

УДК 616.37-002.2-06:616.995.132.8]-071/-078-047.44

В. П. Марценюк¹, д-р техн. наук, проф.,
Л. С. Бабинець², д-р мед. наук, проф.,
Ю. В. Дроняк²

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ DATA MINING С ЦЕЛЬЮ ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКИ КОМОРБИДНЫХ СОСТОЯНИЙ ХРОНИЧЕСКОГО ПАНКРЕАТИТА И АСКАРИДОЗА НА ОСНОВАНИИ ДАННЫХ КЛИНИЧЕСКОЙ СИМПТОМАТИКИ И ИММУНОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

¹ Университет в Бельско-Бялой, Польша,

² Тернопольский государственный медицинский университет имени И. Я. Горбачевского, Украина

Введение

Распространенность заболеваний поджелудочной железы (ПЖ) и их место в общей структуре болезней органов желудочно-кишечного тракта (ЖКТ) значительна. Сегодня распространенность хронического панкреатита (ХП) среди населения Украины, по некоторым данным, составляет 25–30 человек на 100 тыс. населения. За последние 30–40 лет количество больных ХП увеличилось вдвое, первичная инвалидизация достигла 15 % [3; 7].

В большинстве случаев ХП не является изолированным, растет количество лиц с сочетанной патологией, а поражения органов пищеварения отягощают и усложняют течение других заболеваний. Исследования ряда авторов свидетельствуют о том, что в 70–90 % случаев заболевания ЖКТ сочетаются между собой или сопровождаются заболеваниями других органов и систем [1; 6]. В последние годы все большее внимание исследователи уделяют изучению сочетания гастроэнтерологических заболеваний с глистной инвазией. Это, в свою очередь, нуждается в уг-

лублении знаний о патогенетических звеньях при сочетанной патологии и в разработках эффективных лечебных и профилактических мероприятий.

Коморбидность ХП и аскаридоза является достаточно актуальной в наше время: ХП — частый спутник лиц с гельминтозами, особенно с аскаридозом и наоборот.

Клиническая симптоматика ХП чрезвычайно богата и полиморфна в своих проявлениях, однако она является малоспецифической [2; 3; 5; 6]. Классически можно выделить основные критерии клинического течения ХП посиндромно. Клинически ХП характеризуется двумя ведущими синдромами: рецидивирующей и/или постоянной абдоминальной болью и/или экзокринной недостаточностью ПЖ [4; 6]. У некоторых пациентов боли не связаны с приемом пищи, а у 15 % — вообще отсутствуют [7]. При длительном течении ХП, когда отмечается прогрессирование диффузного фиброза в ткани ПЖ, возникает и постепенно углубляется трофологическая недостаточность, которая характеризуется синдромом дефицита энергии, белка, витаминов, микроэлементов и электролитов [1–3]. На фоне прогрессирующей функциональной недостаточнос-

ти ПЖ и углубления трофологической недостаточности исчерпываются запасы нутриентов в тканях, что сначала ведет к биохимическим и функциональным изменениям, а в дальнейшем проявляется многочисленными и сложными клиническими симптомами [1; 6; 7]. Сопутствующий аскаридоз углубляет эти явления, хотя часто остается недиагностированным.

Аскаридоз характеризуется признаками хронической интоксикации и аллергизации человеческого организма, нарушением функционирования иммунной системы, органов дыхания и ЖКТ, миалгией, лимфаденопатией, железодефицитной анемией. Как результат, снижение показателей здоровья, а следовательно, и качества жизни пациентов. Симптомы наличия аскаридов маскируются под такими характерными для многих заболеваний проявлениями, как хроническая усталость; изменение аппетита (чаще его ухудшение); нервозность; нарушение моторики ЖКТ (тошнота, иногда рвота, запоры, понос, метеоризм); слюнотечение; анемия и другие изменения крови — повышение СОЭ, эозинофилия; боль в разных группах мышц, суставах; головная боль; «беспричинное» периодическое повышение температуры тела до субфебрильных цифр (37–38 °С); бессонница и прочие нарушения сна; гиповитаминоз; различные кожные и аллергические проявления; частые заболевания дыхательной системы (иногда вплоть до бронхиальной астмы); скрежет зубами во время сна (бруксизм) [2; 8]. Таким образом, симптоматика коморбидных ХП и аскаридоза часто наслаивается одна на другую, взаимно отягощая общие проявления и добавляя характерные для каждого из этих заболеваний.

Нарушения иммунной системы характерны для обеих исследуемых нозологий. Большинство исследователей свидетельствуют о наличии вторичного иммунодефицита при ХП [1; 4]. Губергриц Н. Б. (2007) доказано наличие при ХП дисбаланса между Т- и В-звеньями иммунитета, который заключается в повышении количества В-лимфоцитов в сочетании с увеличением уровней IgA, IgM, IgG в сыворотке крови при уменьшении Т-лимфоцитов, выявлена сенсibilизация к микробным аллергенам при ХП [3].

Иммунная система пациентов с ХП в фазе обострения характеризуется достоверным уменьшением относительного содержания в крови Т-лимфоцитов, их супрессорной субпопуляции, В-лимфоцитов, дисбалансом между функциональной активностью Т- и В-звеньев иммунитета,

нарастанием циркулирующих иммунных комплексов (ЦИК) [4; 7]. Эти изменения клеточного и гуморального звеньев иммунной системы с нарушением баланса цитокинов. Ряд экспертов установили при ХП наличие следующих показателей метаболической иммуносупрессии: повышение уровня IL-1, IL-6 и TNF- α , снижение уровня CD4 при повышении цитотоксического CD8, увеличение активности апоптоза [8]. Таким образом, иммуновоспалительные реакции, сопровождающиеся дисбалансом иммунной системы на фоне снижения способности лимфоцитов к пролиферации в ответ на неспецифический этиологический фактор, в целом обеспечивают и поддерживают хронизации ХП [4].

Установлено, что CD16 (NK — натуральные киллеры) имеют литическое действие в отношении клеток опухолей, принимают участие в противовирусном и противопаразитарном иммунитете. Реакции с участием NK приводят к спонтанной клеточно-опосредованной цитотоксичности, для которой не нужны антитела, комплемент и др. Снижение NK свидетельствует о негативном прогнозе в течении хронических заболеваний [3].

