

УДК 616.12-008.331.1-037-036:616.379-008.64 (083.3)

С.Н. Коваль, Е.С. Першина, Т.Г. Старченко, А.В. Арсеньев*

*ГУ «Национальный институт терапии им. Л.Т. Малої
НАМН Украины», г. Харьков*

**Национальный фармацевтический университет, г. Харьков*

ПРОГНОЗИРОВАНИЕ ТЯЖЕСТИ ТЕЧЕНИЯ ГИПЕРТОНИЧЕСКОЙ БОЛЕЗНИ У БОЛЬНЫХ САХАРНЫМ ДИАБЕТОМ 2-ГО ТИПА МЕТОДОМ «ДЕРЕВЬЕВ КЛАССИФИКАЦИИ»

Описан метод «деревьев классификации», который позволяет наглядно видеть влияние каждого показателя и его значимость в оценке прогноза течения заболевания. Применение в модели профиброгенного фактора роста – ТФР- β 1 позволяет оценить его вклад в классификацию больных по степени тяжести и указывает на роль данного фактора в развитии структурно-функциональных нарушений левого желудочка сердца.
Ключевые слова: гипертоническая болезнь, сахарный диабет 2-го типа, классификация и ранжирование признаков, прогноз течения заболевания.

Математическое моделирование как нормальных физиологических, так и патологических процессов является в настоящее время одним из самых актуальных направлений в научных исследованиях [1]. Метод моделирования в медицине позволяет устанавливать более глубокие и сложные взаимосвязи между теорией и опытом. Одним из методов многомерного статистического анализа является метод «деревьев классификации», позволяющий исследовать структуру данных изучаемого процесса, построить «дерево классификации», выявить основные прогностические критерии течения заболевания [1, 2]. Метод «деревьев классификации» позволяет установить принадлежность объектов к тому или иному классу в зависимости от значений признаков, характеризующих данный объект, имеет высокий процент правильных ответов, наглядно графически представляет процедуру классификации и ранжирует признаки по степени значимости при определении степени тяжести состояния больных [3, 4].

Цель работы – выявить наиболее значимые прогностические критерии тяжести течения гипертонической болезни (ГБ) у больных с сопутствующим сахарным диабетом (СД) 2-го типа с помощью методов дискриминантного анализа и «деревьев классификации».

Материал и методы. Было обследовано 93 больных ГБ в сочетании с СД 2-го типа легкой и средней степени тяжести. Изучение уровней трансформирующего фактора роста- β 1 (ТФР- β 1) в сыворотке крови проводили иммуноферментным методом. Диастолическую функцию левого желудочка (ЛЖ) оценивали методом доплерографии в импульсном режиме по показателям трансмитрального кровотока. Определяли максимальную скорость кровотока в фазу раннего наполнения (Е), максимальную скорость кровотока в фазу позднего наполнения (А), их соотношение (Е/А), время замедления скорости кровотока в фазу ранней диастолы.

Результаты. С помощью методов кластерного анализа были сформированы две однородные группы среди больных ГБ с СД

© С.Н. Коваль, Е.С. Першина, Т.Г. Старченко, А.В. Арсеньев, 2013

2-го типа, которые можно интерпретировать как группы больных с лёгкой и средней степенью тяжести заболевания [5]. Это хорошо видно на рис. 1, где показаны средние значения наиболее характерных показателей.

Как видно из рис. 1, группа больных со средней степенью заболевания характеризуется более высокими значениями глюкозы

крови, достоверно высокими значениями АД. Больные ГБ с СД 2-го типа при прогрессировании тяжести заболевания имели более высокие значения ИММЛЖ и меньшее значение соотношения Е/А, что свидетельствует о более выраженных структурно-функциональных нарушениях сердца у данной категории больных, прогрессирующих по мере уве-

