

Метаболические факторы риска и формирование мочевых камней. Исследование IV: Прогнозирование химического состава камня *in vivo* по метаболическим показателям

С.А. Голованов, А.В. Сивков, М.Ю. Просяников, В.В. Дрожжева

НИИ урологии и интервенционной радиологии им. Н.А. Лопаткина – филиал ФГБУ «НМИЦ радиологии» Минздрава России

Сведения об авторах:

Голованов С.А. – д.м.н., зав. научно-лабораторным отделом НИИ урологии и интервенционной радиологии им. Н.А. Лопаткина – филиал ФГБУ «НМИЦ радиологии» Минздрава России; e-mail: sergeyGol124@mail.ru

Golovanov S.A. – Dr. Sc., head of scientific Laboratory Department N.A. Lopatkin Scientific Research Institute of Urology and Interventional Radiology – Branch of the National Medical Research Centre of Radiology of the Ministry of Health of Russian Federation; e-mail: sergeyGol124@mail.ru

Сивков А.В. – к.м.н., заместитель директора научно-исследовательского института урологии и интервенционной радиологии им. Н.А. Лопаткина – филиал ФГБУ «НМИЦ радиологии» Минздрава России; e-mail: uroinfo@yandex.ru

Sivkov A.V. – PhD, assistant director of N.A. Lopatkin Scientific Research Institute of Urology and Interventional Radiology – Branch of the National Medical Research Centre of Radiology of the Ministry of Health of Russian Federation; e-mail: uroinfo@yandex.ru

Просяников М.Ю. – к.м.н., зав. отделом мочекаменной болезни НИИ урологии и интервенционной радиологии им. Н.А. Лопаткина – филиал ФГБУ «НМИЦ радиологии» Минздрава России; e-mail: prosyannikov@gmail.com

Prosiannikov M.Yu. – PhD, Head of Department of urolithiasis of N.A. Lopatkin Scientific Research Institute of Urology and Interventional Radiology – Branch of the National Medical Research Centre of Radiology of the Ministry of Health of Russian Federation; e-mail: prosyannikov@gmail.com

Дрожжева В.В. – старший научный сотрудник научно-лабораторного отдела НИИ урологии и интервенционной радиологии им. Н.А. Лопаткина – филиал ФГБУ «НМИЦ радиологии» Минздрава России; e-mail: drozhzhevav@mail.ru

Drozhzheva V.V. – researcher of scientific Laboratory Department of N.A. Lopatkin Scientific Research Institute of Urology and Interventional Radiology – Branch of the National Medical Research Radiological Centre of the Ministry of Health of Russian Federation; e-mail: drozhzhevav@mail.ru

Эпидемиологические исследования, проведенные в ряде индустриально развитых стран в последние годы, указывают на устойчивую тенденцию к росту частоты случаев мочекаменной болезни (МКБ) среди населения по сравнению с предыдущим периодом [1,2]. В России число впервые выявленных случаев МКБ на 100 000 населения за последние годы также заметно выросло с 153,7 в 2002 г. до 182,5 в 2016 г. [3,4].

Важное значение для выбора инвазивных методов лечения и для метафилактики имеют данные о химическом составе мочевого камня [5]. Так, уратные камни, имеющие низкую плотность, достаточно легко разрушаются при дистанционной ударно-волновой литотрипсии (ДУВЛ), по сравнению с оксалатными и фосфатными конкрементами, тогда как брушитные мочевые камни с трудом поддаются фрагментированию при литотрипсии [6-8].

Однако доступные для анализа мочевые камни удается получить только при эндоскопических операциях или в случаях их спонтанного отхождения. Тем не менее, предпринимаются попытки использования двухэнергетической компьютерной томографии для определения минерального состава мочевых камней *in vivo* с многообещающими начальными результатами [9,10]. В то же время, нередко литолитическая терапия является эффективной неинвазивной альтернативой хирургическому лечению уратного уролитиаза [11-13].

Метаболическим нарушениям литогенного характера придает особая роль в развитии мочекаменной болезни, поскольку мочевые камни в значительной степени представляют собой результат длительного воздействия метаболических и физико-химических факторов, ответственных за камнеобразование, известных как факторы риска мочекаменной болезни. Эти факторы

проявляют себя в виде нарушенной экскреции с мочой комплекса камнеобразующих веществ и ионов, сдвигов pH мочи и низкого диуреза [5].

Поэтому основной целью противорецидивного лечения МКБ является снижение риска возникновения мочевых камней путем устранения или ослабления воздействия метаболических литогенных факторов риска [14]. Поскольку метафилактика и мониторинг пациента проводятся в течение длительного периода времени, не исключена вероятность изменения характера обмена веществ и риск формирования мочевого камня другого метаболического типа. В этом случае требуется коррекция противорецидивного лечения в нужном направлении, на основе всесторонней оценки измененных метаболических факторов, участвующих в литогенезе.

Следует отметить, что значения величин показателей экскреции

основных метаболических литогенных факторов, принятые в качестве референсных для больных МКБ, не отражают в полной мере истинной направленности литогенеза. Каждый такой фактор, взятый в отдельности, не может служить надежным предиктором формирования камня определенного химического состава (типа), поскольку характер взаимного влияния друг на друга метаболических и физико-химических факторов мочи достаточность сложен и конечный результат с формированием камня другого химического состава может оказаться неожиданным [15-17].

Отсюда понятен интерес исследователей к поиску методов прогнозирования состава мочевых камней по комплексу метаболических показателей, включая показатели суточной экскреции мочи. Однако работ, посвященных этой проблеме, известно немного, что, по-видимому, связано с низким качеством прогноза, не удовлетворяющим требованиям клинической практики [18-21].

В последние годы активно развиваются особые методы анализа данных, так называемый Data mining (добыча данных, интеллектуальный анализ данных), использующий весь потенциал современных методов классификации, моделирования и прогнозирования для выявления в изучаемом наборе данных новых сведений, которые не могут быть получены путем простого статистического анализа. Другими словами, методы Data mining являются средством извлечения из набора данных ранее неизвестных скрытых взаимосвязей между переменными, характеризующих объект, предоставляя исследователю новые, в практическом смысле полезные и доступные для интерпретации знания, необходимые для принятия решений в различных сферах человеческой деятельности, в том числе и в медицинской диагностике.

