

# Інформаційні технології в медицині

УДК 621.57.673

Е.В. Высоцкая, Н.А. Щукин, А.П. Порван, С.Н. Пушкарь

*Харьковский национальный университет радиоэлектроники, Харьков*

## ИСПОЛЬЗОВАНИЕ ДИСКРИМИНАНТОГО АНАЛИЗА ДЛЯ ДИАГНОСТИКИ ДОБРОКАЧЕСТВЕННЫХ И ЗЛОКАЧЕСТВЕННЫХ ОПУХОЛЕЙ

*В статье авторами предлагается диагностическая модель дифференцирования доброкачественной опухоли и злокачественного процесса, полученная в результате дискриминантного анализа данных клинических, инструментальных и лабораторных исследований. Полученные дискриминантные функции используются при построении территориальной карты дифференциации доброкачественной и злокачественной опухоли, а статистические оценки синтезированных моделей показывают высокую значимость.*

**Ключевые слова:** дискриминантный анализ, злокачественная опухоль, территориальная карта.

### Анализ литературы и постановка задачи

В Украине каждый год обнаруживается более 160 тыс. новых случаев злокачественных новообразований, почти 100 тыс. жителей умирают от рака, причем 35% умерших — лица трудоспособного возраста. Каждый час регистрируется более 20 новых случаев заболевания, а 10 жителей Украины умирают от рака. По расчетам специалистов, до 2020 года количество впервые заболевших раком в Украине превысит 200 тыс. в год. В 2003 году уровень онкозаболеваемости в целом по Украине составил 325 человек на 100 тыс. Что касается показателей смертности от онкологических заболеваний, то каждый второй-третий онкобольной в Украине умирает в первый год болезни, что в 2-10 раз превосходит аналогичный показатель в развитых странах. От 60% до 80% онкологических больных обращаются за медицинской помощью с наличием запущенных (III-IV) стадий заболевания, когда возможности радикального лечения весьма ограничены. Поэтому, несмотря на разработку и применение различных методов лечения, в течение одного года после установления диагноза умирают около 50% пациентов. Таким образом, для населения Украины злокачественные опухоли являются острой медицинской и социальной проблемой.

Одним из приоритетных направлений в борьбе со злокачественными опухолями является разработка принципиально новых методов ранней диагностики злокачественных новообразований.

В связи с вышесказанным, разработка метода диагностики доброкачественных и злокачественных опухолей на ранних стадиях является актуальной научной и практической задачей.

На сегодняшний день известно большое количество методов диагностики злокачественных опухолей, которые можно разделить на методы визуализации и лабораторно-аналитические методы.

Среди методов визуализации наиболее часто применяют:

- рентгенологические;
- магнитно-резонансные;
- радионуклидные;
- ультразвуковые.

Ультразвуковой метод диагностики может дать достаточно большой объем информации об опухолевом процессе – вовлечение в опухолевый процесс сосудов и лимфоузлов, определение характера опухоли, ее плотности, размеров, выявление метастазов.

Для определения распространенности процесса, размеров опухоли и некоторых других параметров используют обзорные снимки: рентгенография грудной клетки в двух проекциях, обзорная рентгенография брюшной полости, рентгенография конечностей, черепа, отдельных костей. Иногда используют внутривенную урографию (например, при опухоли Вилмса).

Наиболее информативным рентгенологическим методом является компьютерная томография (КТ), с помощью которой можно оценить множество параметров опухолевого роста, касающихся локализации, размеров, характера роста, наличия метастазов. В детской онкологии КТ показана при выявлении мелких метастазов, а потому она имеет ценность при обследовании пациентов с герминогенными опухолями, саркомами, опухолями печени, опухолью Вилмса. Благодаря высокой разрешающей способности, снижению дозовых нагрузок в современных аппаратах, КТ также применяется для определения эффективности лечения.

Магниторезонансная томография (МРТ) столь же эффективный и информативный метод визуализации, как КТ. В отличие от последнего, имеет свои преимущества и недостатки. МРТ слабоэффективна при выявлении костных опухолей, опухолей задней черепной ямки, основания черепа. Однако опухоли мягких тканей визуализируются весьма контрастно и порой лучше, чем при КТ. МРТ, также как КТ, достаточно часто используется с применением контрастных веществ, повышающих чувствительность метода.

Радиоизотопные методы диагностики применяются преимущественно для выявления костных, лимфопролиферативных опухолей, нейробластом, а также для проведения некоторых функциональных проб.

Методы визуализации имеют возможность детально исследовать структуру опухоли, состояние ангиоархитектоники и другие параметры, однако не дают возможности уверенно говорить о злокачественности и могут служить лишь вспомогательным методом диагностики опухоли [1].