Как уже указывалось выше, при ХП характерно увеличение уровня ЦИК, что является компонентом нормального иммунного ответа [4]. Это характерно и для аскаридоза. Однако при определенных условиях ЦИК могут фиксироваться в сосудах и вызывать воспалительную реакцию. При хронических процессах они долго циркулируют в крови, поддерживая условия для поражения тканей-мишеней [3]. В малых концентрациях ЦИК способны вызвать пролиферацию В-лимфоцитов, а в больших — подавлять ее. Имеются сообщения, что под влиянием больших концентраций ЦИК происходит повышение активности Т-супрессоров, особенно тех, в состав которых входят IgG, а также подавление активности NK [7]. Аллергизация, характерная как для ХП, так и для аскаридоза, часто сопровождается дисбалансом гуморальных факторов, особенно повышением IgE.

Все приведенные факты относительно иммунодефицита при ХП и аскаридозе, по нашему мнению, могут способствовать поддержанию воспалительного хронического процесса в ПЖ и других вовлеченных органах, требуют дальнейшего изучения и анализа. Являясь общими при ХП и аскаридозе, они могут помочь при оценке состояния больного с коморбидностью, особенно в принятии клинического решения о дальнейшей тактике врача.

Материалы и методы исследования

Обследовано 90 пациентов с ХП в возрасте от 18 до 77 лет (средний возраст — $(50,54 \pm 13,93)$ года, из них женщин 56 (62,2 %), мужчин — 34 (37,7 %). Больные были распределены на две группы: I — 60 пациентов с ХП в сочетании с аскаридозом, II — 30 больных ХП. Всем пациентам с ХП были проведены общеклинические и иммунологические исследования.

Применение алгоритма индукции дерева решений

Математически задача индукции дерева решений формулируется таким образом. Имеем множество D , которое содержит N наборов учебных данных. При этом каждый i -й набор $(A_1^i, A_2^i, \dots, A_p^i, C^i)$ состоит из входных данных — атрибутов A_1, \dots, A_p и выходных данных — атрибута класса C . Атрибуты A_1, \dots, A_p могут принимать как численные, так и категориальные значения. Атрибут класса C принимает одно из K дискретных значений: $C \in \{1, \dots, K\}$.

Целью является прогнозирование деревом решений значения атрибута класса C на основании значений атрибутов A_1, \dots, A_p . При этом следует максимизировать точность прогнозирования атрибута класса, а именно $P\{C = c\}$ на терминальных узлах для любого $c \in \{1, \dots, K\}$. Алгоритмы индукции деревьев решений автоматически разбивают на узлах значения численных атрибутов A_i на два интервала: $A_i \leq x_i$ и $A_i > x_i$, а категориальных атрибутов A_j — на два подмножества: $A_j \in S_j$, $A_j \notin S_j$. Разбиение численных атрибутов основывается, как правило, на энтропии или индексе Джини [9–11]. Процесс разбиения рекурсивно повторяется до тех пор, пока не будет наблюдаться улучшение точности прогнозирования. Последний шаг включает исключение узлов для избежания оверфитинга модели. В результате мы должны получить множество правил, которые идут от корня к каждому терминальному узлу, содержат неравенства для численных атрибутов и условия включения для категориальных атрибутов.

Цель исследования — применение метода индукции дерева решений для программной реализации в системе принятия решений относительно коморбидности ХП и аскаридоза.

Метод индукции дерева решений. За основу взята такая рекурсивная процедура работы [9].

Генерация дерева решений

Входные данные:

D — множество учебных наборов данных $(A_1^i, A_2^i, \dots, A_p^i, C^i)$.

Выходные данные: дерево решений.

Метод:

1. Создать узел N .
2. Если все наборы в D принадлежат к общему классу C , тогда вернуть узел N как листок с названием класса C .
3. Если список атрибутов (а значит, и D) является пустым, тогда вернуть узел N как листок с названием самого распространенного класса в D .
4. Применить *Алгоритм отбора атрибута* из списка атрибутов и для множества D с целью отыскания «наилучшего» атрибута деления.
5. Удалить атрибут деления из списка атрибутов.
6. Для каждого условия деления j для атрибута деления рассмотрим D_j — множество наборов из D , которые удовлетворяют условию деления j .
7. Если D_j — пустое, тогда присоединить к узлу N лист под заглавием самого распространенного класса в D_j , иначе — присоединить к N узел, который возвращается рекурсивным вызовом метода *Генерация дерева решений* с входными данными D_j и список атрибутов.
8. Конец цикла шага 6.
9. Вернуть узел N .

В основу *Алгоритма отбора атрибута* на j -м шаге рекурсии положен такой информационный показатель:

$$Gain(A_i) = Info(D_j) - Info_{A_i}(D_j). \quad (1)$$

Здесь

$$Info(D_j) = - \sum_{k=1}^K p_k^j \log_2(p_k^j) \quad (2)$$

— информация, необходимая для классификации набора (A_1, A_2, \dots, A_p) в D_j ,

$$Info_{A_i}(D_j) = \sum_{l=1}^{K_i} \frac{\#(D_j^l)}{\#(D_j)} Info(D_l) \quad (3)$$

— информация, необходимая для классификации (A_1, A_2, \dots, A_p) в D_j после деления D_j на подмножества D_j^l соответственно к значениям атрибута A_i .

В формуле (2) вероятность того, что любой набор из D_j принадлежит множеству C_k, D_j , оценивается как

$$p_k^j = \frac{\#(C_{k,D_j})}{\#(D_j)},$$

где C_{k,D_j} — множество наборов из D_j , для которых атрибут класса $C = k$. Здесь $\#(\bullet)$ — количество элементов в множестве.

Таблица 1

**Атрибуты для индукции дерева решений
на основании клинической симптоматики**

Обозначение атрибута	Описание атрибута	Тип значений атрибута
A1	Пол	Фактор
A2	Возраст	Целочисленный
A3	Боль в животе	Фактор
A4	Тошнота	Фактор
A5	Рвота	Фактор
A6	Изжога	Фактор
A7	Горечь во рту	Фактор
A8	Отрыжка	Фактор
A9	Вздутие	Фактор
A10	Запор	Фактор
A11	Понос	Фактор
A12	Чередование запора и поноса	Фактор
A13	Плохой аппетит	Фактор
A14	Тяжесть в правом подреберье	Фактор
A15	Чувство переполнения	Фактор
A16	Общая слабость	Фактор
A17	Длительность заболевания	Целочисленный

В формуле (3) $\frac{\#(D_j^l)}{\#(D_j)}$ — оценка вероятности

того, что любой набор из D_j принадлежит множеству D_j^l , где D_j^l — множество наборов из D_j , для которых атрибут $A_i = a_i^l$. Здесь атрибут $A_i \in \{a_i^1, a_i^2, \dots, a_i^{K_i}\}$.

