

УДК 621.315.6

ПРОГНОЗИРОВАНИЕ РЕЦИДИВА РАКА ЯИЧНИКОВ НА ОСНОВЕ КОМПЬЮТЕР-АССИСТИРОВАННОГО МОРФОМЕТРИЧЕСКОГО ИССЛЕДОВАНИЯ

М.В. СПРИНДЖУК, А.В. ШУКЕЛОВИЧ, А.П. КОНЧИЦ

*Объединенный институт проблем информатики НАН Беларуси,
Сурганова, 6, Минск, 220012, Беларусь*

Поступила в редакцию 5 июля 2012

Разработана методика морфометрического исследования на основе обработки изображений с помощью оригинально сконструированного программного обеспечения. Для обоснования внедрения таких методов в клиническую практику необходимо выполнить широкомасштабные исследования, где объектами изучения будут опухоли на ранних стадиях развития.

Ключевые слова: ангиогенез, лимфангиогенез, обработка изображений, метод опорных векторов, графы случайного леса, корреляция.

Введение

Целями исследования были изучение проблемы морфометрии лимфангиогенеза (роста лимфатических капилляров) применением алгоритмов анализа гистологических изображений [1–3] и выяснение возможностей морфометрии для предсказания факта рецидива рака яичников по данным обработки патогистологических изображений, окрашенных различными гистологическими маркерами. Исследование выполнялось в рамках международного проекта МНТЦ (Международный научно-технический центр) № 51682.

Материалы и методы исследования

Исследование выполнялось по материалам Онкологического Диспансера г. Минска в течение 33 месяцев с 2009 по 2012 годы. Препараты были приготовлены из хирургически удаленных опухолей. Образцы опухолей хранились в замороженном виде. Препараты окрашивались маркерами D2-40, CD31, Ki-67. Изображения были получены при помощи микроскопа «Leica DMD 108» с операционной системой Linux. Снимки были получены вручную, без сканера, 150 изображений, которые в основном визуально соответствовали области максимальной васкуляризации по зрительному оценочному решению одного исследователя, как квадрат на увеличении $\times 40$, а затем сняты $\times 200$. После этого были отобраны изображения, исключительно содержащие околоопухолевую зону (10 шт.). Все изображения были сохранены в формат Tiff, приемлемый для программного обеспечения. В исследовании применялось два программных пакета. Первое программное обеспечение было написано на языке Python и использовало библиотеки функций Scipy [4], Mahotas [5] и OpenCV [6]. Скрипт-программа была перемещена в каждую папку с изображениями командами Windows CMD, а затем автоматически запускалась программой на языке Linux Bash (см. листинг 1–4). Второй продукт, написанный на Delphi, вычислял значение цветности, доли мелких объектов сосудистой сети, доли крупных объектов сосудистой сети, относительной площади, занимаемой сосудами, и однородности сосудов для изображений, прокрашенных маркером D2-40. Программное обеспечение произвело список числовых величин, который был импортирован в таблицу CSV специальной программой, написанной на языке Java. Статистическое исследование выполнялось преимущественно программ-

ным пакетом STATISTICA 7.0. и пакетами R (см. листинг 3). Его основная концепция заключалась в моделях случайного леса, опорных векторов, дискриминантного анализа, где параметрам изображений придавалось значение предсказательных факторов, а факту наличия рецидива карциномы яичников – исхода.

Обсуждение результатов исследования

Алгоритмы обработки изображений текстуры Харалика показали относительную наибольшую предиктивную ценность для маркера CD31. Наше исследование выявило сравнительные различия в точностях классификации методом опорных векторов и дискриминантным анализом при применении различных гистохимических маркеров и параметров-свойств изображений. Мы обнаружили значительные предсказательные возможности нескольких признаков Харалика (Haralick) (см. рис 1-2), посчитанных в различных векторах пространства изображений (для маркера Ki-67 – Haralick 9-й признак, 7-ой вектор, CD31 – Haralick_8_8 и для D240 – Haralick_7_10, соответственно). Числовые значения моментов X_u , полученных в результате обработки изображений лимфангиогенеза, показали низкую точность модели классификации. Наибольшая точность классификации модели опорных векторов была выявлена для группы опухолевых образцов, окрашенных маркером CD31. Ki-67 маркер с первым параметром Haralick показал наименьшую неправильную классификацию дискриминантного анализа (см. таблицу).