Среди лабораторно-аналитических методов, которые применяются при диагностике опухоли, наибольшего внимания заслуживают:

- метод радиоиммунологического анализа (РИА);
- анализ крови (общий, биохимический, на онкомаркеры);
- биопсия;
- гистология.

Биопсия и гистология на сегодняшний день являются основными методами для верификации опухолевого процесса, имеющие наибольшую диагностическую точность. Однако к недостаткам данных методов можно отнести: инвазивность, возможное инфицирование исследуемой опухоли и близлежащих органов, травматичность и зависимость качества пробы от квалификации лаборанта.

Радиоиммунологический метод основан на том, что при всех видах злокачественных опухолей в крови образуется неспецифический комплекс онкомаркеров. Низкая иммуногенность белков комплекса обеспечивает их накопление в крови на самых ранних стадиях заболевания.

Количество неспецифических онкогенов в крови можно оценивать по порядку величины соответствующего количества специфических онкогенов. На ранних стадиях заболевания концентрация онкогенов столь мала, что не может быть обнаружена оптическими методами. Комплекс онкогенов взаимодействует с нормальными макромолекулами сыворотки крови и вызывает в них специфические изменения.

Однако данный метод имеет следующий недостаток: раково-эмбриональный антиген определяется

в сыворотке крови только у 80% больных, не позволяя при этом в каждом пятом случае выявить патологию. Поэтому определение титра раково-эмбрионального антигена при первичных злокачественных новообразованиях не имеет существенного значения для диагноза и служит лишь дополнительным тестом для диагностики первичной опухоли. При злокачественных опухолях желудка, поджелудочной железы, толстой и прямой кишки, печени, молочной железы и легких повышенный уровень раково-эмбрионального антигена в основном используется как диагностический критерий лишь при рецидиве опухоли [2].

Нами предлагается использовать цитоморфолого-биофизическую диагностику опухолевого процесса, которая также относится к лабораторно-аналитическим методам исследований и позволяет диагностировать опухоль по здоровым, не участвующим в патологическом процессе, популяциям клеток базального эпителия, установить связь между больными органами и цитоморфолого-биофизическими показателями этих клеток не инвазивно, исключая возможность инфицирования и травм [3].

Особенностью цитоморфологической диагностики, является то, что при диагностике опухолевого процесса врач выстраивает последовательность умозаключений, опирающуюся на представлениях о связи цитоморфолого-биофизических признаков, наблюдаемых у больного, с онкологическими заболеваниями. Проведение цитоморфолого-биофизической диагностики опухолевого процесса без применения вычислительной техники представляет собой очень трудоемкую задачу, а также занимает большое количество времени, так как основным механизмом является сравнение и анализ большого количества цитоморфолого-биофизических параметров пациента с эталонными параметрами онкологических заболеваний.

Для однозначного определения информативных признаков, характеризующих тот или иной диагностический процесс на основе законов распределения и значений вероятностей присутствия, и, соответственно, построения адекватной модели могут использоваться методы регрессионного, кластерного, дискриминантного анализов [4].

Используя методы регрессионного анализа, можно достаточно точно выявить факторы, оказывающие существенное влияние на результирующий показатель, а также определить силу и направление этого влияния. Однако синтезированные модели могут оказаться неточными или, в случае сложных моделей, переобученными, что снижает эффективность их использования [5, 6].

Применение методов кластерного анализа позволяет сопоставить объекты моделирования по их качественным характеристикам, агрегировать экс-

пертные оценки текущего и прогнозируемого уровня развития объектов и т.д. Несмотря на свои широкие возможности, использование данного метода анализа требует значительной затраты времени и сил для проведения соответствующих расчетов, а полученные результаты группирования не всегда адекватно интерпретируются, особенно если информативные признаки, характеризующие объекты исследования, были достаточно разнородны [7].

Указанные выше недостатки позволяет учесть дискриминантный анализ, дающий возможность быстро и качественно классифицировать исследуемые объекты и синтезировать адекватную линейную математическую модель. Данный метод математического анализа имеет ряд преимуществ: учитывает вариативность параметров, рассматривает совокупность всех показателей, взятых со своими коэффициентами, которые отражают удельный вес влияния каждого показателя на постановку диагноза [8, 9, 10].

**Целью статьи** является применение дискриминантного анализа для диагностики злокачественной опухоли на ранних стадиях развития.