Таким образом, $Gain(A_i)$ оценивает уменьшение информации, необходимой для классификации любого набора данных в D_j , за счет известного значения атрибута A_i . Итак, из наличных атрибутов на каждом узле дерева решений для условия деления следует выбирать атрибут A_{i^*} с наибольшим значением $Gain(A_{i^*})$. В результате такого выбора для завершения процесса классификации набора данных в D_j будет требоваться меньше информации. При применении вышеописанного алгоритма для построения дерева решений использованы наборы атрибутов (табл. 1–3). При построении дерева с учетом данных как клинической симптоматики, так и иммунологических параметров использована комбинация атрибутов табл. 1, 2 и 3.

В работе использована реализация алгоритма на языке R с помощью пакета C5.0. Построенные деревья решений представлены на рис. 1–5.

Анализ полученных результатов

Анализируя дерево решений, построенное на основании данных клинической симптоматики (см. рис. 1), установлено, что по информативности на первом месте находится наличие боли в животе. При ее отсутствии алгоритм предлагает вопрос о поносе. При его наличии имеем с вероятностью $P=1$ ХП (5 случаев). При отсутствии поноса система предлагает вопрос о горечи во рту. При ее наличии имеем с вероятностью $P=1$ ХП (2 случая). При отсутствии уточняем пол. Для женщин с вероятностью $P=1$ (4 случая) имеем коморбидность ХП с аскаридозом. Для мужчин уточняем возраст. Для мужчин моложе 45 лет с вероятностью $P=1$ (4 случая) имеем коморбидность ХП с аскаридозом, старше 45 лет — с вероятностью $P=0,75$ (3 случая) отсутствие аскаридоза, с вероятностью $P=0,25$ (1 случай) — наличие.

При наличии абдоминальной боли встает вопрос об общей слабости. При ее наличии — вопрос о тошноте. При наличии тошноты с вероятностью $P=0,96$ (27 случаев из 28) присутствует аскаридоз. При отсутствии тошноты уточняем аппетит. Если аппетит хороший, то с вероятностью $P=0,895$ (17 из 19) имеем аскаридоз. Если аппетит плохой, то уточняем возраст. Если па-

Таблица 2

**Атрибуты для индукции дерева решений
на основании показателей
клеточного иммунитета**

Обозначение атрибута	Описание атрибута	Тип значений атрибута
A32	CD3	Действительное число
A33	CD22	Действительное число
A34	CD16	Действительное число
A35	CD4	Действительное число
A36	CD8	Действительное число

Таблица 3

**Атрибуты для индукции дерева решений
на основании показателей
гуморального иммунитета**

Обозначение атрибута	Описание атрибута	Тип значений атрибута
A37	IgG	Действительное число
A38	IgA	Действительное число
A39	IgM	Действительное число
A40	IgE	Действительное число

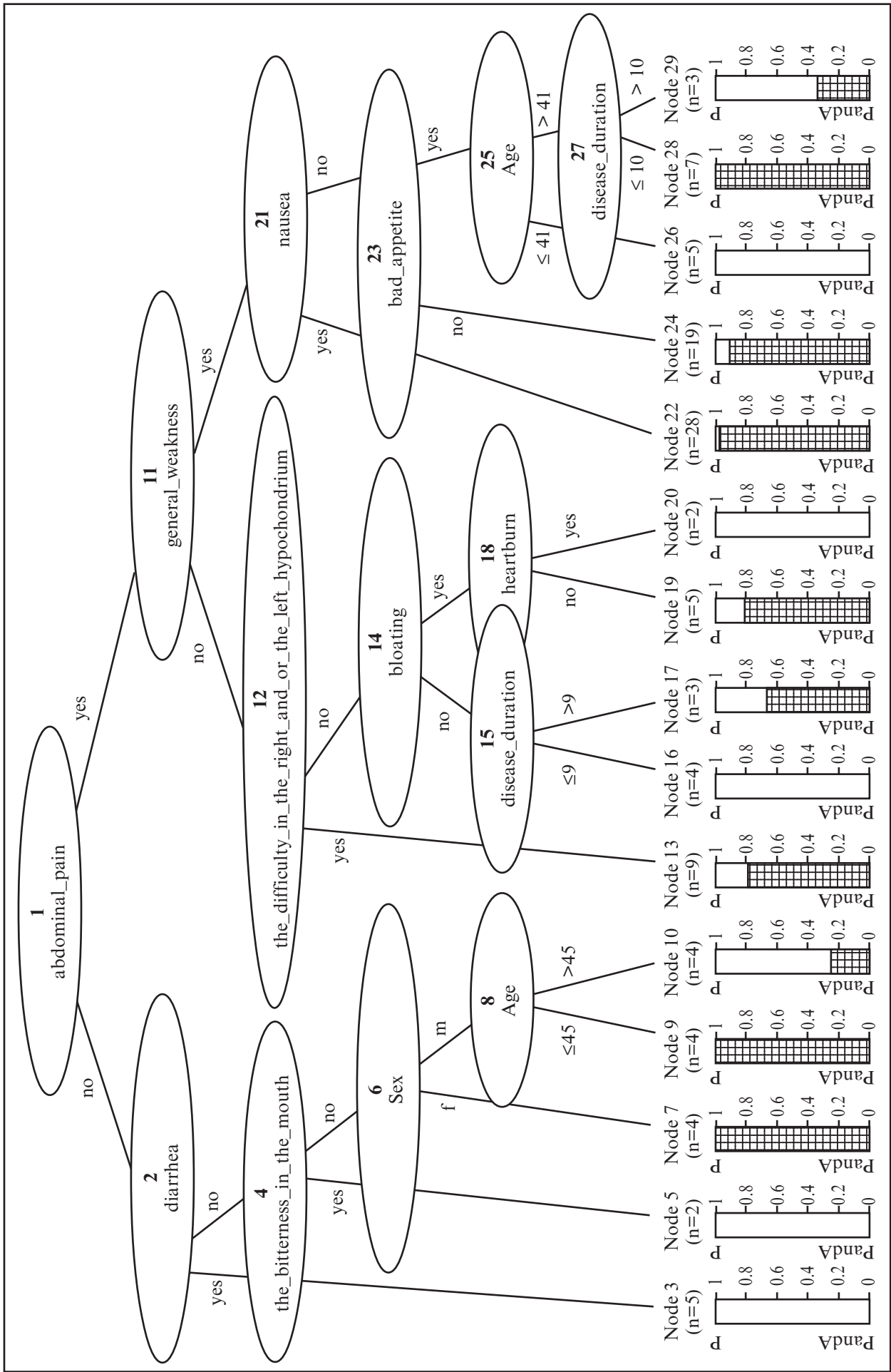


Рис. 1. Дерево решений на основании данных клинической симптоматики.