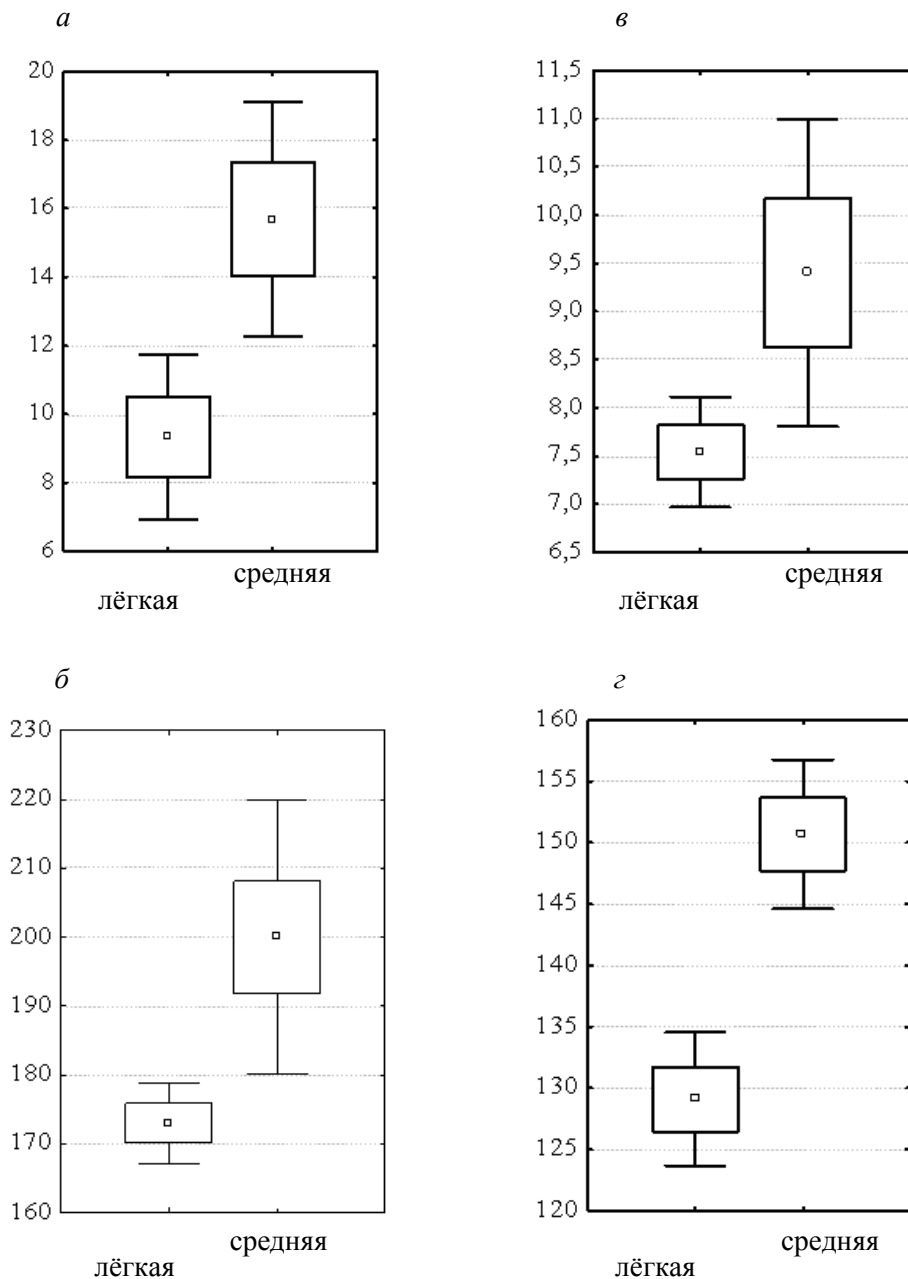


Рис. 1. Диаграмма размаха показателей у больных ГБ с СД 2-го типа с лёгкой и средней степенью заболевания:
 а – длительность АГ, лет; б – САТ, мм рт. ст.; в – глюкоза крови, ммоль/л;
 г – ИММЛЖ, г/м²

личения длительности заболевания и возраста, в то время как больные с лёгким течением заболевания имели достоверно низкие значения ИММЛЖ, глюкозы крови и низкий уровень САД. Уровень ТФР-β1 увеличивался при прогрессировании тяжести заболевания, что свидетельствует об активации профиброгенных процессов.