### Основной материал

При проведении дискриминантного анализа нами были проанализированы данные 150 пациентов с доброкачественными и злокачественными опухолями. Все пациенты были разделены на 3 группы по 50 пациентов в каждой:

- 1-я группа – пациенты с доброкачественными опухолями;
- 2-я группа – пациенты со злокачественными опухолями;
- 3-я контрольная группа –практически здоровые пациенты.

С математической точки зрения все пациенты рассматривались как совокупность объектов с варьирующимися количественными характеристиками. На основании этих характеристик определялась группа, к которой относится объект. Это позволило нам спрогнозировать группы (наличие злокачественной опухоли) для новых объектов из той же совокупности.

Для определения значимых для формирования диагноза свойств, были проанализированы следующие цитоморфологобиофизические показатели: форма клеток и ядер клеток, целостность мембраны клеток и мембран ядер клеток, окрас цитоплазм клеток, окрас ядер клеток, включения в цитоплазму клеток, характер движения ядер, амплитуда движения ядер. Все показатели были закодированы и поставлены в соответствии 41-мерному вектору, что учитывает отсутствие, наличие, направленность и величину каждого показателя. Задачу дискриминации желательно решать с использованием минимального количества функций. Количество функций в каждом конкретном случае зависит от конфигурации классов в многомерном пространстве дискри-

минантных переменных. Чем сложнее конфигурация, тем больше функций необходимо для их распределения и анализа. Функции строили таким образом, чтобы их средние значения для разных классов наиболее различались. При этом совокупность функций образовывала ортогональное пространство, то есть функции были независимыми друг от друга.

Математическая обработка результатов проводилась на персональном компьютере с использованием программ *Microsoft Excel 2007* и *SPSS Statistics 17.0*.

Расчет методом дискриминантных функций значений диагностических коэффициентов позволил выделить 13 значимых для определения диагноза показателей из 41 рассматриваемого (табл. 1).

Таблица 1

Канонические коэффициенты дискриминантных функций для значимых параметров при диагностике доброкачественных и злокачественных опухолей

Признак (%)	Канонические коэффициенты	
	1 функция	2 функция
1	2	3
Форма клетки овальная	0,413	0,31
Форма клетки удлинённая	-0,03	0,303
Мембрана клетки целая	1,070	0,890
Мембрана клетки расслоённая	1,325	0,941
Мембрана клетки разрушена	-0,243	2,960
Мембрана клетки уплотнена	0,156	0,296
Форма ядра клетки овальная	-1,028	2,687
Форма ядра клетки округлая	0,345	0,043
Мембрана ядра клетки целая	0,076	-0,588
Окрас ядра клетки серый	0,357	0,017
Окрас ядра клетки зернистый	0,342	0,37
Характер движения ядер в норме	0,236	0,311
Амплитуда движения ядер низкая	0,78	0,319
Константа	-19,657	-19,188

Таким образом, опухолевый процесс была описан следующими дискриминантными функциями:

$$DF1 = 0,413 \cdot X1 - 0,03 \cdot X2 + 1,070 \cdot X3 + 1,325 \cdot X4 - 0,243 \cdot X5 + 0,156 \cdot X6 - 1,43 \cdot X7 + 0,345 \cdot X8 + 0,076 \cdot X9 + 0,357 \cdot X10 + 0,342 \cdot X11 + 0,236 \cdot X12 + 0,78 \cdot X13 - 19,657;$$

$$DF2 = 0,31 \cdot X1 + 0,303 \cdot X2 + 0,890 \cdot X3 + \\ + 0,941 \cdot X4 + 0,019 \cdot X5 + 0,296 \cdot X6 + 0,073 \cdot X7 + \\ + 0,043 \cdot X8 - 0,588 \cdot X9 + 0,017 \cdot X10 + 0,37 \cdot X11 + \\ + 0,311 \cdot X12 + 0,319 \cdot X13 - 19,188.$$

где X1– форма клетки овальная (%);  
X2– форма клетки удлинённая (%);  
X3– мембрана клетки целая (%);  
X4– мембрана клетки расслоённая (%);  
X5– мембрана клетки разрушена (%);  
X6– мембрана клетки уплотнена (%);  
X7 – форма ядра клетки овальная (%);  
X8 – форма ядра клетки округлая (%);  
X9– мембрана ядра клетки целая (%);  
X10– окрас ядра клетки серый (%);  
X11– окрас ядра клетки зернистый (%);  
X12– характер движения ядер в норме (%);  
X13– амплитуда движения ядер низкая (%).

Оценка меры удачного распределения на группы, полезность дискриминантных функций и их количество функций, были оценены при помощи коэффициентов канонической корреляции (табл. 2).

