

УДК: 577.213:576.362:579.23

Логико-статистический анализ связи клинико-лабораторных показателей с возникновением нарушения мозгового кровообращения у пациентов пожилого возраста с хронической ишемией головного мозга

Кузнецова А.В.^{1*}, Костомарова И.В.^{2**}, Сенько О.В.^{3***}

¹Институт биохимической физики им. Н.М. Эммануэля, 117997, Москва, Косыгина, 4

²Филиал ГБОУ ВПО РНИМУ им. Н.И. Пирогова Минздрава "НКЦ геронтологии", 129226, Москва, Леонова, 16,

³Вычислительный центр им. А.А. Дородницына РАН, 119991, Москва, Вавилова, 40,

Аннотация. В статье представлены результаты анализа связи клинико-лабораторных показателей с развитием транзиторной ишемической атаки и ишемического инсульта у пациентов старших возрастных групп с хронической ишемией головного мозга. В исследованиях использован метод интеллектуального анализа данных, основанный на построении оптимальных разбиений признакового пространства, а также различные методы распознавания. В ходе исследований были выявлены основные предиктивные факторы риска развития транзиторной ишемической атаки: показатели липидного спектра, показатели клеточного состава крови, биохимические показатели, тип гиперлипидемии, II-III стадии хронической ишемии головного мозга, электрокардиографические признаки гипертрофии левого желудочка, признаки диффузных изменений щитовидной железы при ультразвуковом исследовании. Использование методов распознавания с учетом показателей, вошедших в закономерности базовой системы, позволяет достигнуть точности прогноза развития транзиторной ишемической атаки в 81,5%. У больных, перенесших транзиторную ишемическую атаку, факторами риска последующего развития ишемического инсульта являются: ишемическая болезнь сердца с мерцательной аритмией, нарушение ритма, гипергликемия. Прогностическая точность данных факторов составляет около 77%.

Ключевые слова: прогностические факторы риска, транзиторная ишемическая атака, ишемический инсульт, метаболизм липидов, интеллектуальный анализ данных, распознавание.

1. ВВЕДЕНИЕ

В последние годы в большинстве экономически развитых стран наблюдается увеличение доли лиц пожилого и старческого возраста. В данной ситуации на первый

*azfor@narod.ru

**erri06@rambler.ru

***senkoov@mail.ru

план выходит проблема улучшения качества жизни в пожилом возрасте и профилактика возрастной патологии [1].

Хроническая ишемия головного мозга (ХИГМ) превалирует среди сосудистых поражений головного мозга у пожилых. При прогрессировании ХИГМ возможно развитие транзиторных ишемических атак (ТИА) и ишемического инсульта (ИИ), как осложнений при декомпенсации заболевания.

ТИА является наиболее частой формой острой цереброваскулярной патологии у людей старше 60 лет и одним из независимых факторов риска ишемического инсульта (ИИ) и смерти [2]. Показано, что среди перенесших ишемический инсульт (ИИ) предшествующая ТИА встречалась у 7%–40% случаев [3]. Риск развития ТИА увеличивается с возрастом. Согласно данным научного отчета экспертов АНА/ASA «Определение и оценка ТИА», распространенность ТИА в когорте людей зрелого возраста 45–65 лет составляет 0,4%, у мужчин в возрастной группе 65–75 составляет 2,7%, а уже в возрасте 75–79 лет – 3,6%, у женщин увеличивается с 1,6% до 4,1% в соответствующих возрастных группах [3].

Основными причинами возникновения ТИА являются системный атеросклероз, атеросклероз церебральных сосудов, кардиопатии, артериальная гипертензия, инсулиннезависимый сахарный диабет. Транзиторная ишемическая атака является предвестником инсульта. В большинстве случаев, до 50%, она предшествует развитию ишемического инсульта по атеротромботическому типу, около 30% – по кардиоэмболическому типу.

ТИА более прогностически благоприятная патология, чем ИИ. Выявление факторов развития острого церебрального события после ТИА и их коррекция позволит снизить риск развития повторной цереброваскулярной катастрофы и, как следствие, инвалидности пациента или смерти.

ВОЗ определены четыре стратегии, которые могут уменьшить неблагоприятный прогноз инсульта.

В плане нашего исследования интересен аспект уменьшения среднего уровня факторов риска впервые возникшего или повторного инсульта [4].

Факторы риска редко действуют изолированно, чаще наблюдается их совместное воздействие, что значимо увеличивает риск возникновения нарушения мозгового кровообращения [5]. В последние годы становится очевидным, что традиционные факторы риска сердечно-сосудистой патологии, такие как артериальная гипертензия, сахарный диабет, курение и заболевания сердца, не обуславливают полностью риск инсульта.

Прогнозирование риска развития нарушений мозгового кровообращения у больных с ХИГМ старших возрастных групп может иметь некоторые особенности, учитывая изменения структуры факторов риска, наблюдающиеся с возрастом. У больных пожилого и старческого возраста основным этиологическим фактором развития ИИ является атеросклероз. Однако 50 % всех ИИ происходят у больных без гиперхолестеринемии, а 20% всех острых сосудистых событий происходит при отсутствии какого-либо из основных факторов сердечно-сосудистых рисков [6]. Поэтому одним из перспективных подходов к выявлению прогностических факторов риска развития ТИА и/или ишемического инсульта, у больных с ХИГМ старших возрастных групп, является комплексный анализ клинико-лабораторных факторов риска, в частности клеточного и биохимического профиля крови.

Цель настоящего исследования – изучение связи клинических, биохимических и инструментальных показателей с развитием нарушений мозгового кровообращения у пациентов с ХИГМ, включая оценку совместного влияния перенесенной ТИА и выявленных значимых клинико-лабораторных показателей на развитие ишемического инсульта.

В ходе исследований использовали логико-статистические методы интеллектуального анализа данных (DataMining) [7], оценивали эффективность новых модификаций метода анализа данных, основанных на построении оптимальных достоверных разбиений признакового пространства [8]. Логико-статистический подход позволяет проводить анализ, не делая априорных предположений о виде вероятностных распределений, а также такой подход эффективен в условиях малой выборки и большого количества плохо структурированных признаков.

2. ОБЪЕКТЫ И МЕТОДЫ ИССЛЕДОВАНИЯ

2.1. Объекты исследования

В исследование были включены 235 пациентов с хронической ишемией головного мозга, проходивших лечение в стационаре "НКЦ геронтологии". В выборку не вошли данные пациентов с декомпенсированной соматической патологией, в остром и раннем восстановительном периоде ишемического инсульта.