Основой прогностической аналитики является математическое

моделирование, то есть создание модели, как набора правил, формул или уравнений, извлеченных из исходных данных и позволяющих генерировать предсказания на этой основе.

Одним из распространенных методов прогностического моделирования является метод классификации, реализованный во многих компьютерных системах клинического прогнозирования для оценки вероятности и риска возникновения события. Модели классификации используют значения одного или нескольких входных полей (независимых переменных или предикторов), чтобы предсказать значения одного или нескольких выходных целевых полей, или полей назначения (зависимых переменных, спрогнозированных моделью на основе анализа входных предикторов). Во многих областях медицины используются различные способы моделирования и алгоритмы машинного обучения: деревья решений (алгоритмы дерева C&R, QUEST, CHAID и C5.0), регрессии (линейная, логистическая, обобщенная линейная и Кокса), нейронные сети, модели опорных векторов и Байесовские сети [22-24].

Целью данной работы являлось исследование возможности применения методов Data mining для наиболее полной оценки состояния метаболизма у пациентов с МКБ, выявления значимых в литогенезе факторов и скрытых их взаимосвязей, с помощью которых можно было бы достаточно надежно прогнозировать вероятность формирования мочевых камней того или иного метаболического типа и химического состава.


МАТЕРИАЛЫ И МЕТОДЫ

Обследовано 708 больных МКБ (305 мужчин и 403 женщин) в возрасте от 17 до 78 лет, проходивших обследование и лечение в НИИ урологии и интервенционной ра-

диологии им. Н.А. Лопаткина Минздрава России (г. Москва) в период с 2010 г. по 2017 г. Минеральный состав мочевых конкрементов и метаболические показатели определяли методами описанными ранее [15]. Классификацию мочевых камней проводили по преобладающему минеральному компоненту (более 50% всей минеральной основы). Такой подход к классификации типов мочевых конкрементов является наиболее распространенным [25-27]. Интервал времени между выполнением анализа состава мочевого камня и биохимическими исследованиями крови и мочи не превышал 3-х месяцев.

Входными данными для анализа служили биохимические показатели сыворотки крови – общий кальций (Caser), мочева кислота (UAscr), фосфаты (Pser), магний (Mgser); биохимические показатели суточной экскреции с мочой кальция (CaUr), мочево кислоты (UAUr), фосфатов (Pur), магния (MgUr). Учитывали также такие физико-химические показатели мочи, как удельный вес (SpecGrav), pH мочи (pH), суточный объем (Diuresis), рассчитывали Индекс массы тела (Body Mass Index, BMI), представляющего собой отношение массы тела (кг) к квадрату роста пациента (м) [кг/м²].

Принимая во внимание, что степень перенасыщенности мочи и, следовательно, ее литогенный потенциал, в конечном счете определяется концентрацией экскретируемых с мочой веществ, в качестве дополнительных производных метаболических показателей (дополнительных входных переменных в моделях) рассчитывали также концентрационные показатели (в мМол/л), такие, как концентрация в моче кальция (CaUrC), мочево кислоты (UaUrC), фосфатов (PhUrC), магния (MgUrC).

В качестве зависимых переменных (выводных данных моделей) были выбраны метаболические типы мочевых камней, 

классифицируемые по преобладающему (более 50% минеральной основы) составу: оксалатные камни (Ox50), уратные камни, представленные мочевой кислотой и ее гидратом (Ur50), фосфатные из карбонатапатита (даллита, Dh) камни (Dh50), «инфекционные» камни, представленные струвитом (Str50) и уратом аммония (AmUr50). В отдельный класс (тип) Прочие (Others) были выделены смешанные 2-х и 3-х компонентные камни, в которых ни один из указанных компонентов не превышал 50% состава. Для достижения максимальной однородности набора данных из него были исключены случаи с цистиновыми камнями (7 пациентов) и 2 случая с брушитными камнями.

Полученный в конечном счете набор данных, включающих 17 независимых входных переменных и 6 категорий зависимых переменных, соответствие которым предсказывала модель, представлял собой файл программы IBM SPSS Statistica v 25 содержащий 708 записей (или случаев наблюдений).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Для выбора лучших моделей, оптимально и точно описывающих связь метаболических показателей с

химическим составом формирующихся мочевых камней, использовали программный построитель моделей IBM SPSS Modeler v18.0. IBM SPSS Modeler предлагает ряд методов моделирования, взятых из таких областей, как обучение машин, искусственный интеллект и статистика.

Решение задачи прогнозирования вероятности формирования мочевых камней того или иного химического состава требует выявления значимых в литогенезе метаболических факторов и выявления их сложных взаимосвязей, которые не удастся обнаружить обычными методами статистического анализа. Наиболее распространенными подходами в создании прогностических моделей являются методы классификации состояний объекта.

На первом этапе работы был использован специальный модуль IBM SPSS Modeler 18.0 (Узел автоклассификации), который генерирует набор моделей, использующих различные алгоритмы, и ранжирует лучшие кандидаты-модели в соответствии с критериями точности прогноза. Было показано, что наибольшей точностью предсказания (предикции) в 87,1% обладает модель, использующая алгоритм C5.0 (RuleQuest, Австралия), в отличие от других 10 алгоритмов машинного

обучения, точность которых составляла всего 39,5-54,2%,

Особенность алгоритма C5.0 состоит в том, что его можно использовать для построения как дерева решений, так и набора правил. Для повышения степени точности алгоритма C5.0 использовали режим бустинга, (boosting) позволяющий последовательно создавать несколько моделей, где каждая последующая модель учитывает ошибки классификации предыдущей. Это существенно повышает конечную точность модели C5.0. Проводили испытание 10-ти последовательно создаваемых моделей.