Проблема точности морфометрии – относительная случайность выбора участка опухоли и различная репрезентативность разных ее частей. Вторая трудность – субъективный отбор изображений с максимальным количеством сосудов. Третья проблема заключается в обоснованном подборе алгоритмов для обработки изображений. Чтобы программное обеспечение не было похоже на генератор случайных чисел, необходимо применять алгоритмы, имеющие определенный патоморфологический смысл, и желательно имплементировать математическую основу алгоритма в виде кода самому и тестировать программу на предмет одинаковых результатов несколько раз.

Моменты X_u оказались несостоятельными при оценке их прогностической значимости методами опорных векторов, случайного леса, дискриминантного анализа, бинарной логистической регрессии и корреляции Спирмана. Первичное вычисление прогностического значения признаков Харалика в отношении предсказания факта рецидива рака яичников показало положительный результат, но только при применении методов случайного леса, опорных векторов и дискриминантного анализа. Расчеты проводились для усредненных значений признаков в различных векторах. Однако применение корреляционных моделей логистической и пошаговой регрессий не показали значимой предсказательной эффективности параметров Харалика ($p > 0,05$). Таким образом, основной результат настоящего исследования – выявление стабильного прогностического параметра изображений, которым оказалась однородность сосудистой сети лимфатических капилляров (см. рис. 3: Результаты получены на основе вычисления модели графов случайного леса (random forest analysis, переменные подписаны слева направо).

Значения этого фактора показали свою прогностическую значимость при применении различных статистических подходов, что свидетельствует о том, что этот морфометрический признак заслуживает дальнейшего изучения и развития в виде разработки программного обеспечения.

Листинг 1. Программа вычисления признаков Харалика на языке Python.

```
import mahotas
import glob
from scipy import ndimage
import numpy as np
import scipy
import pylab
import pymorph
import mahotas
import numpy
```

```
f = open('batchoutput.csv', 'w') # открываем файл для записи результата вычисления
for filename in glob.iglob ('*.tif'):
img = mahotas.imread (filename)
cancer = mahotas.features.haralick (img)
srednee = sum(cancer)/len(cancer) )# считает среднее от результата
print srednee
f.write(str(srednee))
f.close()
```

Листинг 2. Программа вычисления моментов Ху.

```
import opencv
import cv
import glob
import numpy
f = open('momentsbatchoutput.csv', 'w') # открываем файл для записи результата вычисления
for filename in glob.iglob ('*.tif'):
img = cv.LoadImageM(filename, cv.CV_LOAD_IMAGE_GRAYSCALE)# прочтение изображения
cancer = cv.GetHuMoments(cv.Moments(img)) # вызов функции обработки изображений
arr = numpy.array([cancer])
srednee = numpy.mean(arr)# вычисление среднего
print srednee
f.write(str(srednee))
f.close()
```

Листинг 3. Пример кода для вычисления классификации метода опорных векторов (язык R, библиотека функций e1071) [7, 8].

```
bcdata <- read.table ('C:\\Documents and Settings\\Matvey\\Рабочий
стол\\SVM_R\\ovcd.tab',sep = "\\t",dec = ",",header=TRUE)
names(bcdata)
show(bcdata)
library(e1071)
# Исключаем столбцы с ФИО
bcdata <- bcdata[, -c(9, 10, 11)]
databcall <- subset(bcdata,select=c(-Homogen,-Relapse))
classesbcall <- subset(bcdata,select = Relapse)
databctrain <- databcall[1:55,]
# Тут указываем, что номер класса - это фактор
classesbctrain <- factor(classesbcall[1:55,])
databctest <- databcall[56:81,]
# Для классов в тестовой выборке тоже делаем их факторами
classesbctest <- classesbcall[56:81,]
model <- svm(databctrain, classesbctrain)
pred <- predict(model, databctest)
table(pred,t(classesbctest))
```