I группу составили 117 пациентов с ХИГМ I-III стадии (средний возраст 59,4); II группа – 53 пациента с ТИА в анамнезе (средний возраст 62,0); III группа – 24 пациента с перенесенными ТИА и ишемическим инсультом (средний возраст 64,1) и IV группа – 41 больной с перенесенным ишемическим инсультом более трех лет назад (средний возраст 63,4).

Выраженность ХИГМ определялась как стадия дисциркуляторной энцефалопатии (ДЭ). Диагностика ДЭ проводилась на основании клинических данных и результатов нейровизуализации в соответствии с рекомендациями Научного центра неврологии РАМН [9, 10]. В данном исследовании были проанализированы данные анамнеза, клинико-лабораторных показателей и результаты инструментальных методов исследования (КТ, МРТ, УЗДГ, ЭКГ, УЗИ), полученные в "НКЦ геронтологии".

2.2. Методы анализа данных

Для анализа данных использовался логико-статистический метод оптимально достоверных разбиений (ОДР), а также различные методы распознавания.

Метод интеллектуального анализа данных – ОД основан на построении разбиений областей допустимых значений объясняющих переменных, индуцирующих разбиение исходной выборки на группы объектов с максимальными различиями уровней прогнозируемой величины. Качество разбиений описывается с помощью специального интегрального функционала, описанного в работе [8].

Разбиения с максимальным значением функционала качества ищутся внутри нескольких семейств различного уровня сложности. В настоящем исследовании использовалось простейшее семейство всевозможных одномерных разбиений с одной граничной точкой, а также семейство всевозможных двумерных разбиений с одной граничной по каждой из двух объясняющих переменных. Иными словами, последнее семейство может быть описано как множество двумерных разбиений с двумя границами, каждая из которых параллельна только одной из координатных осей (примеры таких разбиений приведены, в частности, на рис. 1–3). Оптимальное разбиение считается закономерностью только после положительного результата статистической верификации.

Верификация производится с помощью перестановочных тестов, основанных на сравнении качества оптимальных разбиений на исходной анализируемой выборке с качеством разбиений на случайных выборках, полученных из исходной выборки с помощью случайных перестановок значений прогнозируемой величины относительно фиксированных позиций векторов прогностических переменных. Перестановочные тесты позволяют не только оценивать достоверность выявленных закономерностей, но

также исключать из множества выходных закономерностей разбиения с избыточной сложностью.

В настоящей работе использована новая модификация способа оценки достоверности более сложных закономерностей, подробно описанная в приложении 2. В отличие от ранее использовавшегося подхода, изложенного в [8], данная модификация не требует существования достоверных одномерных закономерностей для переменных, используемых при описании верифицируемой сложной закономерности R . Такие переменные далее будут называться релевантными R . Модификация включала также средства, позволяющие оценить достоверность различий, имеющих между закономерностями идентичных наборов показателей, существующими в двух разных группах. Используемый подход основан на сравнении средних значений прогнозируемой величины в двух группах внутри областей, являющихся элементами оптимального разбиения, построенного по одной из групп. Верификация различий производится путём сравнения значений функционала, описывающего суммарные различия по всему разбиению, на исходной выборке и на случайных выборках, полученных с помощью перестановок. Также модификация включала дополнительные средства, позволяющие выделить внутри найденного набора подмножество базовых закономерностей, описывающих основные эффекты.

Метод реализован в рамках программной системы «РАЗБИЕНИЕ».

Для прогноза наличия ТИА и ИИ у пациентов с ХИГМ использовали универсальную систему методов интеллектуального анализа данных «РАСПОЗНАВАНИЕ», включающую коллекцию методов распознавания и анализа данных, основанных на различных хорошо зарекомендовавших себя подходах, включая метод «Логические закономерности», основанный на голосовании по системам логических закономерностей [11,12]. Также был использован метод распознавания «Мультимодельные статистически взвешенные синдромы» [13]. Данный метод основан на принятии коллективных решений по областям оптимальных разбиений, найденных с помощью метода ОДР. Прогностическая способность оценивалась с помощью метода скользящего контроля, дающего объективную несмещённую оценку точности. Под прогностической способностью в данном случае понималось правильное отнесение пациента к группам с уже произошедшим ТИА или без ТИА (с уже произошедшим ИИ или без ИИ) из анализируемой выборки по совокупности других клинических и биохимических показателей.

Изложение полученных результатов производится в следующей последовательности:

1) В разделе 3.1. описаны исследования закономерностей связи риска развития ТИА с набором клинико-лабораторных факторов с использованием одномерных и двумерных моделей ОДР. С помощью методов распознавания изучена прогностическая способность различных групп факторов, включая группу, состоящую только из лабораторных показателей и возраста, а также группу, содержащую все показатели, релевантные базовым закономерностям (см. приложение 2). Показано нарастание прогностической способности по мере расширения групп показателей.

2) В разделе 3.2. проведено прогнозирование развития транзиторной ишемической атаки у пациентов с ХИГМ. С помощью методов распознавания изучена прогностическая способность различных групп всех факторов, релевантных базовым закономерностям. Исследования с помощью ОДР показали, что риск развития повторного ИИ не связан непосредственно с перенесенной ТИА. Однако анализ двумерных закономерностей показал значительный вклад показателя «ТИА в анамнезе». Анализ закономерностей указывает на значительные различия в характере связи ИИ с большой группой показателей у пациентов с перенесенной ТИА и без ТИА.

3) В связи с данным обстоятельством в разделах 3.3. и 3.4. было проведено исследование связи риска развития ИИ с набором клиничко-лабораторных факторов в группах без ТИА и с ТИА.

4) В разделе 3.5. представлено исследование статистической достоверности расхождений между закономерностями в группах без ТИА и с ТИА, проведённое с помощью специальной технологии, изложенной в приложении 2. Исследования подтвердили существенные различия по факторам риска развития первичного ИИ и ИИ после перенесенной ТИА у пациентов с ХИГМ.

3. РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

3.1. Выявление связи клиничко-лабораторных показателей с развитием ТИА у пациентов с ХИГМ

На первом этапе были проведены исследования по выявлению факторов риска развития транзиторной ишемической атаки у пациентов с ХИГМ. С использованием метода ОДР для этой группы были выявлены закономерности, связывающие развитие ТИА с клиническими, биохимическими и инструментальными показателями.