Кроме того, для повышения точности классификации алгоритма C5.0 проводили 10-кратную перекрестную проверку (10-fold cross-validation method), используя набор из 10 моделей, построенных на подмножествах обучающих данных, чтобы оценить точность модели, построенной на полном наборе данных.

Построенная таким образом модель C5.0 обладала высокой точностью предсказания, распределяя типы камней на классы, согласно заранее установленным пользователем критериям (табл. 1). Неверное определение типа камня наблюдалось лишь в 2-х случаях из 708 (0,28%). Для более детального анализа качества модели алгоритма C5.0 проводили ее дальнейшее тестирование.

С помощью модуля Узел раздела полученный набор подготовленных и очищенных данных, содержащий 708 записей (случаев или наблюдений) был разделен на две рандомизированные выборки. Большая часть основного набора данных [70% (n=499)] использовалась для обучения модели, а меньшая ее часть [30% (n=209)] использовалась для проверки модели. Разделение данных на две выборки позволяло обучать модель на одной выборке и тестировать на другой. Сгенерировав модель при помощи одной выборки и испытав ее при помощи другой, отдельной выборки, можно

Таблица 1. Матрица совпадения предсказаний, сделанных моделью (SC-NewbgrN, алгоритм C5.0), в сравнении с фактическим распределением типов камней (пользовательской классификацией)

Результаты для поля вывода NewbgrN

Общие результаты

Сравнение SC-NewbgrN с NewbgrN

Верно	706	99,72%
Неверно	2	0,28%
Всего	708	

Матрица совпадений для SC-NewbgrN (в строках показаны фактические значения)

	AmUr50	Dh50	Others	Ox50	Str50	Ur50
AmUr50	7	0	0	0	0	0
Dh50	0	221	0	1	0	0
Others	0	0	24	0	0	0
Ox50	0	1	0	262	0	0
Str50	0	0	0	0	56	0
Ur50	0	0	0	0	0	136

Примечание: Обозначения классов (типов) мочевых камней: Ox50 - оксалатные камни, Ur50 - уратные камни, Dh50 - фосфатные камни из карбонатапатита, Str50 - камни из струвита, AmUr50 - камни из урата аммония, Others (прочие или смешанные 2-х и 3-х компонентные камни, в которых доля любого компонента не превышала 50% состава)

было сделать вывод об использовании этой прогностической модели в других наборах данных, подобных тому набору данных, который служил для построения настоящей модели.

Модель машинного обучения, построенная на алгоритме C5.0, показывала высокую точность прогнозирования при классификации типов мочевого камней (табл. 2). В обучающей выборке (n=499) отсут-

ствовали ошибки классификации (предикции). В проверочной выборке (n=209) наблюдались две ошибки ложного отнесения: один карбонатапатитный камень (Dh50) был отнесен к группе оксалатных камней (Ox50), и один оксалатный камень (Ox50) был классифицирован моделью как карбонатапатитный (Dh50) (табл. 2).

Таким образом, модель может с высокой точностью предсказывать состав конкрементов в проверочной выборке. Единичные ошибки классификации относились к группе кальций-содержащих камней смешанного состава, представленных оксалатами и фосфатами (карбонатапатитом). По-видимому, эти ошибки классификации обусловлены тем, что оксалатные камни и оксалатно-карбонатные камни имеют некоторые общие механизмы литогенеза, связанные с нарушениями обмена камнеобразующих веществ [15-17].

Для более детальной оценки прогностических возможностей других моделей, проводили скрининг-анализ одиннадцати моделей, построенных на различных прикладных алгоритмах машинного обучения. Использовалась обучающая выборка (n=499). Сравнение проводили с фактическими значениями проверочной выборки (n=209) (табл. 3).

Таблица 2. Результаты классификации типов камней в Обучающей (n=499) и Проверочной (n=209) выборках, полученные моделью машинного обучения \$C-New6grN, алгоритм C5.0

Результаты для поля вывода New6grN

Общие результаты

Сравнение \$C-New6grN с New6grN

'Подмножества'	1_Обучающее		2_Проверка	
Верно	499	100%	207	99,04%
Неверно	0	0%	2	0,96%
Всего	499		209	

Матрица совпадений для \$C-New6grN (в строках показаны фактические значения)

'Подмножества' = 1_Обучающее	AmUr50	Dh50	Others	Ox50	Str50	Ur50
AmUr50	5	0	0	0	0	0
Dh50	0	150	0	0	0	0
Others	0	0	13	0	0	0
Ox50	0	0	0	190	0	0
Str50	0	0	0	0	42	0
Ur50	0	0	0	0	0	99

'Подмножества' = 2_Проверка	AmUr50	Dh50	Others	Ox50	Str50	Ur50
AmUr50	2	0	0	0	0	0
Dh50	0	71	0	1	0	0
Others	0	0	11	0	0	0
Ox50	0	1	0	72	0	0
Str50	0	0	0	0	14	0
Ur50	0	0	0	0	0	37

Поля вывода*	Кол-во в наборе данных (n [%])	Выборка			
		Обучающая 70% (n=499)		Проверочная 30% (n=209)	
		Верно (n, [%])	Неверно (n, [%])	Верно (n, [%])	Неверно (n, [%])
AmUr50	7 [1,0%]	5 [100%]		2 [100%]	
Dh50	222 [31,4%]	150 [100%]		71 [98,6%]	1 [1,4%]
Others	24 [3,4%]	13 [100%]		11 [100%]	
Ox50	263 [37,1%]	190 [100%]		72 [98,6%]	1 [1,4%]
Str50	56 [7,9%]	42 [100%]		14 [100%]	
Ur50	136 [19,2%]	99 [100%]		37 [100%]	
Всего	708 [100%]	499 [100%]	0 [0,0%]	207 [99,04%]	2 [0,96%]

Примечание: * обозначения типов (классов) камней те же, что и в таблице 1

Таблица 3. Сравнение общих результатов прогноза различных прикладных алгоритмов машинного обучения с фактическими значениями проверочной выборки (n=209)