Для классификации групп больных в зависимости от степени тяжести заболевания можно использовать ряд методов многомерного статистического анализа, таких как дискриминантный анализ, деревья классификации, нейронные сети и др. Исходя из специфики рассматриваемой в работе задачи были выбраны дискриминантный анализ как один из наиболее точных и формализованных методов и деревья классификации – менее точные, но позволяющие наглядно интерпретировать результаты классификации.

Использование дискриминантного анализа предполагает прогнозирование тяжести состояния пациента (принадлежность к одной из двух групп) путём вычисления значений каждой классификационной функции для этого больного [2]. При этом больной относился к той группе тяжести состояния, где классификационная функция была максимальной. Классификационные функции вычисляли по формулам:

$$Y_{л} = -100,1 + 0,185 \text{ ТФР-}\beta 1 + 16,2 \text{ Е/А} + 0,61 \text{ возраст} + 1,2 \text{ глюкоза крови} - 0,6 \text{ длительность АГ} + 1,4 \text{ ДАД};$$

$$Y_{т} = -102,2 + 0,67 \text{ ТФР-}\beta 1 + 9,16 \text{ Е/А} + 0,79 \text{ возраст} + 1,7 \text{ глюкоза крови} - 0,36 \text{ длительность АГ} + 1,25 \text{ ДАД}.$$

Как следует из анализа канонических корней и матрицы факторной структуры, можно ранжировать дискриминантные переменные в модели по мере их вклада в классификацию: ТФР-β1, Е/А, возраст, глюкоза крови, длительность АГ, возраст.

Из матрицы классификации (таблица) видно, что точность классификации при лёг-

кой степени тяжести заболевания составила 93,1 %, при тяжёлой – 90,5 %.

Для получения более наглядной модели классификации использовали метод «деревьев классификации» [3, 6, 7]. Качество построенной модели с помощью дерева решений характеризуется двумя основными признаками: точностью распознавания и ошибкой. Процесс построения дерева происходит сверху вниз, то есть по нисходящей. В ходе процесса алгоритм должен иметь такой критерий разветвления, который бы ассоциировался с узлом проверки. В работе использован полный перебор деревьев с одномерным ветвлением по алгоритму CART, направленный на построение бинарного дерева решений. Каждый узел бинарного дерева при разбиении имеет только двух потомков, называемых дочерними ветвями. Дальнейшее распределение ветви зависит от количества исходных данных, которые описывает данная ветвь. На каждом шаге построения дерева согласно правилу, формируемому в узле, происходит деление указанного множества объектов на две части. Левая его часть – это та часть множества, в которой правило выполняется; правая – та, для которой правило не выполняется. Остановка ветвления осуществлялась прямой остановкой по методу FACT с установленной долей неклассифицированных наблюдений менее 5 %. Априорные вероятности принадлежности наблюдений в классы оценивались по выборке, а величина ошибки классификации установлена одинаковой для обоих классов. Для оценки качества осуществлялась V-кратная кросс-проверка.

Полный перебор деревьев с одномерным ветвлением по алгоритму CART свидетельствует о наличии четырёх близких по составу предикторов модели с ценой глобальной кросс-проверки в диапазоне 0,29–0,33. Это можно интерпретировать, как точность классификации на достаточно высоком уровне – 67–71 %. Поэтому была выбрана модель, в ко-

Матрица классификации

Группа прогноза	Правильно классифицированные наблюдения, %	Число больных со степенью тяжести	
		лёгкой	средней
Лёгкая степень	93,1	49	3
Средняя степень	90,5	4	36
Всего	92,0	53	39

торой использовались такие же предикторы, как и при дискриминантном анализе. При этом точность классификации на выборке, которая изучалась (93 больных), составила 100 %, тогда как результаты кросс-проверки показали меньшую точность – 71 %, что можно считать достаточно хорошим результатом.

На рис. 2 показано дерево решений с указанием условия ветвления. При этом число рядом с прямоугольником – это количество больных, которые в него входят, а гистограммы в прямоугольниках – соотношения больных по степени тяжести в данном узле.