Во всех таблицах далее используются обозначения с пояснениями в приложении 3.

3.1.1. Одномерные закономерности

Закономерности, выявленные в рамках простейшей одномерной модели с одной граничной точкой [8], с достоверностью не хуже 0,01 представлены в таблице 1. Для всех показателей, принимающих вещественные числовые значения, граница ставилась автоматически по базе данных. Однако для большого числа таких показателей существует граница, выставляемая экспертом на основании общепринятых референтных значений или данных научных источников. При этом граница, выставленная экспертом, чаще всего не совпадает с границей, рассчитанной по методу ОДР, вследствие особенностей конкретной анализируемой базы данных. Поэтому для таких показателей формировались две переменных: одна – непрерывная, другая – бинарная (в названии помечено буквой «ч», см. приложение 3). Информативность и статистическая достоверность связи с прогнозируемой величиной (ТИА и/или ИИ) в обоих случаях оценивалась с помощью метода ОДР. Бинарные и категориальные переменные в рассматриваемой базе данных задавались с помощью числовых функций. Поэтому границы, разделяющие задаваемые ими группы, также являются числовыми. Для категориального показателя стадия ХИГМ граница 1.5 формирует две группы: группу с первой стадией и группу со второй и третьей стадиями. Для категориального показателя ГЛП граница 3.5 отделяет тип Пв от других типов или случаев отсутствия ГЛП. В правом крайнем столбце дана оценка р-значения по методу ОДР [8], которое по смыслу совпадает с р-значениями, рассчитываемыми в стандартных статистических тестах.

H-коэффициент показывает степень достоверности закономерности. Он указывает, во сколько раз значение функционала, описывающего качество разбиения на исходной выборке, превышает максимальное значение функционала качества разбиений на случайных выборках. *H*-коэффициенты позволяют сравнить достоверность для двух закономерностей, в которых значение функционала качества на случайных выборках ни разу не превысило значение на реальных данных. Показатель «вес» позволяет оценить вклад признака в некоторый набор закономерностей, и, по сути, является суммой *H*-коэффициентов, оценивающих вклад данного признака в те закономерности из набора, которым он релевантен (см. приложение 2).

Таблица 1. Основные клинико-лабораторные и инструментальные показатели, связанные с развитием ТИА у пациентов с ХИГМ (выявлены при одномерных разбиениях методом ОДР) (Показатели расположены по убыванию функционала качества *H*-информативности)

Показатели	Ниже границы (число случаев, %)		Граница	Выше границы (число случаев, %)		<i>H</i> - коэффициент	<i>p</i>
	ХИГМ	ХИГМ, ТИА		ХИГМ	ХИГМ, ТИА		
1. ХС-ЛПВП (меньше/больше границы, найденной ОДР)	10 31%	22 69%	0,55	84 84%	16 16%	1,987544	< 0,0005
2. ХС-ЛПВП-ч (меньше/больше границы, выставленной экспертом)	33 51%	32 49%	1,0	61 90%	7 10%	1,694723	< 0,0005
3. ХИГМ-стадии (1 – I-я, 2 – II-я, 3 – III-я)	42 93%	3 7%	1,5 (I-II ст.)	75 60%	50 40%	1,511683	< 0,0005
4. ХС – ЛПОНП (меньше/больше границы, найденной ОДР)	87 81%	22 19%	0,84	7 30%	16 70%	1,327511	< 0,0005
5. ЭКГ-глж (нет -0/есть -1)	37 95%	2 5%	0,5	80 62%	50 38%	1,234568	< 0,0005
6. ЭКГ-ИМ (нет/есть)	109 73%	40 27%	0,5	8 32%	13 68%	1,000000	0,001
7. Фибриноген (меньше/больше границы, найденной ОДР)	84 79%	22 21%	4,6	33 52%	31 48%	0,875286	0,0005
8. ТГ (меньше/больше границы, найденной ОДР)	91 78%	25 22%	2,0	8 8%	13 92%	0,770320	0,002
9. ГЛП (нет/есть 1 – нет ГЛП, 2 – тип I, 3 – тип IIa, 4 – тип IIb)	79 87%	20 13%	3,5 (IIb тип)	20 32%	18 68%	0,763010	0,005
10. ХС-ЛПНП (меньше/больше границы, найденной ОДР)	75 82%	17 18%	6,15	19 48%	21 52%	0,717993	0,004
11. ХС-ЛПОНП- ч (меньше/больше границы, выставленной экспертом)	66 80%	16 20%	0,63	28 56%	22 44%	0,626944	0,0035
12. Прием Вазилипа (0 – нет, 1 – да).	100 74%	35 26%	0,5	16 47%	18 53%	0,617490	0,006
13. АСТ (меньше/больше границы, найденной ОДР)	38 90%	4 10%	20,5	79 62%	49 38%	0,556530	0,006

14. Фибриноген-ч (меньше/больше границы, выставленной экспертом)	48 83%	10 17%	4,0	69 62%	43 38%	0,548046	0,005
15. ОХС-ч (меньше/больше границы, выставленной экспертом)	26 90%	3 10%	5,0	91 65%	50 35%	0,454713	0,0095
16. ТГ-ч (меньше/больше границы, выставленной экспертом)	74 64%	19 36%	1,65	25 57%	19 43%	0,396244	0,008

Анализ данной таблицы показал, что у пациентов с ХИГМ наиболее значимым фактором риска развития ТИА является низкий уровень ХС-ЛПВП ($< 0,55$ ммоль/л) ($H = 1,98$, $p < 0,0005$). Также значительно выражена связь риска развития ТИА со стадиями ХИГМ ($H = 1,51$, $p < 0,0005$) и ЭКГ-признаками гипертрофии левого желудочка ($H = 1,23$, $p < 0,0005$). Так, у больных II-III-й стадиями ХИГМ риск развития ТИА повышен в 6 раз (OR = 6,0 ДИ95% [1,97–18,27]), а при наличии гипертрофии левого желудочка – в 7,5 раз (OR = 7,5 ДИ95% [1,91–29,45]).

3.1.2. Двумерные закономерности и базовая система закономерностей

Использование модели двумерных разбиений с одной граничной точкой по каждой из релевантных переменных (модель III, см.[8]) выявило 55 достоверных двумерных закономерностей. При этом в случаях, когда по крайней мере для одной из релевантных переменных существовала простейшая одномерная закономерность с уровнем достоверности не хуже 0,01, отбор двумерных закономерностей производился согласно условию $p_1 < 0,01$, $p_2 < 0,01$. В случаях, когда значимые закономерности для релевантных переменных отсутствовали, отбор двумерных закономерностей производился согласно условию $p_1 < 0,02$, $p_2 < 0,02$, (см. приложение 2).