На рис. 1–5: Panda — коморбидность хронического панкреатита и аскаридоза; P — хронический панкреатит

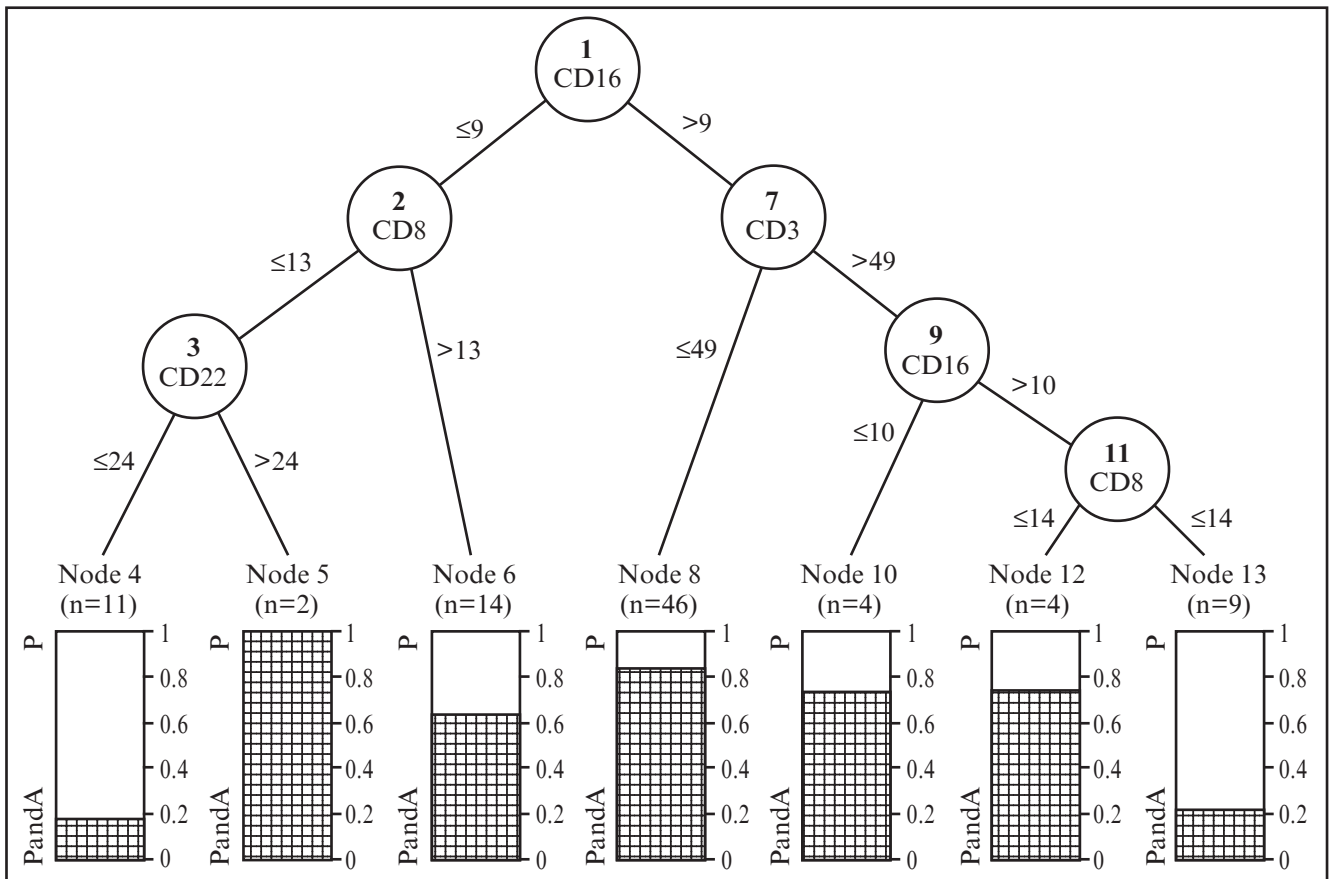


Рис. 2. Дерево решений на основании показателей клеточного иммунитета

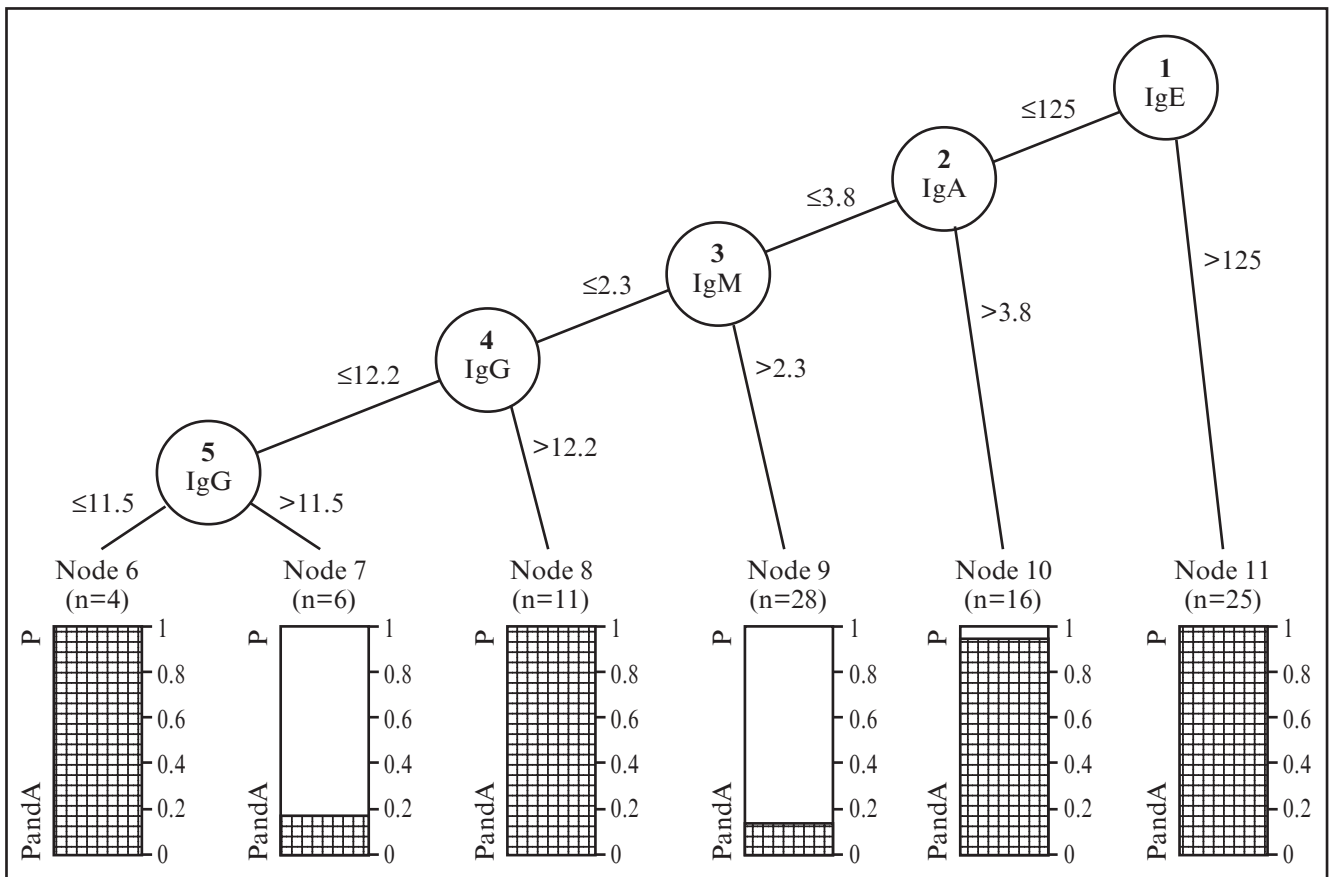


Рис. 3. Дерево решений на основании показателей гуморального иммунитета

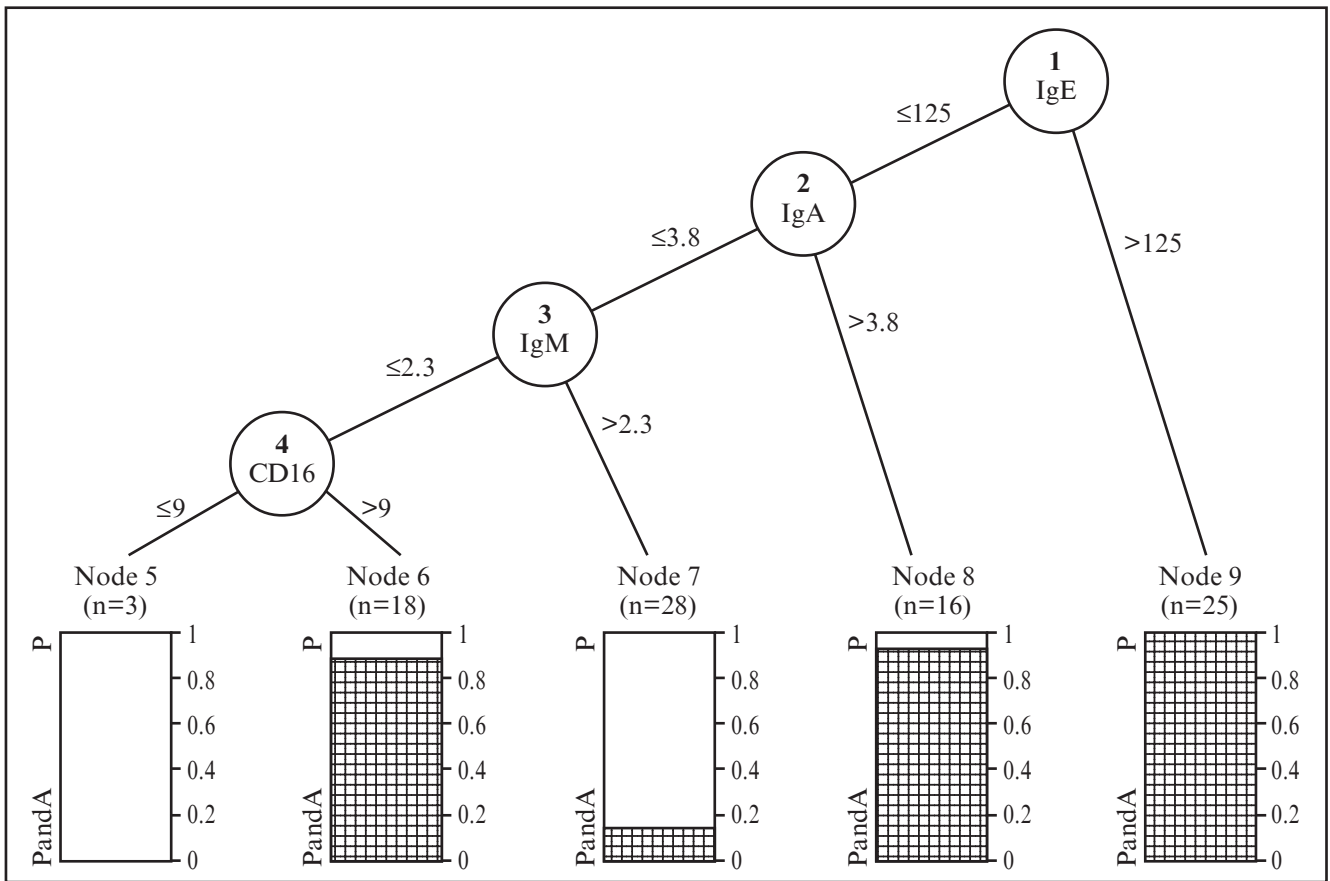


Рис. 4. Дерево решений на основании показателей иммунограммы

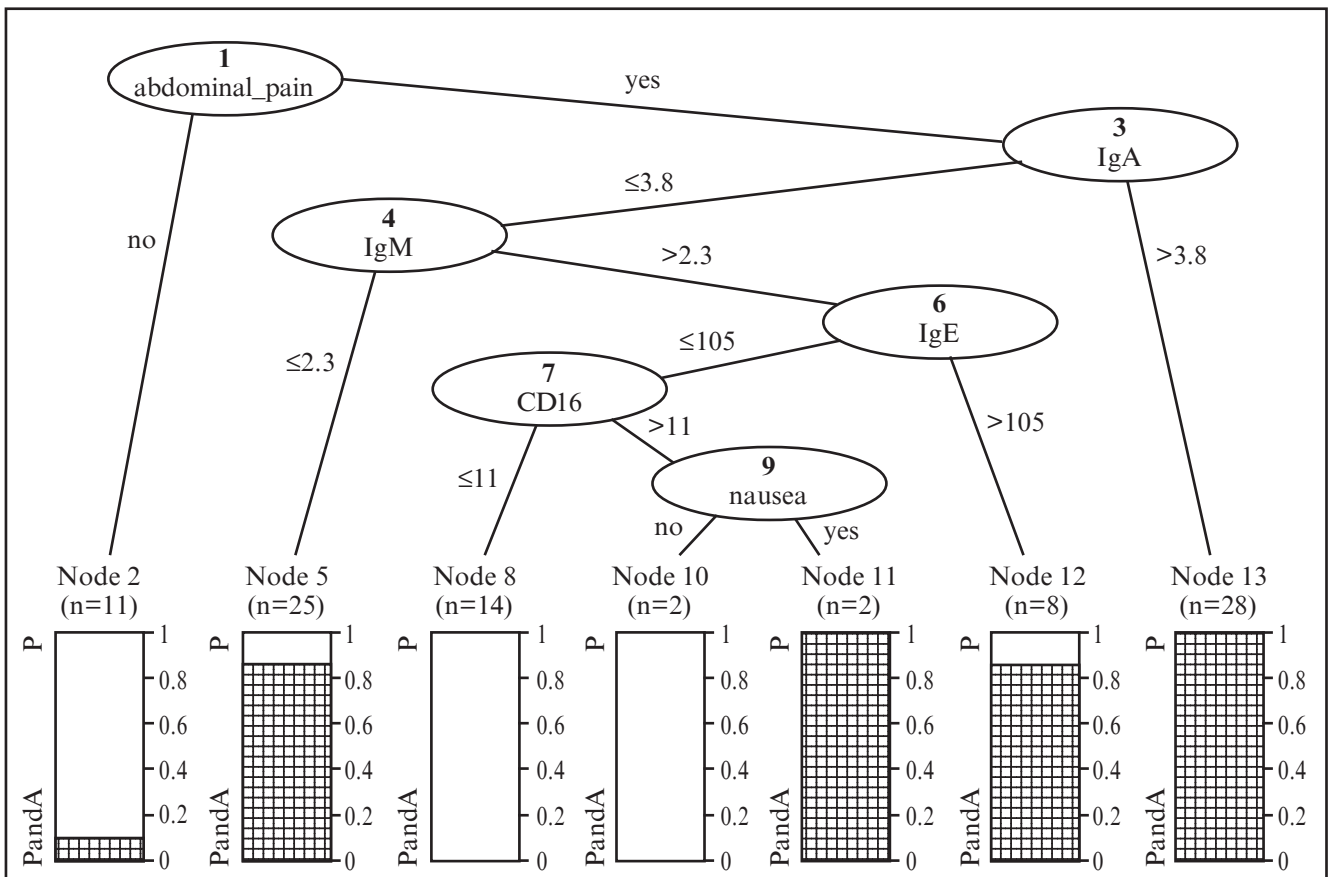


Рис. 5. Дерево решений на основании данных клинической симптоматики и иммунограммы

циент моложе 41 года (включительно), то с вероятностью $P=1$ аскаридоза нет (5 случаев). Если старше, то уточняем длительность болезни. Если длительность болезни меньше либо равна 10 лет, то с вероятностью $P=1$ (7 случаев) имеем наличие аскаридоза. Если больше 10 лет, то с вероятностью $P=0,66$ (2 случая из 3) аскаридоза нет. И наоборот.

При отсутствии общей слабости имеем соответствующее поддереву. Первым уточняем вопрос о тяжести в правом/левом подреберье. При ее наличии имеем с вероятностью $P=0,78$ аскаридоз. При ее отсутствии спрашиваем о вздутии. Если вздутия нет, то уточняем длительность болезни. Если меньше 9 лет, то с вероятностью $P=1$ аскаридоза нет. Если больше, то с вероятностью $P=0,66$ аскаридоз есть (2 случая из 3). Если вздутие имеется, то встает вопрос об изжоге. Если изжоги нет, то с вероятностью $P=0,8$ имеем аскаридоз (4 случая из 5), если есть, то с вероятностью $P=1$ аскаридоза нет.

Анализируя дерево решений, построенное на основании данных иммунограммы (см. рис. 4), видим, что первое место по информативности для диагностики коморбидности занимает IgE. При значениях IgE больше 125 с вероятностью $P=1$ имеем коморбидность ХП и аскаридоза (в нашем исследовании определено у 25 пациентов). При значениях IgE меньше либо равных 125 следующим по информативности является IgA. При значениях IgA больше 3,8 вероятность коморбидного аскаридоза составляет 0,94 (16 из 17 пациентов), вероятность его отсутствия 0,06 (1 пациент из 17). При значениях IgA меньше либо равных 3,8 алгоритм предлагает анализировать IgM. При значениях IgM больше 2,3 аскаридоз отсутствует с вероятностью 0,86 (28 пациентов из 32) и с вероятностью 0,14 — присутствует (4 пациента из 32).

При значениях IgM меньше либо равных 2,3 далее проверяем CD16. При значениях меньше либо равных 9 аскаридоз отсутствует с вероятностью $P=1$ (3 случая). При больше 9 — присутствует с вероятностью $P=0,9$ (18 случаев из 20) и наоборот — отсутствует с вероятностью $P=0,1$ (2 случая из 20). Заметим, что вероятность ошибки при использовании такого дерева решений, построенного только на данных иммунограммы, составляет 7,8 % (7 случаев из 90). Частота использования атрибутов, используемых в соответствии с деревом рис. 4, имеет вид: 100,00 % — IgE; 72,22 % — IgA; 54,44 % — IgM; 23,33 % — CD16.

Подобным же образом может быть проанализировано дерево решений с учетом как данных

клинической симптоматики, так и иммунологических исследований (см. рис. 5). Важно отметить, что с использованием этого дерева решений получаем частоту использования атрибутов для классификации коморбидных состояний (табл. 4). В табл. 4 указано использование атрибутов в порядке убывания частоты (в процентах). Исследована погрешность диагностирования с использованием дерева, представленного на рис. 5. В целом на исследуемых данных погрешность составляет 5,6 % (т. е. 5 случаев из 90). При этом ошибочно диагностируются четыре случая из 30 при отсутствии аскаридоза и один случай из 60 при его наличии.

В приложении приведен детальный отчет программы C5.0 относительно анализа деревьев решений, построенных на рис. 1–5 с указанием вероятности диагностирования коморбидных состояний, погрешности алгоритма и частоты использования атрибутов.