Алгоритмы моделей#	Выборка					
	Обучающая 70% (n=499)			Проверочная 30% (n=209)		
	Верно (n, [%])	Неверно (n, [%])	Всего (n)	Верно (n, [%])	Неверно (n, [%])	Всего (n)
C5.0	499 100,0%	0 0,0%	499	207 99,06%	2 0,96%	209
SVM	402 80,56%	97 19,44%	499	168 80,38%	41 19,62%	209
CHAID исчерпывающий	362 72,55%	137 27,45%	499	147 70,33%	62 29,67%	209
Случайные деревья	359 71,94%	140 28,06%	499	142 67,94%	67 32,06%	209
Нейронная сеть	341 68,34%	158 31,66%	499	73 34,93%	136 65,07%	209
C&RT	312 62,53%	187 37,47%	499	112 53,59%	97 46,41%	209
Байесовская сеть	291 58,32%	208 41,68%	499	126 60,29%	83 39,71%	209
Логистическая регрессия	252 50,5%	247 49,5%	499	102 48,8%	107 51,2%	209
Дискриминантный анализ	252 50,5%	247 49,5%	499	142 67,94%	67 32,06%	209
KNN	246 49,3%	253 50,7%	499	99 47,37%	110 52,63%	209
QUEST	241 48,3%	258 51,7%	499	94 44,98%	115 55,02%	209

Примечание: # - полные названия условных сокращений некоторых алгоритмов:

C5.0 – особый алгоритм для построения деревьев решений; **SVM** – механизм опорных векторов, один из способов классификации и регрессии, **CHAID** - алгоритм генерирования дерева решений с использованием статистик хи-квадрат, **C&RT** - алгоритм построения дерева решений для предсказания будущих наблюдений, **KNN** - метод классификации на основе сходства наблюдений, **QUEST** - метод бинарной классификации для построения деревьев решений

Можно видеть, что наибольшей точностью предикции обладает модель, построенная на алгоритме C5.0, с общей точностью классификации в проверочной выборке 99,06%. Заметно меньшую точность классификации в 70-80% демонстрировали модели алгоритмов SVM, CHAID исчерпывающий и Random Forest (Случайные деревья).

Проверочная выборка (n=209) использовалась также для оценки качества прогностических возможностей моделей по показателям чувствительности, специфичности и точности в отношении отдельных типов (классов) мочевых камней. Сравнение точности прогнозирования различных моделей машинного обучения при классификации типов мочевых камней свидетельствует о явных преимуществах модели, построенной на алгоритме C5.0. В некоторых моделях (C&RT, CHAID, QUEST, Random Forest) показатель чувствительности достигал 100%, однако это наблюдалось при верной классификации только двух из трех основных классов камней (Ox50, Dh50, Ur50), составляющих 34,9%, 34,4% и 17,7% соответственно от всех мочевых камней в проверочной выборке (n = 209), то есть, суммарно представляющих большую часть мочевых камней – 87,0% (табл. 2). Точность прогноза мочевых камней малочисленных классов Str50, Others и AmUr50, на долю которых во всей проверочной выборке (n=209) приходилось 7,9%, 4,7% и 0,8% от всех камней соответственно, в большинстве исследуемых моделей была низкой, за исключением модели C5.0.

Модель машинного обучения, построенная на алгоритме C5.0 показывала самые высокие показатели чувствительности (97,5-100%) и специфичности (99,3-100%), а также высокую точность прогнозирования при классификации всех шести типов мочевых камней (99,5-100%). По сравнению с другими моделями модель C5.0 показывала самый высокий результат верных прогнозов (100% распознавание) типа мочевых

камней не только в общей проверочной выборке (n=209), но также давала 100% верных прогнозов внутри каждого соответствующего класса конкрементов, включая плохо распознаваемые другими моделями малочисленные типы камней (Str50, AmUr50, Others).

Дальнейшая задача состояла в том, чтобы объективно оценить прогностические возможности модельного алгоритма C5.0, используя новый набор данных, не имеющий в целевом поле меток-идентификаторов типа камня; то есть, протестировать модель на наборе данных, не связанном с главным набором данных, послужившим для построения модели.

В связи с этим из основного набора данных (n=708) было выделено шесть небольших рандомизированных проверочных выборок, каждая из которых соответствовала своему типу мочевых камней. Пять таких выборок были представлены десятью случаями в своем классе (типе) мочевых камней (Ox50 [n=10]; Dh50 [n=10]; Ur50

[n=10]; Str50 [n=10]; Others [n=10]). Камней из урата аммония (AmUr50) насчитывалось всего 7 случаев в основном наборе данных (n=708), то есть 1% от общего количества камней. Поэтому в проверочную выборку вошло всего 3 случая (AmUr50) [n=3]), а остальные 4 случая были включены в новый модельный набор данных (n=655). Шесть проверочных выборок (n=53) были извлечены из основного набора данных (n=708). Оставшихся данные использовались для формирования нового набора данных (n=655), на котором была построена новая модель C5.0.

Каждая из тестирующих выборок не имела в своем целевом поле идентификаторов, определяющих тип мочевых камней, и поэтому воспринималась моделью как новый набор неизвестных данных. Это позволяло более объективно оценить точность предсказания моделью C5.0 метаболического типа мочевого камня *in vivo*.