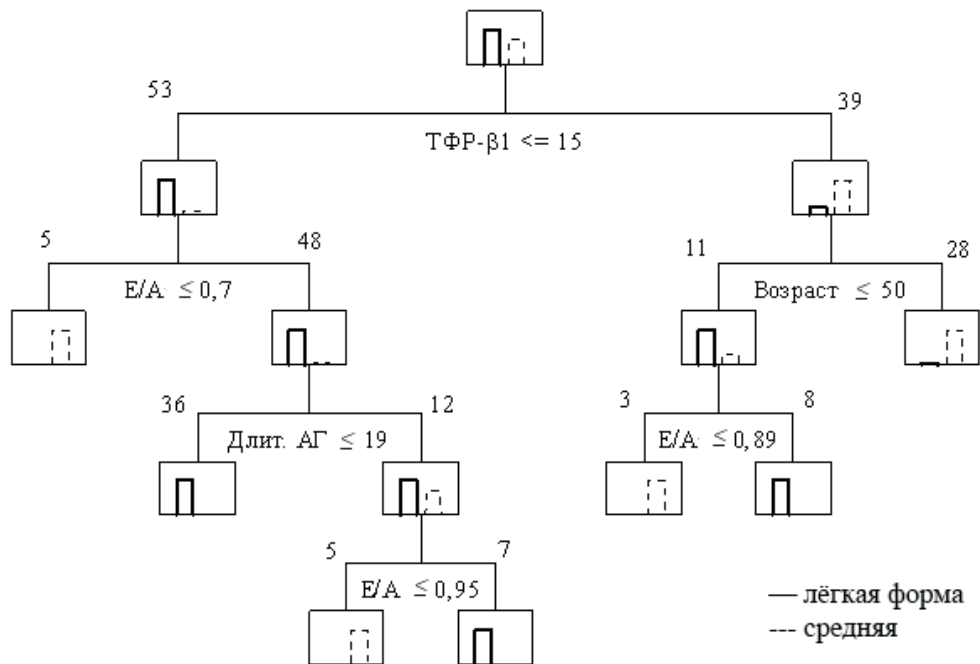


Рис. 2. Дерево решений с указанием условия ветвления

Как свидетельствуют данные рис. 2, тяжесть ГБ в сочетании с СД интерпретировалась как зависимая, группирующая переменная. В результате проведенного анализа составляется алгоритм, на первом шаге которого больные ГБ с СД 2-го типа подразделяются в зависимости от уровня ТФР-β1 (до 15,01 нг/мл).

Первую группу составили 53 больных ГБ с СД 2-го типа с ТФР-β1 до 15,01 нг/мл преимущественно с лёгким течением заболевания, вторую – 39 больных ГБ с СД 2-го типа и ТФР-β1 > 15,01 нг/мл, в основном с более тяжёлым течением заболевания. В подгруппе с ТФР-β1 до 15,01 нг/мл последующим

фактором разбиения был показатель диастолической дисфункции ЛЖ (соотношение Е/А), а в группе с ТФР-β1 > 15,01 нг/мл – возраст. При соотношении Е/А ≤ 0,7 тяжёлую степень заболевания имели 5 больных. Последующим фактором разбиения 48 больных, имеющих преимущественно лёгкую степень тяжести заболевания, была длительность АГ. При длительности АГ ≤ 19 лет 36 пациентов имели лёгкую степень тяжести. Дальнейшее разделение больных с длительностью заболевания более 19 лет осуществлялось также в зависимости от соотношения Е/А. Из 12 па-

циентов при Е/А ≤ 0,95 тяжёлую степень заболевания имели 5 пациентов, лёгкую – 7. Вторую группу с уровнем ТФР-β1 > 15,01 нг/мл составили 39 больных. В основе разделения этой группы был возраст: пациенты моложе 50 лет характеризовались лёгким течением заболевания. Последующим критерием разделения данных больных было соотношение Е/А. При этом 3 пациента с Е/А ≤ 0,89 имели тяжёлую степень заболевания, 8 – лёгкую.

Дерево имеет 7 терминальных узлов, точность кросс-проверки (степень точности) составляет 96 %, стандартная ошибка классификации – 72 %. На каждом этапе этот диагно-

стический алгоритм имеет два варианта решений, которые ведут к решению вопроса о принадлежности больного к определённой степени тяжести заболевания.