Выводы

Проведенное исследование показало, что сопутствующий аскаридоз усложнял течение хронического панкреатита, увеличивая частоту болевого, диспепсического, энтеропанкреатического синдромов, углубляя проявление трофологической недостаточности (похудение, анемия, гиповитаминозы, иммунодефицит и другие проявления), что отражается на качестве жизни пациента. Заподозрить и выявить сопутствующий аскаридоз важно для правильной оценки состояния больного с хроническим панкреатитом и составления наиболее эффективной программы лечения с учетом этиологии и коморбидности.

Для диагностирования аскаридоза может быть использована методика на основании построения дерева решений. Это одна из составляющих технологии data mining, основанная на алгоритме C5.0. При этом могут быть использованы данные как клинической симптоматики,

Таблица 4

Использование атрибутов в дереве решений с учетом данных клинической симптоматики и иммунологических параметров (рис. 5)

Атрибут	Использование в дереве рис. 3, %
Боль в животе	100,00
IgA	87,78
IgM	56,67
IgE	28,89
CD16	20,00
Тошнота	4,44

так и иммунологических исследований. Для каждого из видов исследований может быть построено отдельное дерево.

Анализ построенных деревьев доказал, что погрешность диагностирования коморбидных состояний уменьшалась при совместном использовании данных разного вида исследований — например, клинической симптоматики и показателей клеточного и гуморального иммунитета.

В перспективе дальнейших исследований планируем использовать другие алгоритмы data mining, например, метод последовательного покрытия с построением классификационных правил.

Ключові слова: хронічний панкреатит, аскарідоз, коморбідність, клінічна симптоматика, імунограма, клітинний імунітет, гуморальний імунітет, дерево рішень, data mining, R, C5.0.

Приложение

> summary(treeModel_clinsymp)

```
Call:
C5.0.default(x = df[, 1:17], y = factor(df$diagnose))
C5.0 [Release 2.07 GPL Edition] Fri Feb 02 18:14:28 2018
```

```
-----
Class specified by attribute `outcome'
Read 90 cases (18 attributes) from undefined.data
```

```
Decision tree:
abdominal_pain = no: P (11/1)
abdominal_pain = yes:
...general_weakness = no: [S1]
  general_weakness = yes:
  ...nausea = yes: PandA (26/1)
    nausea = no:
    ...bad_appetite = no: PandA (18/2)
      bad_appetite = yes:
      ...Age <= 41: P (5)
        Age > 41:
        ...Sex = m: PandA (4)
          Sex = f:
          ...eructation = no: PandA (2)
            eructation = yes: P (2)
```

```
SubTree [S1]
the_difficulty_in_the_right_and_or_the_left_hypochondrium
= yes: PandA (9/2)
the_difficulty_in_the_right_and_or_the_left_hypochondrium
= no:
...Sex = f:
  ...disease_duration <= 10: P (4/1)
  : disease_duration > 10: PandA (2)
Sex = m:
  ...heartburn = yes: P (3)
  heartburn = no:
  ...disease_duration <= 9: P (2)
  disease_duration > 9: PandA (2)
```

Evaluation on training data (90 cases):

Decision Tree

```
-----
Size  Errors
13   7(7.8%) <<
(a) (b) <-classified as
----  ----
25    5 (a): class P
  2   58 (b): class PandA
```

Attribute usage:

```
100.00% abdominal_pain
87.78%  general_weakness
63.33%  nausea
34.44%  bad_appetite
24.44%
```

the_difficulty_in_the_right_and_or_the_left_hypochondrium

```
23.33% Sex
14.44% Age
11.11% disease_duration
7.78%  heartburn
4.44%  eructation
```

> summary(treeModel_cellImmun)

```
Call:
C5.0.default(x = df[, 32:36], y = factor(df$diagnose))
C5.0 [Release 2.07 GPL Edition] Fri Feb 02 18:14:28 2018
```

```
-----
Class specified by attribute `outcome'
Read 90 cases (6 attributes) from undefined.data
```

```
Decision tree:
CD16 <= 9:
...CD8 > 13: PandA (14/5)
: CD8 <= 13:
: ...CD22 <= 24: P (11/2)
:   CD22 > 24: PandA (2)
CD16 > 9:
...CD3 <= 49: PandA (46/7)
  CD3 > 49:
  ...CD16 <= 10: PandA (4/1)
  CD16 > 10:
  ...CD8 <= 14: PandA (4/1)
  CD8 > 14: P (9/2)
```

Evaluation on training data (90 cases):

Decision Tree

```
-----
Size  Errors
 7  18(20.0%) <<
(a) (b) <-classified as
----  ----
16  14 (a): class P
  4  56 (b): class PandA
```

Attribute usage:

```
100.00% CD16
70.00%  CD3
44.44%  CD8
14.44%  CD22
```

```
> summary(treeModel_humorImmun)
Call:
C5.0.default(x = df[, 37:40], y = factor(df$diagnose))
C5.0 [Release 2.07 GPL Edition] Fri Feb 02 18:14:28 2018
```

```
-----
Class specified by attribute `outcome'
Read 90 cases (5 attributes) from undefined.data
Decision tree:
IgE > 125: PandA (25)
IgE <= 125:
...IgA > 3.8: PandA (16/1)
  IgA <= 3.8:
  ...IgM > 2.3: P (28/4)
    IgM <= 2.3:
    ...IgG > 12.2: PandA (11)
      IgG <= 12.2:
      ...IgG <= 11.5: PandA (4)
        IgG > 11.5: P (6/1)
```

Evaluation on training data (90 cases):

```
Decision Tree
-----
Size Errors
 6 6( 6.7%) <<
(a) (b) <-classified as
-----
 29 1 (a): class P
 5 55 (b): class PandA
Attribute usage:
100.00% IgE
 72.22% IgA
 54.44% IgM
 23.33% IgG
```

```
> summary(treeModel_immunogram)
Call:
C5.0.default(x = df[, 32:40], y = factor(df$diagnose))
C5.0 [Release 2.07 GPL Edition] Fri Feb 02 18:14:28 2018
```

```
-----
Class specified by attribute `outcome'
Read 90 cases (10 attributes) from undefined.data
Decision tree:
IgE > 125: PandA (25)
IgE <= 125:
...IgA > 3.8: PandA (16/1)
  IgA <= 3.8:
  ...IgM > 2.3: P (28/4)
    IgM <= 2.3:
    ...CD16 <= 9: P (3)
      CD16 > 9: PandA (18/2)
```