Результаты, представленные в таблице 4, свидетельствуют о высо-

Таблица 4. Результаты классификации типов камней в шести тестовых выборках (всего n=53), полученные моделью машинного обучения C5.0 на обучающей выборке (n=655)

Результаты для поля вывода New6grN

Общие результаты

Сравнение \$C-New6grN с New6grN

Верно	655	92,51%
Неверно	53	7,49%
Всего	708	

Матрица совпадений для \$C-New6grN (в строках показаны фактические значения)

	AmUr50	Dh50	Others	Ox50	Str50	Ur50
AmUr50	4	0	0	0	0	0
Dh50	0	212	0	0	0	0
Others	0	0	14	0	0	0
Ox50	0	0	0	253	0	0
Str50	0	0	0	0	46	0
Ur50	0	0	0	0	0	126
\$null\$	0	15	1	23	4	10

Тип (класс) камней	Модель C5.0 (\$null\$), n	Фактически, n	Общая тестовая выборка (n=53)				Se (%)	Sp (%)	Ac (%)
			ИП	ЛО	ЛП	ИО			
AmUr50	0	3	0	3	0	50	0,0	100,0	94,3
Dh50	15	10	10	0	5	43	100,0	89,6	91,4
Others	1	10	1	9	0	43	10,0	100,0	83,0
Ox50	23	10	10	0	13	43	100,0	76,8	80,3
Str50	4	10	4	6	0	43	40,0	100,0	88,7
Ur50	10	10	10	0	0	40	100,0	100,0	100,0
Всего	53	53	35	18	18	0	66,0	0,0	49,3

Примечание: \$null\$ - созданное моделью C5.0 (\$C-New6grN) поле для записи классифицированных ею типов мочевых камней; Se, Sp, Ac - показатели чувствительности (Se), специфичности (Sp) и точности (Ac) модельного прогноза соответственно; ИП, ЛО, ЛП, ИО - истинно положительные, ложно отрицательные, ложно положительные и истинно отрицательные результаты соответственно. AmUr50, Dh50, Others, Ox50, Str50, Ur50 - обозначения типов (классов) мочевых камней те же, что и в Таблице 1

ких показателях чувствительности, специфичности и точности модельного прогноза у трех основных типов камней (Ox50, Dh50, Ur50), составляющих большую часть (87,4%) всех мочевых камней. В тестовой выборке из 10 камней класса (типа) Str50 точное распознавание наблюдалось только в 4 случаях (прогностическая чувствительность 40%, специфичность 100%, точность 88,7%, табл. 4). По-видимому, это связано с низкой весовой долей струвитных камней, составляющей всего 8,1% от общего количества мочевых камней в модели. Можно полагать, что это обстоятельство объясняет также все ложно отрицательные результаты предсказания камней типа Others (табл. 4), представляющих всего 3,4% от всех мочевых камней. Следует отметить, что в проверочной выборке из 10 камней типа Others среднее содержание оксалатов (Ox) и карбонатапатита (Dh) в камнях приближалось к 50% (составляло 43,3% и 41,7% соответственно). Возможно, именно поэтому 90% камней класса Others были ошибочно распределены моделью между классами Ox50 и Dh50 (табл. 4). Как отмечалось, камни из урата аммония (AmUr50) встречаются крайне редко, всего 1% от общего количества мочевых конкрементов (табл. 2). Поэтому модельный прогноз формирования такого типа камней, выполняемый на малых тестовых выборках, пока недостаточно надежен.

Немногочисленные работы, посвященные прогнозированию химического состава камня *in vivo* по метаболическим показателям, демонстрируют невысокую точность предикции.

Так, в работе D.M. Moreira и соавт. попытка прогнозирования состава мочевых камней по показателям суточной экскреции мочи с отнесения камней к одному из 4-х типов камней (оксалатным, уратным, фосфатным и смешанным) показывала точность предикции 61%, 69%, 56% и 71% для оксалатных, уратных, фосфатных и смешанных

камней, соответственно, а общая точность прогноза составляла всего 64% [19]. Сокращение типов классифицируемых камней до двух (уратных и оксалатных) позволяло повысить диагностическую чувствительность в прогнозировании до 87,5-100%, однако это касалось только уратных камней и специфичность при этом оставалась неприемлемо низкой (42-68%) [20,21]. Применяя метод логистической регрессии, авторам удалось достаточно надежно предсказывать лишь вероятность наличия уратного камня по сравнению с оксалатным конкрементом. В отношении других типов камней предикция была неудовлетворительной. Использование другого статистического метода – дискриминантного анализа в собственной модификации с разработкой оригинального набора решающих правил и развернутого алгоритма было предложено Н.Н. Поповкиным с соавт. [18]. В 58,1% случаев состав камня был определен точно, в 22,5% – удовлетворительно, то есть верно был распознан основной минеральный компонент. Авторы использовали набор данных из 125 наблюдений с разделением их по химическому составу камня на 8 групп, 6 из которых представляли собой различные сочетания смесей камней, иногда не связанных между собой общими механизмами литогенеза (например, группа камней струвит с карбонатапатитом, группа камней струвит с карбонатапатитом и уратами). Прогностическая точность применяемого в работе метода, по-видимому, была бы выше при увеличении численности набора исходных данных и при выборе более четких критериев разделения типов камней на классы, например, используя общепринятый подход по преобладающему минеральному компоненту.

Несмотря на явные преимущества методов Data Mining, не всегда имеется возможность использовать в диагностических и прогностических целях такой сложный про-

граммный продукт, как построитель моделей IBM SPSS Modeler v18.0.

Следует отметить, что тестируемый в настоящей работе алгоритм C5.0 позволяет использовать его для построения как дерева решений, так и набора правил, которые можно использовать автономно без построителя моделей IBM SPSS Modeler v18.0. Таким образом, можно создать систему классификации на основе модели C5.0 с построением Деревя решений, которое будет служить для предсказания или классификации будущих наблюдений на основе имеющегося набора решающих правил. При этом выбор заключительного диагностического решения может быть выполнен путем применения специального программного обеспечения.

Как показано в настоящей работе, использование самой модели C5.0, построенной на подготовленном наборе данных, намного удобнее, поскольку дает возможность получить надежный диагностический вывод высокого качества без значительных затрат времени на анализ.

В целом, предлагаемая модель машинного обучения, построенная на алгоритме C5.0 и использующая метаболические показатели мочи и крови пациента с МКБ, может использоваться в клинике для достаточно надежного диагностического прогнозирования (предсказания) *in vivo* минерального состава мочевых камней наиболее распространенных метаболических типов (оксалатов, уратов, фосфатов, представленных карбонатапатитом), на долю которых приходится более 87% от всех камней.