Листинг 4. Код Linux Bash для автоматического запуска скрипта обработки изображений.

```
ls -ld /media/SAMSUNG/Hot_spots_ovarian_cancer_2012/*/* | while read i
do
pushd "$i"
python ./Sof1.py # это название скрипта
popd
done
```

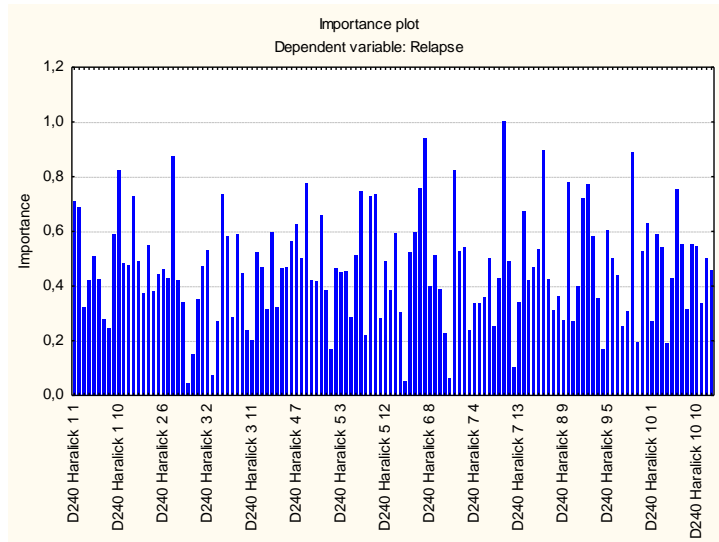


Рис. 1. Сравнительный предикторный вес значений параметров Харалика для маркера D240

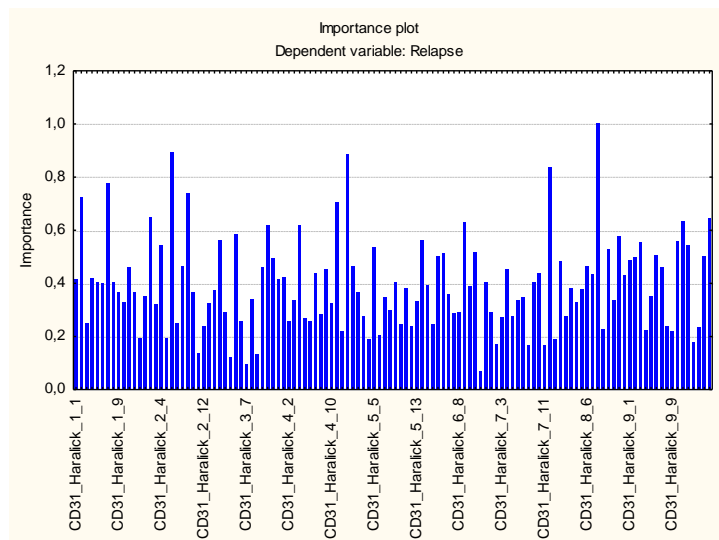


Рис. 2. Сравнительный предикторный вес значений параметров Харалика для маркера CD31

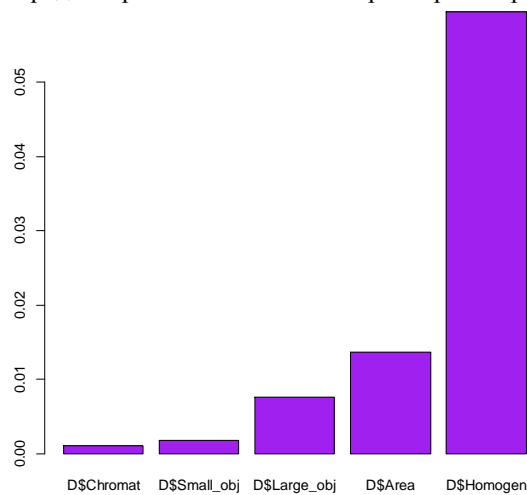


Рис. 3. График, представляющий сравнительное прогностическое значение цветности, доли мелких объектов сосудистой сети, доли крупных объектов сосудистой сети, относительной площади, занимаемой сосудами, и однородности сосудов