Выводы

Для определения с высокой степенью точности степени тяжести состояния больных ГБ с СД 2-го типа предложены математические модели на основе методов дискриминантного анализа и «деревьев классификации». Используемый метод «деревьев классификации» (деревья решений) представляет более

упрощённую и удобную картину алгоритма классификации, что позволяет наглядно видеть влияние каждого показателя и его значимость в оценке прогноза течения заболевания. Так, применение в модели профиброгенного фактора роста – ТФР- β 1 – позволяет оценить его вклад в классификацию больных по степени тяжести и указывает на роль данного фактора роста в развитии структурно-функциональных нарушений левого желудочка сердца, прогрессивно нарастающих по мере увеличения возраста и длительности артериальной гипертензии.

Список литературы

1. *Резниченко Г.Ю.* Лекции по математическим моделям в биологии. Ч. 1. – Москва–Ижевск: Научн.-изд. центр «Регулярная и хаотическая динамика», 2002. – 231 с.
2. *Rencher Alvin C.* Methods of multivariate analysis / Alvin C. Rencher, 2nd ed. (Wiley series in probability and mathematical statistics). A. Wiley-Interscience publication, 2002. – 732 p.
3. Методы и модели анализа данных: OLAP и Data Mining / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко [и др.] // СПб.: БХВ-Петербург, 2004. – 336 с.
4. *Елисеева Л.Н.* Классификация больных, страдающих хронической сердечной недостаточностью методом «Деревья классификации» / Л.Н. Елисеева, А.А. Халафян, С.Г. Сафонова // Успехи современного естествознания. – 2006. – № 11. – С. 16–18.
5. *Марчук Г.И.* Математические модели в иммунологии. — М.: Наука, 1985. – 240 с.
6. StatSoft. Inc. (2001) Электронный учебник по статистике. Москва. StatSoft. WEB: <http://www.StatSoft.ru/home/textbook/default.htm>
7. *Елисеева Л.Н.* Применение методов классификационного анализа для определения функционального класса хронической сердечной недостаточности больных, страдающих ишемической болезнью сердца / Л.Н. Елисеева, А.А. Халафян, С.Г. Сафонова // Экологический вестник научных центров Черноморского экономического сотрудничества. – 2005. – № 3 – С. 79–83.

С.М. Коваль, К.С. Першина, Т.Г. Старченко, О.В. Арсеньев

ПРОГНОЗУВАННЯ ТЯЖКОСТІ ПЕРЕБІГУ ГІПЕРТОНІЧНОЇ ХВОРОБИ У ХВОРИХ З ЦУКРОВИМ ДІАБЕТОМ 2-ГО ТИПУ МЕТОДОМ «ДЕРЕВ КЛАСИФІКАЦІЇ»

Описаний метод «дерев класифікації», який дозволяє за допомогою графічного зображення визначити вплив кожного показника та його значущість в оцінці прогнозу перебігу захворювання. Застосування в моделі профіброгенного фактора росту – ТФР- β 1 – дозволяє оцінити його внесок в класифікацію хворих за ступенем тяжкості та вказує на роль даного фактора росту в розвитку структурно-функціональних порушень лівого шлуночка серця.

Ключові слова: гіпертонічна хвороба, цукровий діабет 2-го типу, класифікація і розташування ознак по ранжирі, прогноз перебігу захворювання.

S.N. Koval, K.C. Pershyna, T.G. Starchenko, A.B. Arsenyev

PREDICTION OF SEVERITY OF ESSENTIAL HYPERTENSION IN PATIENTS WITH DIABETES MELLITUS TYPE 2 BY METHOD «CLASSIFICATION TREE»

The article describes a method of classification trees, which allows visually see the effect of each parameter and its importance in the assessment of prognosis of the disease. The use of a model of profibrogenic growth factor – TGF- β 1 – to evaluate its contribution to the classification of patients according to the severity of disease. The role of this growth factor in the development of structural and functional disorders of left ventricle of the heart was established.

Key words: essential hypertension, type 2 diabetes mellitus, classification and range of parameters, prognosis of disease.

Поступила 11.06.13