Evaluation on training data (90 cases):

```
Decision Tree
-----
Size Errors
 5 7(7.8%) <<
(a) (b) <-classified as
-----
 27 3 (a): class P
 4 56 (b): class PandA
```

```
Attribute usage:
100.00% IgE
 72.22% IgA
 54.44% IgM
 23.33% CD16
```

```
> summary(treeModel)
Call:
C5.0.default(x = df[, c(1:17, 32:40)],
y = factor(df$diagnose))
C5.0 [Release 2.07 GPL Edition] Fri Feb 02 18:14:28 2018
```

```
-----
Class specified by attribute `outcome'
Read 90 cases (27 attributes) from undefined.data
Decision tree:
```

```
abdominal_pain = no: P (11/1)
abdominal_pain = yes:
...IgA > 3.8: PandA (28)
  IgA <= 3.8:
  ...IgM <= 2.3: PandA (25/3)
    IgM > 2.3:
    ...IgE > 105: PandA (8/1)
      IgE <= 105:
      ...CD16 <= 11: P (14)
        CD16 > 11:
        ...nausea = no: P (2)
          nausea = yes: PandA (2)
```

Evaluation on training data (90 cases):

```
Decision Tree
-----
Size Errors
 7 5(5.6%) <<
(a) (b) <-classified as
-----
 26 4 (a): class P
 1 59 (b): class PandA
```

```
Attribute usage:
100.00% abdominal_pain
 87.78% IgA
 56.67% IgM
 28.89% IgE
 20.00% CD16
 4.44% nausea
```

ЛИТЕРАТУРА

1. Babinets L., Kotsaba Y., Jugheli K. The Correlations Between Trophology Status and the Main Disease Characteristics of Patients with Biliary Genesis Chronic Pancreatitis. *Bulletin of Akaki Tsereteli state University*. 2016. № 2 (8). P. 180–184.
2. Бабінець Л. С., Дроняк Ю. В. Аскаридоз і хронічний панкреатит: спільні етіопатогенетичні аспекти, оптимізація

лікування в практиці сімейного лікаря. *Сімейна медицина*. 2014. № 2. С. 101–104.

3. Губергріц Н. Б., Скопиченко С. В. Практична панкреатологія. Донецьк : Либідь, 2007. 244 с.

4. Бабінець Л. С., Сабат З. І., Шайген О. Р. Синдром ендогенної інтоксикації при хронічному панкреатиті та коморбідних станах. *Ліки України плюс*. 2017. № 3 (32). С. 27–29.

5. Сучасні методи дослідження в гастроентерології: метод. посібник / Ю. М. Степанов та ін. Дніпропетровськ : Тип. Свідлер, 2006. 155 с.

6. Buchler M. W., Uhl W., Malfertheiner P., Sarr M. G. Diseases of pancreas. *Karger* 2014. 212 p.

7. Kocher H. M. Chronic pancreatitis. *Am. Fam. Physician*. 2008. Vol. 77, № 5. P. 661–662.

8. Pancreatic duct ascariasis: sonographic diagnosis — a case report / A. Agarwal et al. *Trop. Gastroenterol*. 2005. Vol. 26, № 4. P. 197–198.

9. Martsenyuk V. P., Andrushchak I. Y., Gvozdetska I. S. Qualitative analysis of the antineoplastic immunity system on the basis of a decision tree. *Cybern Syst Anal*. 2015. Vol. 51. P. 461. <https://doi.org/10.1007/s10559-015-9737-6>.

10. Qualitative analysis of compartmental dynamic system using decision-tree induction / A. S. Lyapandra et al. Proceedings of the 2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2015. Vol. 2, art. no. 7341391. P. 688–692. DOI: 10.1109/IDAACS.2015.7341391.

11. On multivariate method of qualitative analysis of Hodgkin-Huxley model with decision tree induction (2017) / V. Martsenyuk et al. International Conference on Control, Automation and Systems, art. no. 7832365. P. 489–494. DOI: 10.1109/ICCAS.2016.7832365.

Поступила в редакцію 06.02.2018

Рецензент д-р мед. наук, проф. Л. С. Годлевский,
дата рецензії 26.02.2018

УДК 616.37-002.2-06:616.995.132.8]-071/-078-047.44

В. П. Марценюк, Л. С. Бабінець, Ю. В. Дроняк

ИСПОЛЬЗОВАНИЕ ТЕХНОЛОГИИ DATA MINING С ЦЕЛЬЮ ДИФФЕРЕНЦИАЛЬНОЙ ДИАГНОСТИКИ КОМОРБИДНЫХ СОСТОЯНИЙ ХРОНИЧЕСКОГО ПАНКРЕАТИТА И АСКАРИДОЗА НА ОСНОВАНИИ ДАННЫХ КЛИНИЧЕСКОЙ СИМПТОМАТИКИ И ИММУНОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ

Для диагностирования коморбидности хронического панкреатита и аскаридоза использована методика построения дерева решений, основанная на алгоритме C5.0. При этом могут быть использованы данные как клинической симптоматики, так и иммунологических исследований. Для каждого из видов исследований (клинические симптомы, показатели клеточного и гуморального иммунитета), а также для их совокупности построены отдельные деревья решений. Исследована погрешность алгоритма.

Ключевые слова: хронический панкреатит, аскаридоз, коморбидность, клиническая симптоматика, иммунограмма, клеточный иммунитет, гуморальный иммунитет, дерево решений, data mining, R, C5.0.

UDC 616.37-002.2-06:616.995.132.8]-071/-078-047.44

V. P. Martsenyuk, L. S. Babinets, Yu. V. Dronyak

APPLICATION OF DATA MINING TECHNOLOGY WITH AIM OF DIFFERENTIAL DIAGNOSTICS OF THE COMORBIDITY OF CHRONIC PANCREATITIS AND ASCARIDOSIS ON THE BASIS OF DATA OF CLINICAL SYMPTOMATOLOGY AND IMMUNOLOGICAL RESEARCHES

For diagnostics of comorbidity of chronic pancreatitis and ascariidosis methodology of construction of decision tree is used, which is based on C5.0 algorithm. Data of both clinical symptomatology and immunological research can be used. For each of types of researches (clinical symptoms, cell immunity, humoral immunity) and also for their totality the separate decision trees are built. The error of algorithm is investigated.

Key words: chronic pancreatitis, ascariidosis, comorbidity, clinical symptomatology, immunogram, cell immunity, humoral immunity, decision tree, data mining, R, C5.0.

Передплатуйте
і читайте
журнал



ДОСЯГНЕННЯ
БІОЛОГІЇ та МЕДИЦИНИ

У випусках журналу:

Передплата приймається
у будь-якому передплатному
пункті

Передплатний індекс 08205

- ◆ Фундаментальні проблеми медицини та біології
- ◆ Нові медико-біологічні технології
- ◆ Оригінальні дослідження
- ◆ Огляди
- ◆ Інформація, хроніка, ювілеї