

И.П. Тарасова

## Возможности математического моделирования и прогнозирования в биологии и медицине

Все биологические объекты, изучаемые в биологии и медицине, такие, как микроорганизмы, лабораторные животные, отдельные трудовые коллективы, здоровые или больные люди и т.п., являются сложными стохастическими системами (рис. 1), функционирующими при воздействии на них множества входных факторов. Часть таких факторов  $X_1, X_2, \dots, X_k$  являются контролируруемыми, они измеряются количественно, оцениваются в баллах или номинально. Другая часть относится к группе неконтролируемых, случайных факторов и зачастую неизвестных, они не подлежат измерению, но оказывают воздействие на изучаемую биологическую систему, результатом которого является случайность ее состояния и функционирования.



Рис. 1. Представление объекта биологического исследования в виде «черного ящика».

Состояние системы характеризуется множеством выходных параметров  $Y_1, Y_2, \dots, Y_k$ , которые также измеряются количественно или в баллах и представляют собой случайные величины, подчиненные нормальному или иному закону распределения с соответствующими числовыми характеристиками.

В силу того, что неконтролируемые и случайные факторы для каждого объекта наблюдения принимают различные случайные значения, выходные параметры, характеризующие состояние и функционирование сложной стохастической (вероятностной) системы, являются случайными величинами, для исследования которых необходимо применять методы теории вероятности и математической статистики.

Одной из важных задач медико-биологического исследования является изучение связи между фактором, воздействующим на организм, и параметром-откликом на это воздействие, а также моделирование этого параметра в зависимости от действующего фактора. Эта задача реализуется методами корреляционного, регрессионного и дискриминантного анализа.

Корреляционный анализ для оценки связи признаков в медико-биологических исследованиях применяется не часто. Одной из причин является то, что коэффициенты корреляции – непостоянная и капризная мера, которая с трудом поддается интерпретации [1]. Более интересен в биологических исследованиях регрессионный и дискриминантный анализы.

Модели, описывающие различные патологические сдвиги на основе биохимических характеристик, могут быть построены с применением детерминистского подхода [2]. Для этого используются методы многомерного регрессионного и дискриминантного видов статистического анализа [3, 4]. Дискриминантный анализ является статистическим методом, который позволяет изучать различия между двумя и более группами объектов по отношению к нескольким переменным одновременно [4]. Данный анализ предусматривает создание экспертных систем (ЭС), под которыми понимают системы, способные выносить суждения и принимать решения без участия человека-эксперта [5]. В основе идеологии создания ЭС лежат системный подход, концепция различия и статистические методы распознавания образов.

Разработка диагностических экспертных систем (алгоритмов) каждого синдрома позволит резко улучшить результаты диагностики болезней в медицине. Алгоритмическая диагностика осуществляется на основе привычных общепринятых качественных критериев: наличия или отсутствия признака и степени его выраженности. Таким образом, диагностическую тактику можно свести к указанному алгоритму.

На сегодняшний момент клинических экспертных систем сравнительно не много, известные из них касаются в большей степени теоретических аспектов. Следует отметить попытки прогноза течения и моделирования вирусных поражений печени [6]. Построены количественные модели обострений бронхиальной астмы в зависимости от погодных-климатических условий [7–9], существуют эффективные модели в кардиологии [10].

Тем не менее, специалисты вновь и вновь возвращаются к ней в связи с нерешенностью многих проблем. Существует потребность в методах диагностики, обладающих высокой чувствительностью и специфичностью, которые давали бы четкие формальные критерии диагноза даже в тех случаях, когда клинические признаки заболевания не выражены.

В связи с вышесказанным актуальным является внедрение и использование новых способов анализа полученной информации с помощью компьютера. Одним из подходов в разработке новой системы диагностики является создание ЭС, которые позволили бы проводить компьютерное различение здоровых и патологических образов и компьютерную диагностику процессов, происходящих в организме.

*Цель работы.* Создание экспертных систем, способных на биохимическом уровне различить здоровый организм и больной ишемической болезнью сердца (ИБС).

*Материал и методы.* Для анализа нами была использована база данных Республиканского липидного лечебно-диагностического центра метаболической терапии (г. Витебск). Она включала в себя результаты комплексного биохимического обследования 129 практически здоровых людей и 2863 больных ИБС. Объектом анализа явились следующие биохимические показатели крови: триацилглицерины (ТГ), общий холестерол (ОХС), холестерол липопротеинов высокой плотности (ХС ЛПВП), холестерол липопротеинов низкой плотности (ХС ЛПНП), холестерол липопротеинов очень низкой плотности (ХС ЛПОНП), мочевины, билирубин, глюкоза, общий белок, альбумин, мочевая кислота, аланинаминотрансфераза (АЛТ), аспаратаминотрансфераза (АСТ), щелочная фосфатаза (ЩФ) и  $\gamma$ -глутамилтранспептидаза (ГГТ).

Математико-статистический анализ проводили с помощью прикладных компьютерных пакетов Microsoft Excel 2000, а также универсального статистического пакета STATGRAPHICS Plus 2.1 для Windows [3]. На основании полученной информации были созданы исходные базы данных (Microsoft Excel 2000 и STATGRAPHICS Plus 2.1).

Для решения поставленной задачи был применен пошаговый дискриминантный анализ, который позволил оценить чувствительность, специфичность и безошибочность созданной ЭС, а также был рассчитан коэффициент канонической корреляции (сап r). При проведении анализа указанной базы данных в качестве классифицирующего признака использовались перекодированные ряды, которые рассматривались как входные данные: 1 – больные ИБС, 2 – практически здоровые люди. В качестве входных параметров использовались пол, возраст, а также перекодированные биохимические показатели крови.