Значение p_1 характеризует значимость вхождения в двумерную закономерность переменной X. Соответственно, через p_2 обозначено p -значение, характеризующее значимость вхождения в двумерную закономерность переменной Y.

С помощью методики, описанной в приложении 2, из полученных наборов одномерных и двумерных закономерностей была выделена базовая система из 35 двумерных и одной одномерной закономерности. Найденные двумерные закономерности, вошедшие в базовую систему, представлены в приложении 1 (табл. 1).

Некоторые из двумерных закономерностей представлены на рис. 1–3. На диаграммах в каждом квадранте указаны соотношения числа случаев, обозначенных «+», к числу случаев, обозначенных «о».

Показано, что среди пациентов, перенесших ТИА, преобладают больные с низким уровнем ХС-ЛПВП (менее 0,45 ммоль/л) или высоким уровнем фибриногена (выше 4,87 г/л) (рис. 1). Подобные результаты, о влиянии низких уровней ХС-ЛПВП и высоких уровней фибриногена на риск развития нарушений мозгового кровообращения, были показаны в исследовании Rizzo M. с соавторами [14].

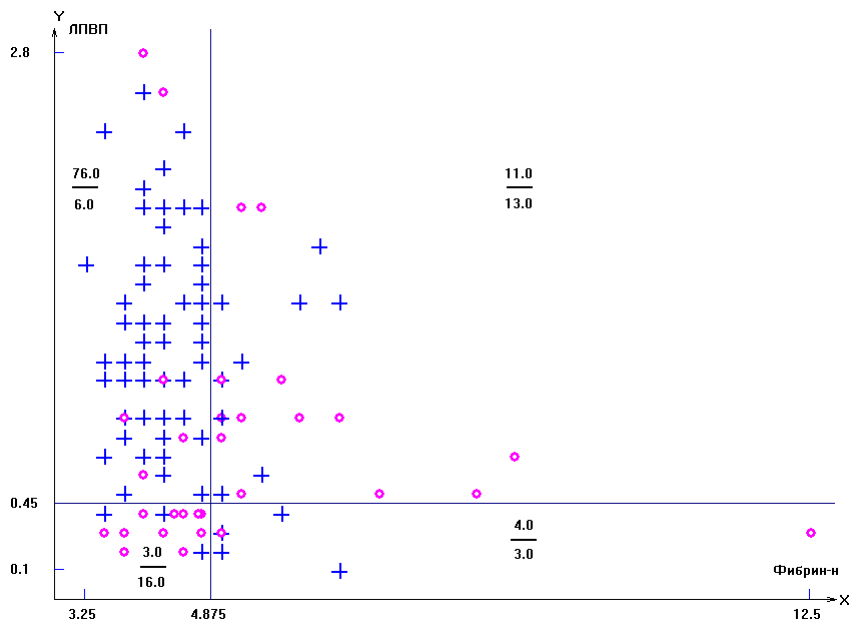


Рис. 1. Диаграмма рассеяния, отражающая связь уровня фибриногена (г/л) – ось X и уровня ХС-ЛПВП (ммоль/л) – ось Y с наличием («o») или отсутствием («+») ТИА у пациентов с ХИГМ, $p1 < 0,0005$, $p2 = 0,0005$.

Низкий уровень ХС-ЛПВП может усугублять атеросклеротический процесс через влияние на процесс воспаления. Низкие уровни ХС-ЛПВП, по-видимому, способствуют переводу процесса воспаления в острую фазу. Воспаление, как известно, играет решающую роль в эволюции от эндотелиальной дисфункции до разрыва атеросклеротической бляшки и атеротромбоза. Фибриноген, в свою очередь, также оказывает влияние на прогрессирование атеросклероза через участие в раннем формировании и росте атероматозной бляшки и способствует усилению воспаления. Обнаружена достоверная обратная корреляция между уровнями ХС-ЛПВП и концентрацией фибриногена [15].

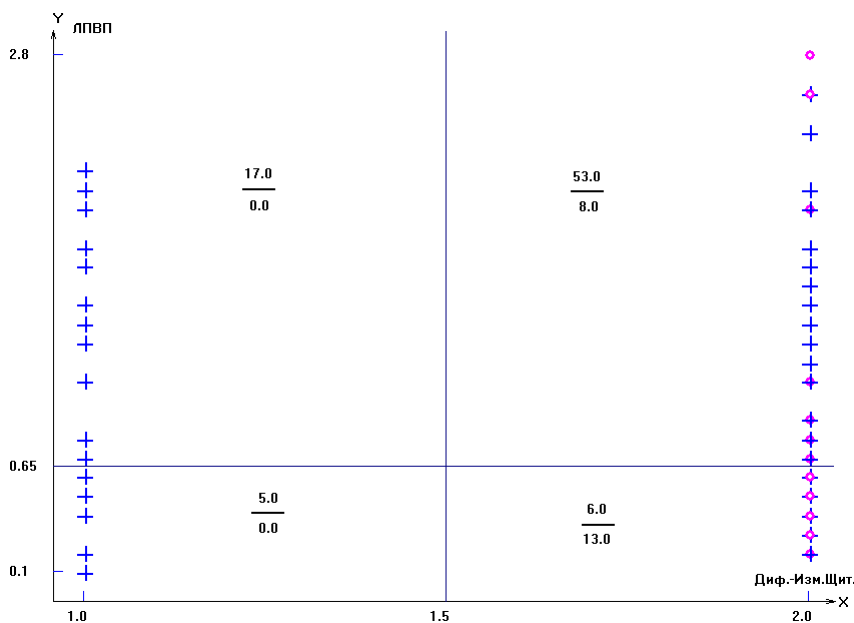


Рис 2. Диаграмма рассеяния, отражающая связь диффузных изменений щитовидной железы (1 – нет, 2 – есть) – ось X – и уровня ХС-ЛПВП (ммоль/л) – ось Y с наличием («o») или отсутствием («+») ТИА у пациентов с ХИГМ, $p1 = 0,0095$, $p2 < 0,0005$.