Таблица 2  
Характеристика дискриминантных возможностей функций

Функция	1	2
Собственное значение	431,014	5,238
% дисперсии	98,8	1,2
Общий %	98,8	100,0
Каноническая корреляция	0,999	0,916

Величина собственного значения связана с дискриминирующими возможностями функции: чем больше собственное значение, тем лучше различие. Таким образом, первая функция обладает наибольшими дискриминирующими возможностями, вторая – обеспечивает максимальное различие после первой.

Чтобы облегчить такое сравнение, собственным значениям приписывают процент дисперсии. Для этого сначала суммируют все собственные значения, затем делят каждое собственное значение на общую сумму. Так, в приведенной системе уравнений первая функция содержит 98,8 % общих дискриминантных возможностей.

Учитывая значение канонического коэффициента корреляции первой функции 0,999, можно сделать вывод о существовании высокой связи между диагнозом и значениями первой дискриминантной функции. Также нами была обнаружена высокая связь между диагнозом и значениями второй дискриминантной функции (0,916). Качественная оценка силы связи  $r_{xy}$  величин X и Y была сделана на основании шкалы Чеддока (табл. 3).

Таблица 3

Шкала Чеддока

Теснота связи	$r_{xy}$ при прямой связи	$r_{xy}$ при обратной связи
Слабая	0,1 – 0,3	(-0,1) – (-0,3)
Умеренная	0,3 – 0,5	(-0,3) – (-0,5)
Заметная	0,5 – 0,7	(-0,5) – (-0,7)
Высокая	0,7 – 0,9	(-0,7) – (-0,9)
Весьма высокая	0,9 – 0,99	(-0,9) – (-0,99)

Оценка значимости дискриминантных функций была проверена  $\lambda$  – статистикой Уилкса (табл. 4).

Таблица 4

Определение значимости дискриминантных функций методом  $\lambda$  – статистики Уилкса

Тест функций	Лямбда Уилкса	$\chi^2$ - квадрат	Степени свободы, df	Значимость
1	0,000	1113,769	26	0,000
2	0,160	258,116	12	0,000

Число степеней свободы k-й дискриминантной функции зависит от количества дискриминантных переменных p и количества классов g:

$$df_k = (p - (k - 1)) \cdot (g - (k - 1) - 1).$$

Тогда:

$$df_1 = (26 - (1 - 1)) \cdot (3 - (1 - 1) - 1) = 26,$$

$$df_2 = (13 - (2 - 1)) \cdot (3 - (2 - 1) - 1) = 12.$$

Так как  $\chi^2$  больше критического, то можно говорить о достоверном различии между группами ( $P < 0,001$ ).

На основании значений обеих дискриминантных функций построена территориальная карта распределения больных в зависимости от диагноза (рис. 1), где 1 – группа пациентов, которые имеют доброкачественную опухоль; 2 – группа пациентов, которые имеют злокачественную опухоль; 3 – группа здоровых пациентов; \* – центр группы.

Анализ полученных данных показал, что из 50 пациентов первой группы двое были ошибочно отнесены ко второй группе. Из 50 пациентов второй группы один ошибочно был отнесен к первой. Все пациенты третьей группы были определены верно.

Таким образом, разработанная математическая модель корректно классифицирует 98% пациентов с доброкачественными и злокачественными опухолями.

## Выводы

Применение дискриминантного анализа позволило классифицировать пациентов с доброкачественными и злокачественными опухолями.