Менее надежный прогноз ожидаем в случаях, редко встречающихся 3-х компонентных камней, в составе которых отсутствует преобладающий минеральный компонент камня. Невысокая точность прогноза наличия «инфекционных» камней из струвита и аммония урата, по-видимому, отчасти также объясняется более редкой встречаемостью

камней такого типа. Однако, можно полагать, что для повышения точности прогноза образования этих камней в качестве дополнительных переменных-предикторов, помимо значений pH мочи, целесообразно включать в модель показатели наличия инфекции мочевых путей, данные двухэнергетической компьютерной томографии и, возможно, другие показатели.

Применение методов прогностической аналитики имеет боль-

шие возможности для оптимизации диагностики и лечения урологических заболеваний. Прогностическая модель, построенная на алгоритме C5.0, может быть эффективно использована для определения химического состава мочевого камня перед выполнением ДУВЛ. Кроме того, модельное прогнозирование следует проводить в случаях выявления литогенных нарушений обмена веществ, которые пока не

привели к формированию мочевого камня, но указывают на необходимость начала превентивной персонализированной метафилактики, направленной на предотвращение образования камня конкретного метаболического типа. Это особенно важно при мониторинге больных после удаления у них камня, поскольку мочекаменная болезнь характеризуется рецидивирующим течением. ■

Ключевые слова: мочекаменная болезнь, метаболизм, компьютерное моделирование, прогнозирование химического состава камней.

Key words: urolithiasis, metabolism, computer modeling, prediction of the chemical composition of stones.

Резюме:

Введение. Важное значение для выбора как инвазивных методов лечения мочекаменной болезни (МКБ), так и методов предупреждения повторного камнеобразования имеют данные о химическом составе мочевого камня, точнее о его метаболическом типе. Однако мочевые камни не всегда доступны для анализа, что затрудняет выбор оптимальных методов лечения и приводит к необходимости поиска (разработки) способов оценки химического состава мочевых камней у пациента *in vivo*. Одним из направлений поиска является всесторонний анализ литогенных метаболических факторов, длительное воздействие которых ведет к образованию мочевого камня.

Цель данной работы оценить возможность некоторых алгоритмов машинного обучения в прогнозировании метаболического типа мочевого камня *in vivo* по комплексу метаболических показателей.

Материал и методы. Исследован минеральный состав 708 мочевых конкрементов (от 305 мужчин и 403 женщин в возрасте от 16 до 81 лет), а также биохимические показатели сыворотки крови (кальций, мочевая кислота, фосфаты, магний) и показатели суточной экскреции с мочой кальция, мочевой кислоты, фосфатов, магния. Определяли удельный вес мочи, ее pH, суточный объем. Рассчитывали Индекс массы тела пациентов.

Результаты и обсуждение. Использование программного построителя моделей IBM SPSS Modeler v18.0 и подготовленного набора данных (n=708) позволило определить, что алгоритм C5.0 является лучшим для более точного предсказания химического типа мочевого камня по метаболическим показателям.

Совпадения предсказаний, сделанных моделью (алгоритм C5.0), в сравнении с фактическим распределением типов камней было практически абсолютным. Неверное определение типа камня наблюдалось лишь в 2-х случаях из 708 (0,28%).

При дальнейшем тестировании модели C5.0 с разделением основного набора данных (n=708) на две рандомизированные выборки Обучающую (70%, n=499) и Проверочную (30%, n=209) отмечено высокое качество предикции модели. В Проверочной выборке верное определение типа камня наблюдалось в 207 случаях из 209 (99,04%).

Summary:

Metabolic risk factors and urinary stone formation. Study IV: Prediction of the chemical composition of the stone *in vivo* according to metabolic parameters

S.A. Golovanov, A.V. Sivkov, M. Yu. Prosyannikov, V.V. Drozhzheva

Introduction. Chemical composition of the urinary stone, that is, of its metabolic type, are of great importance for the choice of both invasive methods of treatment of urolithiasis and methods of stone recurrence prevention. However, urinary stones are not always available for analysis, which makes it difficult to select the best treatment methods and leads us to the need to find methods for assessing the chemical composition of patient urinary stones in a patient *in vivo*. One of the search directions is a comprehensive analysis of lithogenic metabolic factors, the long-term effect of which leads to the urinary stone formation.

Goal. The purpose of this paper is to evaluate the possibility of some machine learning algorithms in predicting the metabolic type of urinary calculi *in vivo* in a set of metabolic parameters.

Material and methods. The mineral composition of 708 urinary calculi (from 305 men and 403 women aged 16 to 81 years), as well as biochemical parameters of blood serum (calcium, uric acid, phosphates, magnesium) and daily excretion in urine of calcium, uric acid, phosphates, magnesium. The specific gravity of urine, its pH, daily volume was determined. The body mass index of the patients was calculated too.

Results and discussion. Using the IBM SPSS Modeler v18.0 (software for models building) and the prepared data set (n = 708), it was possible to determine that the C5.0 algorithm is the best for accurate prediction of the chemical type of urinary stone by metabolic parameters.

The coincidence of predictions made by the model (algorithm C5.0), in comparison with the actual distribution of types of stones, was almost absolute. Incorrect definition of the type of stone was observed only in 2 cases out of 708 (0.28%).

Further testing of the C5.0 model with separation of the main data set (n = 708) into two randomized samples – Teaching (70%, n = 499) and Testing (30%, n = 209), marked the high quality of the model prediction. In the Testing sample the correct definition

Точности прогнозирования других моделей машинного обучения при классификации типов мочевых камней была заметно ниже (45-80%), что свидетельствует о явных преимуществах модели, построенная на алгоритме C5.0. Эта же модель показывала самый высокий процент верных распознаваний типов камней (100%), в то время как в других моделях машинного обучения верные предсказания типов камней составляли 34,8-90,2%.

Для практического применения в клинике модели C5.0 имеется возможность построения диагностического дерева решений или набора решающих правил, позволяющих выполнять модельный прогноз метаболического типа камня *in vivo*.