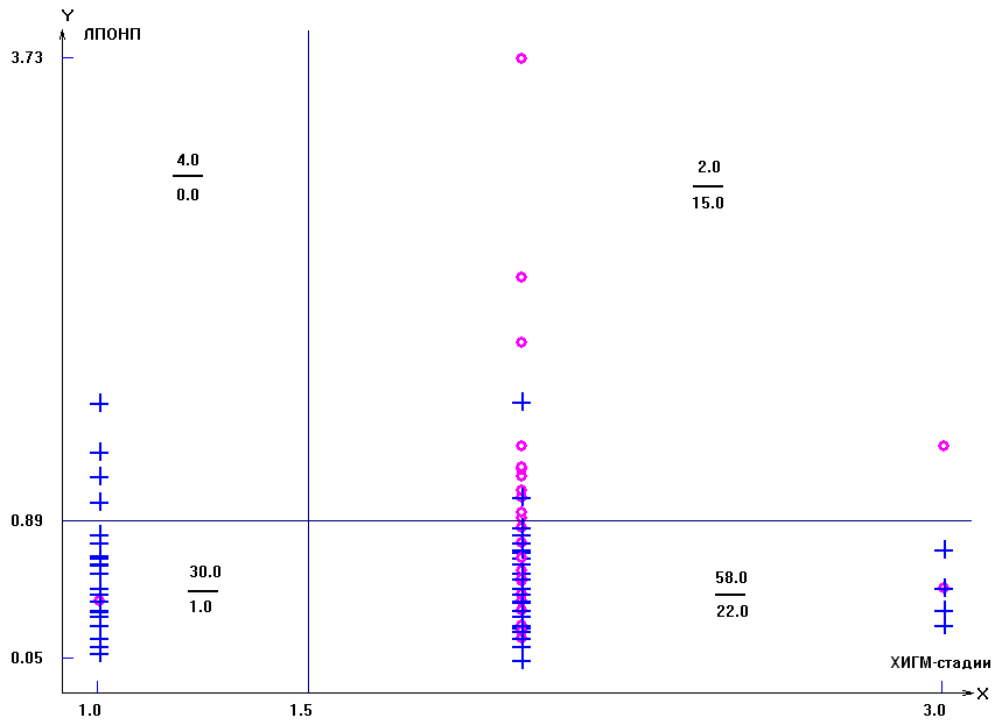


Рис 3. Диаграмма рассеяния, отражающая связь стадии заболевания – ось X – и уровня ХС-ЛПОНП (ммоль/л) – ось Y – с наличием («o») или отсутствием («+») ТИА у пациентов с ХИГМ, $p_1 = 0,009$, $p_2 < 0,0005$.

Значимым фактором для развития ТИА является сочетанное наличие у пациента низкого уровня ХС-ЛПВП (менее 0,65 ммоль/л) и диффузных изменений щитовидной железы (по данным ультразвукового исследования) (рис. 2). Это, по-видимому, связано как с влиянием на риск низкого уровня ХС-ЛПВП, так и влиянием вероятного «гипотиреоидного статуса» пациента, на УЗИ проявляющегося в виде признаков диффузных изменений щитовидной железы. Связи риска развития ТИА с уровнями тиреоидных гормонов выявлено не было, что можно объяснить наличием медикаментозной ремиссии у пациентов.

Также, у больных с развернутой клинической картиной ХИГМ повышенный риск развития ТИА связан с уровнями ХС-ЛПОНП. У пациентов ХИГМ II-III стадий с уровнями ХС-ЛПОНП выше 0,89 ммоль/л в 3 раза выше риск развития ТИА (OR = 3,2 ДИ95% [2,16–4,77]) по сравнению с пациентами с более низкими уровнями ХС-ЛПОНП (рис. 3).

Далее из полученных показателей, по одномерным и двумерным разбиениям, была выделена базовая система для прогнозирования.

В таблице 2 представлен вклад различных показателей базовой системы, связанных с развитием ТИА, вместе с границами оптимальных разбиений. Здесь и далее в случае, если для признака существует достоверное одномерное разбиение, в таблице приводится его граница. В противном случае приводится чаще встречающаяся граница в двумерной закономерности. Области повышенного риска ТИА, задаваемые границами, могут быть прослежены по таблице 1 и таблице 1 из приложения 1.

Таким образом, наши данные показывают, что основным параметром, влияющим на риск развития ТИА, является уровень ХС-ЛПВП. Отметим, что весовой вклад бинарного показателя ХС-ЛПВП-ч с границей, выставленной экспертом, оказывается больше вклада показателя ХС-ЛПВП с границей, найденной с помощью ОДР. Высокая величина вклада ХС-ЛПВП-ч связана с высоким уровнем разделения групп с помощью экспертной границы на анализируемой выборке, который близок к максимальному уровню разделения с помощью ОДР границы. Однако статистическая значимость

границы ХС-ЛПВП-ч оказывается выше значимости границы, найденной с помощью ОДР, из-за того, что используемая процедура оценки значимости учитывает эффект подгонки на анализируемых данных. Вклад (вес) показателя имеет смысл суммы H -коэффициентов, оценивающих статистическую значимость.

Таблица 2. Основные показатели, вошедшие в закономерности базовой системы, для прогнозирования развития ТИА у больных с ХИГМ

№	Показатели	Вес	Граница
1.	ХС-ЛПВП-ч	15,23061	1,5
2.	ХИГМ-стадии	11,21818	1,5
3.	ОХС	4,546871	5,25 ммоль/л
4.	ХС-ЛПВП	4,448803	0,55 ммоль/л
5.	ЭКГ-глж	4,150513	1,5
6.	Диф.Изм.Щит.	3,393896	1,5
7.	Тип ГЛП	2,710497	3,5
8.	ХС-ЛПОНП-ч	2,493449	1,5
9.	Извитость сосудов (по данным УЗДГ)	2,481803	1,5
10.	ХС-ЛПОНП	2,157131	0,84 ммоль/л
11.	Фибриноген	2,052072	4,6 г/л
12.	ХС-ЛПНП	2,03552	6,15 ммоль/л
13.	ЭКГ- инфаркт миокарда	1,78517	1,5
14.	Триглицериды	1,695146	2,0 ммоль/л
15.	Фибриноген - ч	1,647036	1,5
16.	АСТ	1,030622	20,5 ед./л
17.	ОХС-ч	0,946256	1,5
18.	ИБС-атер.	0,8968	1,5
19.	Признаки лейкоареоза (по данным МРТ)	0,889046	1,5
20.	Наследственность	0,851438	1,5
21.	Триглицериды - ч	0,840753	1,5
22.	Группы по возрасту	0,826005	1,5
23.	Лимфоциты	0,773862	25,5 %
24.	СОЭ	0,741266	12,7 мм/час
25.	АЛТ-ч	0,648316	1,5

Подобные результаты получены в исследовании SPARCL. У пациентов, недавно перенесших ТИА или ишемический инсульт, показано предиктивное значение низкого уровня ХС-ЛПВП. Показано, что с каждым повышением ХС-ЛПВП на 0,35 ммоль/л риск развития инсульта снижается на 13% [16].