Таблица дискриминантного анализа для предсказания факта рецидива рака яичников (указано название маркера и номер признака Харалика)

Параметр-свойство	Количество неправильно классифицированных	Процент неправильно классифицированных	2LogXВероятность
D240__1	31	33,7	54,77
D240__2	29	31,52	56,46
D240__3	35	38,04	58,18
D240__4	35	38,04	55,45
D240__5	33	36,26	53,02
D240__6	27	30	53,47
D240__7	34	38,2	56,13
D240__8	32	35,6	55,95
D240__9	32	35,96	55,95
D240__10	28	32,56	52,67
CD31__1	23	27,38	49,28
CD31__2	19	22,62	48,05
CD31__3	28	33,33	47,68
CD31__4	26	30,95	51,22
CD31__5	27	32,14	45,53
CD31__6	29	34,94	50,44
CD31__7	26	32,1	49,2
CD31__8	26	32,1	44,57
CD31__9	26	32,1	49,2
CD31__10	27	33,75	50,97
Ki67__1	19***	24,68	41,32
Ki67__2	25	32,47	44,23
Ki67__3	25	32,47	47,4
Ki67__4	28	36,36	47,54
Ki67__5	28	36,36	46,71
Ki67__6	30	38,96	45,54
Ki67__7	26	34,21	45,08
Ki67__8	21	28,38	38,34
Ki67__9	21	28,77	42,54
Ki67__10	22	30,14	40,27

Заключение

Мы получили свидетельства того, что применение алгоритмов анализа изображений может иметь прогностическое значение для предсказания рецидива рака яичников. Также мы выявили факт, что окраска опухолей различными маркерами приводит к неодинаковым численным значениям параметров изображений. Кроме того, применив различные прогностические статистические модели, мы выявили неодинаковую стабильность прогностических факторов. Однако ограничениями нашего исследования было число пациентов (в двух выборках – 72 и 81 опухолевых образцов), изображений, и тот факт, что целесообразнее исследовать пациентов на ранних стадиях исследования.

PROGNOSIS OF THE RELAPSE OF OVARIAN CANCER BASED ON THE COMPUTER-ASSISTED MORPHOMETRIC STUDY OF OVARIAN CARCINOMA SAMPLES STAINED WITH VARIOUS HISTOCHEMICAL MARKERS

M.V. SPRINDZUK, A.V. SHUKELOVICH, A.P. KONCHITS

Abstract

An experience in lymphangiogenesis morphometry applying their original software is described. For the implementation in clinical practice protocols, it is necessary to carry out a large-scale studies, where the objects of research are in the earliest stages of pathology.

Список литературы

1. Cruz-Roa A., Caicedo J.C., Gonzalez F.A. // *Artif Intell Med*. 2011. № 2. P. 91–106.
2. Di Ieva A. // *Microvascular Research*. 2010. № 3. P. 522–533.
3. Wootton R., Springall D.R., Polak J.M. *Image analysis in histology: Conventional and confocal microscopy*. Cambridge, 1995.
4. Библиотека функций Scipy. [Электронный ресурс]. – Режим доступа: <http://numpy.scipy.org/>. – Дата доступа: 01.09.2012.
5. Библиотека функций Mahotas. [Электронный ресурс]. – Режим доступа: <http://pyri.python.org/pyri/mahotas>. – Дата доступа: 01.09.2012.
6. Библиотека функций OpenCV. [Электронный ресурс]. – Режим доступа: <http://opencv.willowgarage.com/wiki/PythonInterface>. – Дата доступа: 01.09.2012.
7. Пакет языка R для вычисления опорных векторов. [Электронный ресурс]. – Режим доступа: <http://cran.rproject.org/web/packages/e1071/index.html/>. – Дата доступа: 01.09.2012.
8. Ресурсы языка R. [Электронный ресурс]. – Режим доступа: <http://www.r-project.org>. – Дата доступа: 01.09.2012.