Заключение. Предлагаемая модель машинного обучения, построенная на алгоритме C5.0 и использующая метаболические показатели мочи и крови пациента с МКБ, может использоваться в клинике для достаточно надежного диагностического прогнозирования *in vivo* химического состава мочевых камней наиболее распространенных метаболических типов. Эта модель может быть использована для оценки химического состава мочевого камня перед выполнением литотрипсии, а также при мониторинге больных после удаления у них камня с целью прогнозирования риска формирования камня конкретного метаболического типа и своевременного начала проведения целенаправленной метафилактики.

Авторы заявляют об отсутствии конфликта интересов.

of the type of stone was observed in 207 cases out of 209 (99.4%).

The prediction accuracy of other machine learning models in classification of urinary stone types was significantly lower (45-80%), which indicates significant advantages of the model, built on C5.0 algorithm. The same model showed the highest percentage of correct stone type recognition (100%), while in other machine learning models the correct predictions of types of stones were 34.8-90.2%.

For practical clinical use of C5.0 model, there is an opportunity to build a diagnostic decision tree or create a set of decision rules that allow perform a model forecast of the metabolic type of stone *in vivo*.

Conclusion. The proposed machine learning model based on the C5.0 algorithm and using the metabolic parameters of urine and blood of a patient with urolithiasis can be used in the clinic for sufficiently reliable diagnostic prediction *in vivo* the chemical composition of urinary stones of the most common metabolic types. This model can be used for evaluation the urinary stone chemical composition before performing lithotripsy, as well as monitoring patients after removing stones from them in order to predict risk of forming a specific metabolic type stone and timely to initiate targeted metaphylaxis.

Authors declare lack of the possible conflicts of interests.

ЛИТЕРАТУРА

- Romero V, Akpınar H, Assimos DG. Kidney stones: a global picture of prevalence, incidence, and associated risk factors. *Rev Urol* 2010;12(2-3):e86-96.
- Sorokin I, Mamoulakis C, Miyazawa K, Rodgers A, Talati J, Lotan Y. Epidemiology of stone disease across the world. *World J Urol* 2017;35(9):1301-1320. doi: 10.1007/s00345-017-2008-6
- Заболеемость населения России в 2003 году: Статистические материалы. М.; 2004 г. С. 153-154
- Заболеемость населения в России в 2016. (электронная версия). URL: <http://mednet.ru/ru/statistika/zabolevaemost-naseleniya.html>
- Türk C, Neisius A, Petrik A, Seitz C, Skolarikos A, Tepeler A, Thomas K. EAU Guidelines on Urolithiasis. *EAU*, 2017 URL: http://uroweb.org/wp-content/uploads/Guidelines_WebVersion_Complete-1.pdf].
- Nakasato T, Morita J, Ogawa Y. Evaluation of Hounsfield Units as a predictive factor for the outcome of extracorporeal shock wave lithotripsy and stone composition. *Urolithiasis* 2015;43(1):69-75. doi: 10.1007/s00240-014-0712-x.
- Stewart G, Johnson L, Ganesh H, Davenport D, Smelser W, Crispin P, Venkatesh R. Stone size limits the use of Hounsfield units for prediction of calcium oxalate stone composition. *Urology* 2015;85(2):292-5. doi: 10.1016/j.urology.2014.10.006.
- Tailly T, Larish Y, Nadeau B, Violette P, Glickman L, Olvera-Posada D, et al. Combining mean and standard deviation of hounsfield unit measurements from preoperative CT allows more accurate prediction of urinary stone composition than mean hounsfield units alone. *J Endourol* 2016;30(4):453-9. doi: 10.1089/end.2015.0209.
- Kulkarni NM, Eisner BH, Pinho DE, Joshi MC, Kambadakone AR, Sahani DV. Determination of renal stone composition in phantom and patients using single-source dual-energy computed tomography. *J Comput Assist Tomogr* 2013;37(1):37-45. doi: 10.1097/RCT.0b013e3182720f66.
- Руденко В.И., Серова Н.С./, Капанадзе Л.Б. Возможности двухэнергетической компьютерной томографии в диагностике мочекаменной болезни. Материалы 3-й научно-практической конференции урологов Северо-Западного федерального округа РФ 20-21 апреля 2017 г. *Урологические ведомости* 2017;7(S): 92-93.
- Cicerello E, Merlo F, Maccatrozzo L. Urinary alkalization for the treatment of uric acid nephrolithiasis. *Arch Ital Urol Androl* 2010;82:145-148.
- Trinchieri A, Esposito N, Castelnuovo C. Dissolution of radiolucent renal stones by oral alkalization with potassium citrate/potassium bicarbonate. *Arch Ital Urol Androl* 2009;81:188-191.
- Аль-Шукри СХ, Слесаревская МН, Кузьмин ИВ. Литолитическая терапия уратного нефролитиаза. *Урология* 2016;(2):23-27
- Eisner BH, Sheth S, Dretler SP, Herrick B, Pais VM Jr. Abnormalities of 24-hour urine composition in first-time and recurrent stone-formers. *Urology* 2012;80(4):776-9. doi: 10.1016/j.urology.2012.06.034. Epub 2012 Aug 22
- Голованов СА, Сивков АВ, Дрожжева ВВ, Анохин НВ. Метаболические факторы риска и формирование мочевых камней. Исследование I: влияние кальцийурии и урикурии. *Экспериментальная и клиническая урология* 2017;(1):52-57
- Голованов СА, Сивков АВ, Дрожжева ВВ, Анохин НВ. Метабо-