3.2. Прогнозирование развития транзиторной ишемической атаки у пациентов с ХИГМ с помощью методов распознавания.

Были проведены эксперименты по оценке точности прогноза развития ТИА по трём группам переменных, входящих в базовые закономерности и представленных в таблице 2. Исследовалась прогностическая эффективность следующих методов: линейного дискриминанта Фишера, линейной машины, Q -ближайших соседей, метода опорных векторов, мультимодельных статистических взвешенных синдромов, метода, основанного на голосовании по наборам логических закономерностей (“Логические закономерности”) [11, 12]. Точность прогноза оценивалась с помощью метода скользящего контроля. Целью экспериментов была оценка прогностической ценности добавляемых групп показателей. В данном случае прогноз понимается только лишь как отнесение пациента в группу с ТИА или без ТИА по совокупности показателей.

Для первого эксперимента из всех показателей, вошедших в базовую систему, были взяты все лабораторные показатели, а также возраст и тип гиперлипидемии.

Во втором эксперименте были добавлены показатели, характеризующие состояние сердечно-сосудистой системы и щитовидной железы, также вошедшие в базовую систему.

В третьем эксперименте использовались все показатели, вошедшие в базовые закономерности, связанные с развитием ТИА у больных с ХИГМ. Показатели, используемые в первом эксперименте, представлены в таблице 3. Область значений повышенного риска ТИА для показателей не указана, т.к. для показателей, полученных при двумерных разбиениях, данные области могут варьировать и отличаться в зависимости от пары признака.

Таблица 3. Показатели, используемые в первом эксперименте

<i>№</i>	<i>Показатели</i>	<i>№</i>	<i>Показатели</i>
1.	ХС-ЛПВП-ч	10.	Фибриноген-ч
2.	ОХС	11.	АСТ
3.	ХС-ЛПВП	12.	ТГ-ч
4.	Тип гиперлипидемии	13.	Группы по возрастам
5.	ХС-ЛПОНП-ч	14.	ХС-ч
6.	ХС-ЛПОНП	15.	Лимфоциты
7.	Фибриноген	16.	СОЭ
8.	ХС-ЛПНП	17.	АЛТ-ч
9.	Триглицериды		

Результаты по первому эксперименту представлены в таблице 4. Наряду со стандартными показателями – чувствительность и специфичность, – в таблице также приводится средняя точность прогноза в группе пациентов, перенесших ТИА, и без ТИА. Средняя точность связана со специфичностью и чувствительностью соотношением:

$$\text{Средняя точность} = (1 - \text{Специфичность} + \text{Чувствительность}) / 2.$$

Средняя точность позволяет наглядно сравнивать эффективность различных методов распознавания.

Таблица 4. Прогнозирование развития ТИА с помощью различных методов машинного обучения (по переменным первого эксперимента)

<i>Метод</i>	<i>Точность в группе с ТИА - чувствительность прогноза</i>	<i>Специфичность</i>	<i>Средняя точность</i>
МСВС	71,7%	20,5%	75,6%
Линейная машина	56,6%	18,8%	68,9%
Линейный дискриминант Фишера	58,5%	12,0%	73,3%
Метод опорных векторов	64,2%	12,8%	75,7%
Q-ближайших соседей	81,1%	28,2%	76,4%
Логические закономерности	73,6%	19,7%	77,0%
Среднее по всем методам	67,6%	18,7%	74,5%

Среднее по всем методам значение чувствительности составило 67,6%, среднее значение специфичности – 18,7%. Наивысшую среднюю точность (77%) показал метод поиска закономерностей. Близкую точность показал метод Q-ближайших соседей (76,4%), а также метод опорных векторов (75,7%) и МСВС (75,6%).

Отметим, что метод МСВС основан на голосовании по системам закономерностей, вычисляемых с помощью оптимальных разбиений.

Во втором эксперименте (табл.5) к прогностическим переменным, используемым в первом эксперименте (табл. 3.), были добавлены 4 параметра из таблицы 2: ИБС-атеросклероз коронарных артерий (ИБС-атер), ЭКГ признаки гипертрофии левого желудочка и перенесенного инфаркта миокарда (ЭКГ.глж., ЭКГ-им), наличие признаков диффузных изменений щитовидной железы при УЗИ (Диф.Изм.Щит.).

Таблица 5. Основные показатели, используемые во втором эксперименте

<i>№</i>	<i>Показатели</i>	<i>№</i>	<i>Показатели</i>
1.	ХС-ЛПВП-ч	12.	Фибриноген-ч
2.	ОХС	13.	АСТ
3.	ХС-ЛПВП	14.	ТГ-ч
4.	Тип гиперлипидемии	15.	Группы по возрастам
5.	ХС-ЛПОНП-ч	16.	ХС-ч
6.	ХС-ЛПОНП	17.	Лимфоциты
7.	Фибриноген	18.	СОЭ
8.	ХС-ЛПНП	19.	ИБС-атер *
9.	Триглицериды	20.	ЭКГ- им *
10.	Диф.Изм.Щит.*	21.	АЛТ-ч
11.	ЭКГ.глж *		

* Добавленные показатели из табл. 2.

Таблица 6. Прогнозирование развития ТИА с помощью различных методов машинного обучения (по переменным второго эксперимента)

<i>Метод</i>	<i>Точность в группе с ТИА – чувствительность прогноза</i>	<i>Специфичность</i>	<i>Средняя точность</i>
МСВС	73,6%	16,2%	78,8%
Линейная машина	67,9%	20,5%	73,7%
Линейный дискриминант Фишера	62,3%	17,9	72,2%
Метод опорных векторов	67,9%	14,5%	76,7%
Q - ближайших соседей	84,9%	23,9%	78,8%
Логические закономерности	77,4	20,5	78,5%
Среднее по всем методам	72,3%	18,9%	76,5%

По таблице 6 можно увидеть заметное улучшение чувствительности прогноза по сравнению с первым экспериментом. Среднее по всем методам значение чувствительности возросло с 67,6% до 72,3%, при сохранении среднего значения специфичности – 18,9%. Наивысшую среднюю точность (78,8%) показали метод МСВС и Q - ближайших соседей. Близкую точность продемонстрировали метод «Логические закономерности» (78,5%) и метод опорных векторов (76,7%).