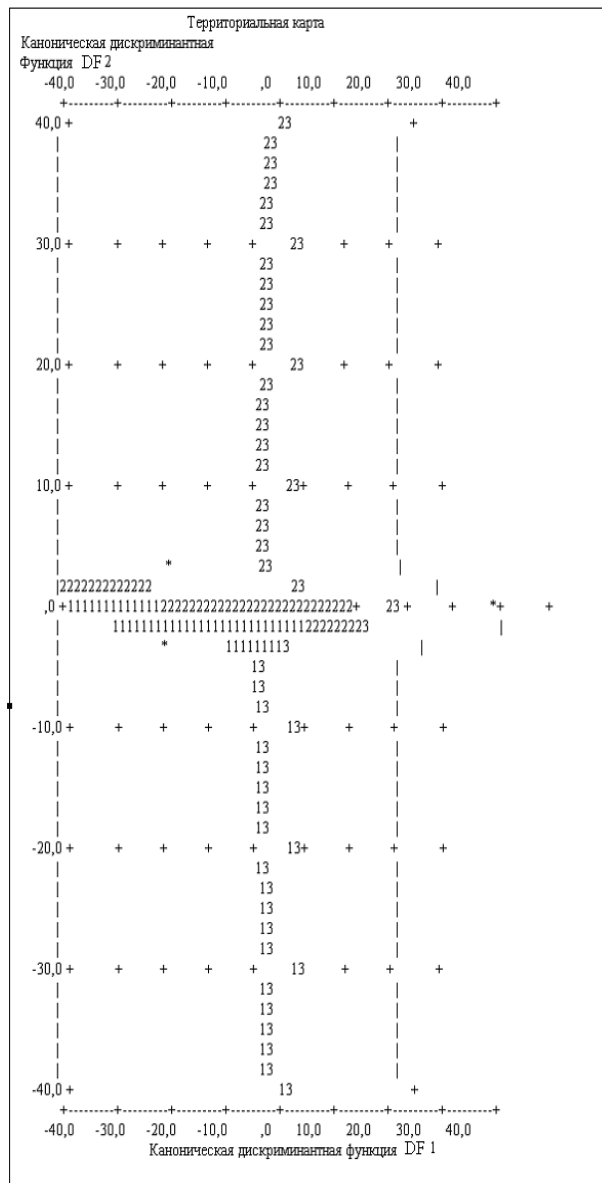


Рис. 1. Территориальная карта распределения больных в зависимости от диагноза

Впервые были использованы количественные характеристики результатов цитоморфологобиофизической диагностики которые позволили объективизировать исследования и получить диагностическую информацию независимо от показателей, которые сильно варьируют.

Применение дискриминантного анализа при диагностике злокачественных и доброкачественных опухолей позволило повысить ее качество.

### Список литературы

1. Зубарев А.В. Методы медицинской визуализации – УЗИ, КТ, МРТ – в диагностике опухолей и кист печени / А.В. Зубарев. – М.: Наука, 1995. – 369 с.
2. Виторган Ю.Е. Клиническое применение иммунологических тестов при ранних формах рака молочной железы / Ю.Е. Виторган // Сборник: Иммунологические аспекты диагностики злокачественных новообразований – 1988. – С. 35-36.
3. Высоцкая Е.В. Разработка информационной системы дифференциальной диагностики состояния организма человека на основе цитоморфологобиофизических показателей / Е.В. Высоцкая, Г.В. Щукина, А.П. Порван, Н.А. Щукин // Вестник НТУ «ХПИ» Тематический выпуск «Информатика и моделирование». – Х.: НТУ «ХПИ», 2007. – № 39. – С. 17-20.
4. Каримов Р.Н. Обработка экспериментальной информации. Уч. пособие. Ч. 3. Многомерный анализ / Р.Н. Каримов. – Саратов: СГТУ, 2000. – 108 с.
5. Андерсон Т. Введение в многомерный статистический анализ / Т. Андерсон. – М.: Феникс, 1963. – 516 с.
6. Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: Издательский дом «Вильямс», 2007. – 118 с.
7. Ким Дж. Щ. Факторный, дискриминантный и кластерный анализ / Дж. Щ. Ким, Ч.У. Мюллер, У.Р. Клекка – М.: Финансы и статистика, 1989. – 215 с

Поступила в редколлегию 23.02.2011

Рецензент: д-р техн. наук, проф. А.И. Бых, Харьковский национальный университет радиоэлектроники, Харьков.

### ВИКОРИСТАННЯ ДИСКРИМІНАНТНОГО АНАЛІЗУ ДЛЯ ДІАГНОСТИКИ ДОБРОЯКІСНОЇ І ЗЛОЯКІСНОЇ ПУХЛИН

О.В. Висоцька, М.О. Щукін, А.П. Порван, С.М. Пушкар

У статті авторами пропонується діагностична модель диференціювання доброякісної пухлини та злоякісного процесу, отримана в результаті дискримінантного аналізу клінічних, інструментальних і лабораторних даних. Отримані дискримінантні функції використовуються при побудові територіальної карти диференціації доброякісної і злоякісної пухлини, а статистичні оцінки синтезованих моделей показують високу значимість.

**Ключові слова:** дискримінантний аналіз, злоякісна пухлина, територіальна карта.

### USING DISCRIMINANT ANALYSIS FOR DIAGNOSIS OF BENIGN AND MALIGNANT TUMORS

O.V. Visotskaja, N.O. Shukin, A.P. Porvan, N.S. Pushkar

In this paper authors propose a diagnostic model of the differentiation benign and malignant tumors of the process resulting from discriminant analysis of clinical, instrumental and laboratory data. The obtained discriminant functions are used to construct maps of territorial differentiation of benign and malignant tumors, and statistical evaluation of the synthesized models show a high significance.

**Keywords:** discriminant analysis, a malignant tumor, the territorial map