ЛИТЕРАТУРА

- лические факторы риска и формирование мочевых камней. Исследование II: влияние фосфатурии и магниурии. *Экспериментальная и клиническая урология* 2017;(2):42 – 48
17. Голованов СА, Сивков АВ, Поликарпова АМ, Дрожжева ВВ, Андрияхин МИ, Просяников МЮ. Метаболические факторы риска и формирование мочевых камней. Исследование III: влияние рН мочи. *Экспериментальная и клиническая урология*, № 1, 2018; (1):84 – 91
18. Поповкин НН, Гришкова НВ, Чудновская МВ, Даренков АФ, Голованов СА. Дифференциальная диагностика состава мочевого камня in vivo по метаболическим показателям у больных нефролитиазом. Актуальные вопросы урологии и оперативной нефрологии: Сборник научных трудов. М., 1994. С. 32-42.
19. Moreira DM, Friedlander JI, Hartman C, Elsamra SE, Smith AD, Okeke Z. Using 24-hour urinalysis to predict stone type. *J Urol* 2013;190(6):2106-11. doi: 10.1016/j.juro.2013.05.115.
20. Torricelli FC, De S, Liu X, Calle J, Gebreselassie S, Monga M. Can 24-hour urine stone risk profiles predict urinary stone composition? *J Endourol* 2014;28(6):735-8. doi: 10.1089/end.2013.0769. Epub 2014 Feb 14.
21. Torricelli FC, Brown R, Berto FC, Tarplin S, Srougi M, Mazzucchi E, et al. Nomogram to predict uric acid kidney stones based on patient's age, BMI and 24-hour urine profiles: A multicentre validation. *Can Urol Assoc J* 2015;9(3-4):E178-82. doi: 10.5489/auaj.2682
22. Tseng CJ, Lu CJ, Chang CC, Chen GD, Cheewakriangkrai C. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. *Artif Intell Med* 2017;78:47-54. doi: 10.1016/j.artmed.2017.06.003.
23. Shahmoradi L, Langarizadeh M, Pourmand G, Fard ZA, Borhani A. Comparing Three Data Mining Methods to Predict Kidney Transplant Survival. *Acta Inform Med* 2016;24(5):322-327. doi: 10.5455/aim.2016.24.322-327.
24. Tavares M, Paredes S, Rocha T, Carvalho P, Ramos J, Mendes D, et al. Expert knowledge integration in the data mining process with application to cardiovascular risk assessment. *Conf Proc IEEE Eng Med Biol Soc* 2015;2015:2538-42. doi: 10.1109/EMBC.2015.7318909
25. Rendina D, De Filippo G, De Pascale F, Zampa G, Muscariello R, De Palma D, et al. The changing profile of patients with calcium nephrolithiasis and the ascendancy of overweight and obesity: a comparison of two patient series observed 25 years apart. *Nephrol Dial Transplant* 2013;28 Suppl 4:iv146-51. doi: 10.1093/ndt/gft076.
26. Cho ST, Jung SI, Myung SC, Kim TH. Correlation of metabolic syndrome with urinary stone composition. *Int J Urol* 2013 Feb;20(2):208-13. doi: 10.1111/j.1442-2042.2012.03131.x.
27. Daudon M, Lacour B, Jungers P. Influence of body size on urinary stone composition in men and women. *Urol Res* 2006;34(3):193-9. doi: 10.1007/s00240-006-0042-8].
28. Daudon M, Lacour B, Jungers P. Influence of body size on urinary stone composition in men and women. *Urol Res* 2006;34(3):193-9. doi: 10.1007/s00240-006-0042-8].

REFERENCES (3, 4, 10, 15-18)

3. Zaboлева most naseleniya Rossii v 2003 godu: Statisticheskie materialy. M.; 2004 g. P. 153-154. (In Russian)
4. Zaboлева most naseleniya v Rossii v 2016. (elektronnaya versiya). URL: <http://mednet.ru/ru/statistika/zaboлева-most-naseleniya.html>. (In Russian)
10. Rudenko V.I., Serova N.S., Kapanadze L.B. Vozmozhnosti dvuhenergeticheskoy kompyuternoy tomografii v diagnostike mochekamennoy bolezni. [Possibilities of dual energy computed tomography in the diagnosis of urolithiasis]. Materialy 3-y nauchno-prakticheskoy konferentsii urologov Severo-Zapadnogo federalnogo okruga RF 20–21 aprelya 2017 g. *Urologicheskie vedomosti* 2017;7(S): 92-93. (In Russian)
13. Al-Shukri SH, Slesarevskaya MN, Kuzmin IV. Litoliticheskaya terapiya uratnogo nefrolitiaz. [Litholytic therapy for urate nephrolithiasis]. *Urologiya* 2016;(2): 23-27. (In Russian)
15. Golovanov SA, Sivkov AV, Drozhzheva VV, Anokhin NV. Metabolicheskie faktory riska i formirovanie mochevyih kamney. Issledovanie I: vliyanie kaltsiyurii i uriku-rii. [Metabolic risk factors and urinary stone formation. The first study: role of calciuria and uricuria]. *Eksperimentalnaya i klinicheskaya urologiya* 2017;(1):52 – 57. (In Russian)
16. Golovanov SA, Sivkov AV, Drozhzheva VV, Anokhin NV. Metabolicheskie faktory riska i formirovanie mochevyih kamney. Issledovanie II: vliyanie fosfaturii i magniyurii. [Metabolic risk factors and formation of urinary stones. Study II: effects of phosphaturia and magnesium]. *Eksperimentalnaya i klinicheskaya urologiya* 2017;(2):42–48. (In Russian)
17. Golovanov SA, Sivkov AV, Polikarpova AM, Drozhzheva VV, Andryuhin MI, Prosyannikov MYu. Metabolicheskie faktory riska i formirovanie mochevyih kamney. Issledovanie III: vliyanie rN mochi. [Metabolic risk factors and formation of urinary stones. Study III: Effect of urine pH]. *Eksperimentalnaya i klinicheskaya urologiya* 2018;(1):84–91. (In Russian)
18. Popovkin NN, Grishkova NV, Chudnovskaya MV, Darenkov AF, Golovanov SA. Diferentsialnaya diagnostika sostava mochevogo kamnya in vivo po metabolicheskim pokazatelyam u bolnykh nefrolitiazom. [Differential diagnosis of the composition of the urinary stone in vivo according to metabolic parameters in patients with nephrolithiasis]. Aktualnye voprosy urologii i operativnoy nefrologii: Sbornik nauchnykh trudov. M., 1994. P. 32-42. (In Russian)