Использование, в третьем эксперименте (табл.7), всех показателей, вошедших в закономерности базовой системы, привело к небольшому повышению точности прогноза. Наивысшая средняя точность прогноза была достигнута с помощью метода опорных векторов (82%). Близкое значение продемонстрировали также метод МСВС (81,5%) и «Логические закономерности» (80,1%) .

Таблица 7. Прогнозирование развития ТИА с помощью различных методов машинного обучения (по всем переменным, релевантным базовым закономерностям – третий эксперимент)

<i>Метод</i>	<i>Точность в группе с ТИА – чувствительность прогноза</i>	<i>Специфичность</i>	<i>Средняя точность</i>
Линейная машина	71,7%	15,4%	78,2%
Q-ближайших соседей	71,7%	21,4%	75,2%
Линейный дискриминант Фишера	69,8%	12,8%	78,5%
МСВС	77,4%	14,5%	81,5%
Метод опорных векторов	73,6%	9,5%	82,0%
Логические закономерности	75,5%	15,4%	80,1%
Среднее по всем методам	73,2%	14,8%	79,3%

3.3. Выявление связи клинико-лабораторных показателей с развитием острого ишемического инсульта у пациентов с ХИГМ (по всей исследуемой выборке)

Далее нами были проведены исследования по выявлению факторов риска развития ишемического инсульта у пациентов с ХИГМ по всей изучаемой выборке. Это было проведено для оценки значения перенесенной ТИА, как одного из основных независимых факторов риска повторного ишемического инсульта и смерти. С использованием метода ОДР были выявлены клинические, биохимические и инструментальные показатели риска развития ИИ.

3.3.1. Одномерные закономерности

Результаты исследования представлены в таблице 8. Таблица содержит закономерности с достоверностью не хуже 0,01. В таблице используются обозначения, аналогичные обозначениям, используемым в таблице 1.

При анализе показателей, связанных с развитием ИИ, выявленных при одномерных разбиениях методом ОДР, самой значимой ($H = 3,0$) является III стадия ХИГМ, вероятно, как наиболее клинически выраженная и прогностически неблагоприятная.

Выраженной является связь крайне низкого уровня ХС-ЛПВП – менее 0,35 ммоль/л ($H = 0,81$) с развитием ИИ. Этот уровень ХС-ЛПВП является даже более низким, чем уровень ХС-ЛПВП (0,55 ммоль/л), связанный с риском развития ТИА. Возможно, уровень ХС-ЛПВП обратно коррелирует с тяжестью нарушения мозгового кровообращения. Выявлена связь высокого уровня систолического давления (более 197 мм рт. ст.) с развитием ишемического инсульта. Подобные результаты были получены Simons L.A. с соавторами, показавшими, что предикторами развития ИИ у пожилых, в возрастной группе 60–69 лет, являются сниженный уровень ХС-ЛПВП и артериальная гипертензия [17].

Следует отметить отсутствие значимой, даже на уровне 0,1, прямой взаимосвязи ТИА с развитием ИИ.

Таблица 8. Основные клинико-лабораторные и инструментальные показатели, связанные с развитием ишемического инсульта у пациентов с ХИГМ (выявлены при одномерных разбиениях методом ОДР) (Показатели расположены по убыванию коэффициента качества H - информативности)

Показатели	Ниже границы (число случаев, %)		Граница	Выше границы (число случаев, %)		H- коэфф.	P
	ХИГМ	ХИГМ, ИИ		ХИГМ	ХИГМ, ИИ		
1. ХИГМ-стадии (1 – I-я, 2 – II-я, 3 – III-я)	163 85%	29 15%	2,5	7 16%	36 84%	3,0	<0,0005
2. Давление сист. (меньше/больше границы, найденной ОДР)	139 80%	35 20%	197,5	31 51%	30 49%	1,02	<0,0005 1
3. ХС-ЛПВП (меньше/больше границы, найденной ОДР)	15 45%	18 55%	0,35	117 65%	32 35%	0,81	0,0005
4. P _г -аорты (атеросклероза нет – 0/ есть –1)	32 89%	4 11%	0,5	76 65%	41 35%	0,75	0,0045
5. ИБС-им в анамнезе (нет –1/есть –2)	162 75%	54 25%	1,5	8 40%	12 60%	0,75	0,0010
6. СД-2-типа (нет – 1/есть – 2)	161 75%	54 25%	1,5	9 45%	11 55%	0,71	0,0085
7. УЗИ-щит.уз. (нет – 1/есть – 2)	49 91%	5 9%	1,5	66 71%	27 29%	0,67	0,0085
8. Давление диаст. (меньше/больше границы, найденной ОДР)	152 76%	47 24%	115,0	18 50%	18 50%	0,66	0,0085
9. ЭКГ – нп (нет – 1/есть – 2)	112 79%	30 21%	1,5	58 62%	35 38%	0,46	0,0060

3.3.2. Двумерные закономерности и базовая система закономерностей

С использованием модели двумерных разбиений с одной граничной точкой по каждой из релевантных переменных было выявлено 45 двумерных закономерностей с уровнем достоверности, удовлетворяющим условию $p_1 < 0,01$, $p_2 < 0,01$. При этом отбор двумерных закономерностей производился способом, описанном в пункте 3.1.2 (см. приложение 2).

Найденные двумерные закономерности из базовой системы представлены в приложении 1 (табл. 2).

Некоторые из двумерных закономерностей представлены на рисунках 4 и 5.

