: A new explanation for the effectiveness of voting methods // The Annals of Statistics. – V. 26. – № 5. – 1998. – P. 1651–1686.
14. Breiman L. Bagging predictors // Machine Learning. – V. 24. – № 2. – 1996. – P. 123–140.
15. Wehrens R., Putter H., Buydens L.M.C. The bootstrap: a tutorial // Chemometrics and intelligent laboratory systems. – V. 54. – 2000. – P. 35–52.
16. Breiman L. Random Forests // Machine Learning. – V. 45. – № 1. – 2001. – P. 5–32.
17. Breiman L., Cutler A., Wiener M. randomForest: Breiman and Cutler's random forests for classification and regression, R package version 4.5-4, 2005. – 25 p.
18. Cox T.F., Cox M.A. Multidimensional scaling. – London: Chapman & Hall, 1994. – 213 p.
19. Cailliez F. The analytical solution of the additive constant problem // Psychometrika. – V. 48. – 1983. – P. 343–349.

Поступила 06.04.05

<sup>1</sup>Университет Суррей Гилфорд,  
Великобритания  
e-mail: v.kovalev@surrey.ac.uk

<sup>2</sup>ГУ «НИИ онкологии и медицинской радиологии им. Н.Н. Александрова»,  
Белорусская медицинская академия последипломного образования,  
п/о Лесной-2, Минск  
e-mail: skhoruzhik@nld.by

<sup>3</sup>Отделение медицинских наук НАН Беларуси,  
Белорусская медицинская академия последипломного образования,  
Минск, П. Бровка, 3

**V.A. Kovalev, S.A. Khoruzhik, A.N. Mikhailov**

**STUDYING THE EFFICIENCY OF INTENSITY INFORMATION  
FOR DISCRIMINATING ATELECTASIS AND CANCER TUMOUR REGIONS  
IN THE COMPUTED TOMOGRAPHY IMAGES OF LUNG**

The purpose of this study was to evaluate the potential usefulness of the intensity information of lung CT images for discriminating atelectasis and cancer tumour regions, which are known to be hardly distinguishable in visual examination. Results of the statistical assessment of the significance of intensity differences as well as the classification accuracy, achieved with the help of hierarchical clustering, support vector machines, and random forests methods, are reported. It was demonstrated that the usefulness of the intensity information varies from patient to patient and depends mostly upon the voxel sample size.