*Результаты и обсуждение.* Полученные результаты обследований были подвергнуты стандартному статистическому анализу (межгрупповое сопоставление средних значений, расчет попарных коэффициентов корреляции). После анализа матрицы коэффициентов корреляции несколько уменьшился объем материала, подлежащего дальнейшему рассмотрению. Во-первых, содержание ОХС сильно коррелировано ( $r=0,95$ ,  $p<0,0001$  для двух групп) с содержанием ХС ЛПНП. Во-вторых, обнаружена функциональная корреляционная зависимость между триацилглицеринами сыворотки крови и холестерином ЛПОНП у практически здоровых людей ( $r=1,00$ ,  $p<0,0001$ ) и весьма тесная взаимосвязь этих же показателей у больных ИБС ( $r=0,99$ ,  $p<0,0001$ ). Так как в дискриминантный анализ [5] может быть взят только один член пары признаков, для которой коэффициент корреляции близок к 1, поэтому ХС ЛПОНП и ХС ЛПНП не были использованы в качестве входных параметров.

При пошаговом дискриминантном анализе из общего набора входных параметров остались следующие переменные: пол ( $X_1$ ), возраст ( $X_2$ ), ОХС ( $X_3$ ), ХС-ЛПВП ( $X_4$ ), ТГ ( $X_5$ ), общий белок ( $X_9$ ), мочевая кислота ( $X_{10}$ ), АЛТ ( $X_{14}$ ) и альбумин ( $X_{15}$ ). Вместе они обеспечивают безошибочность экспертной системы в 95,03% (сап  $r=0,711$ ,  $p<0,0001$ ) случаев, при этом больные ИБС правильно диагностируются в 95,20%, а практически здоровые люди в 93,61% случаев.

Экспертная система имеет следующее математическое выражение:

$$Y = -3,868 - 0,697 \cdot X_1 + 0,092 \cdot X_2 + 0,246 \cdot X_3 - 0,19 \cdot X_4 + 0,158 \cdot X_5 - 0,049 \cdot X_9 + 0,001 \cdot X_{10} + 0,008 \cdot X_{14} + 0,03 \cdot X_{15}$$

Координаты групповых центроидов следующие: для больных ИБС – (0,345), для практически здоровых людей – (-2,882).

Графическое представление эффективности диагностики показано на рис. 2.



Рис. 2. Эффективность различения практически здоровых людей и людей, больных ИБС.

С учетом значений групповых центроидов, критической точкой является значение (-1,614): ее превышение свидетельствует о биохимических изменениях, характерных для ИБС (рис. 2), а цифры ниже критической точки доказывают о принадлежности к группе практически здоровых людей.

Данная ЭС может иметь важное практическое значение, так как применение ее на практике позволит врачу не только правильно диагностировать заболевание, но и выявлять людей, имеющих ранние, доклинические изменения метаболизма и организовать раннюю профилактику выявленных нарушений. Возможно также ее использовать при массовом обследовании населения.

По нашему мнению, сегодня медицинская общественность не готова к широкому использованию экспертных систем. Мы не можем полностью полагаться на ЭС, в связи с возможными ошибками, что определяется уровнем наших знаний в области той либо иной патологии. Поэтому сегодняшний день медицинской диагностики можно охарактеризовать как промежуточный. Машинные системы должны использоваться в клинической диагностике, но в качестве вспомогательных инструментов.

В заключение необходимо отметить, что прогресс в области биологии и медицины во многом зависит от понимания новых возможностей современных компьютеров и от смены парадигмы статистической обработки клинико-лабораторных данных парадигмой их глубокого системного анализа. Экспертные системы, полученные различными математическими методами, не только средство обработки экспериментальных данных, но и инструменты медико-биологического познания.

## ЛИТЕРАТУРА

1. **Генкин А.А.** Новая информационная технология анализа медицинских данных (программный комплекс ОМИС). – СПб., 1999. – 191 с.
2. **Саркисов Д.С., Гельфанд В.Б., Туманов В.П.** Проблемы ранней диагностики болезней человека // *Клин. мед.*, 1993, № 7. – С. 6–14.
3. **Дюк В.** Обработка данных на ПК в примерах. – СПб., 1997. – 240 с.
4. **Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р., Олдендерфер М.С., Блешфилд Р.К.** Факторный, дискриминантный и кластерный анализ / Под ред. **И.С. Енюкова.** – М., 1989. – 215 с.
5. **Нэйлор К.** Как построить свою экспертную систему. – М., 1991. – 288 с.
6. **Доценко М.Л.** Разработка и оценка эффективности диагностических экспертных систем для прогноза течения некоторых вирусных инфекций: Автореферат дис. ...доктора мед. наук. – М., 1998. – 42 с.
7. **Доценко Э.А., Фетисов И.Н., Пчельников Ю.В., Доценко М.Л.** Математический анализ влияния погодных факторов на характер течения бронхиальной астмы / *Иммунодиагностика и иммунотерапия.* – Витебск, 1995. – С. 36–43.
8. **Доценко Э.А., Ефременко И.И., Прищепа И.М.** Количественные модели обострений БА в зависимости от погодноклиматических условий // *Экологический вестник*, 1997, № 3. – С. 44–47.
9. **Доценко Э.А., Прищепа И.М.** Биоклиматология и экология бронхиальной астмы: абиотические факторы: Монография. – Витебск, 2001. – 353 с.
10. **Кузнецов Е.А., Негашева М.А., Пирунджан А.Л.** Новая стратегия в диагностике ишемической болезни сердца (математико-статистические аспекты) // *Экологическая антропология: Ежегодник*, 2003. – С. 120–125.

## S U M M A R Y

*The present paper is aimed at creating expert systems, which could give possibility to distinguish the healthy organism and the organism suffering from ischemic heart disease.*

*Поступила в редакцию 2.07.2004*