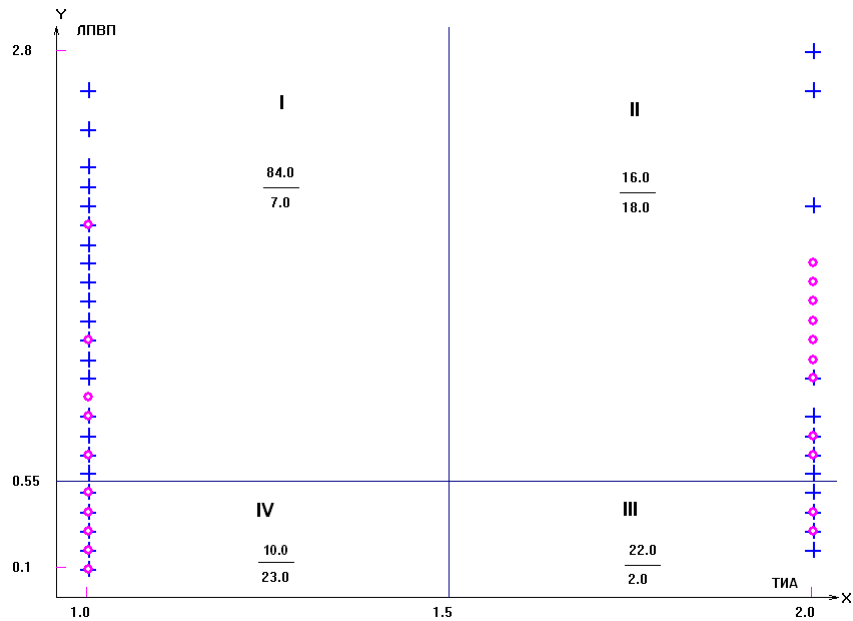


Рис. 4. Диаграмма рассеяния, отражающая связь перенесенной ТИА (1 – нет, 2 – есть) – ось X – и уровня ХС-ЛПВП (ммоль/л) – ось Y – с наличием « o » или отсутствием « $+$ » повторного ИИ, $p_1 < 0,0005$, $p_2 < 0,0005$. В каждом квадранте указаны соотношения числа случаев, обозначенных « $+$ », к числу случаев, обозначенных « o ».

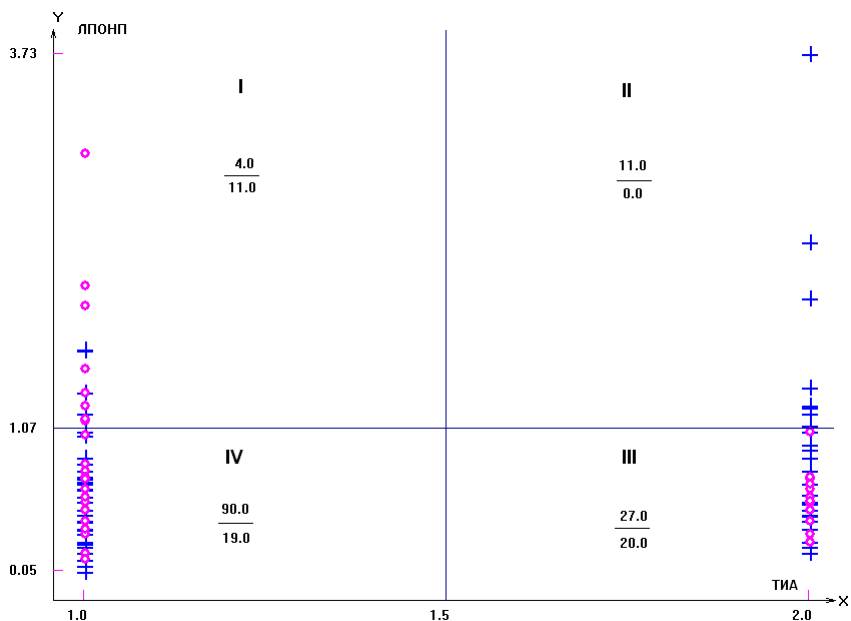


Рис. 5. Диаграмма рассеяния, отражающая связь перенесенной ТИА (1 – нет, 2 – есть) – ось X и уровня ХС-ЛПОНП, (ммоль/л) – ось Y с наличием « o » или отсутствием « $+$ » повторного ИИ, $p_1 < 0,0005$, $p_2 < 0,0005$. В каждом квадранте указаны соотношения числа случаев, обозначенных « $+$ », к числу случаев, обозначенных « o ».

При анализе рисунка 4 видно, что значения низкого уровня ХС-ЛПВП ($< 0,55$ ммоль/л) являются значимыми для риска развития ИИ на этапе заболевания до развития ТИА (I и IV квадранты диаграммы). При наличии у пациента перенесенной ТИА низкий уровень ХС-ЛПВП уже не влияет на риск развития инсульта (II и III квадранты диаграммы).

При оценке связи развития ИИ с уровнем ХС-ЛПОНП (рис. 5) выявляется та же закономерность. Основным фактором риска – является перенесённая ТИА. Уровень ХС-ЛПОНП не является значимым. Даже при нормальных значениях уровня ХС-ЛПОНП, но при наличии ТИА в анамнезе, у больных достоверно чаще отмечено

развитие ИИ. Таким образом, уровни ХС-ЛПВП и ХС-ЛПОНП не связаны с риском развития повторного ИИ у больных, перенесших ТИА.

Далее, из множества всех достоверных одномерных и двумерных закономерностей была выделена базовая система, содержащая закономерности, над которыми не доминируют никакие другие закономерности (подробно описано в приложении 2). Полностью данная система представлена в приложении 1.

Вклад наиболее значимых переменных вместе с границами разбиений представлен в таблице 9. Области повышенного риска ИИ, выделяемые границами, могут быть прослежены по таблице 8 и таблице 2 приложения 1.

Таблица 9. Основные показатели, вошедшие в закономерности базовой системы, для прогнозирования развития ишемического инсульта по всей исследуемой выборке

<i>№</i>	<i>Показатели</i>	<i>Вес</i>	<i>Границы</i>
1.	ХИГМ-стадии	12,04626	2,5
2.	ТИА	7,996631	1,5
3.	ХС-ЛПВП	3,882096	0,35 ммоль/л
4.	ХС-ЛПВП-ч	3,484922	1,5
5.	Св.х	2,197244	71,7 мг/дл
6.	Давление сист.	2,137515	197,5 мм рт.ст.
7.	Степень ожирения	1,85386	2,5
8.	УЗИ-щит.уз.	1,829759	1,5
9.	Лимфоциты	1,774802	23,5%
10.	Рг-аорты (признаки атеросклероза аорты)	1,671126	0,5
11.	Св.х-ч	1,573741	1,5
12.	ХС-ЛПОНП	1,229221	1,07 ммоль/л
13.	Тип ГЛП	1,120431	3,5
14.	Бр	0,988502	6,75 мкмоль/л
15.	Рост	0,815072	178 см
16.	УЗИ-п/ж	0,80393	1,5
17.	МС	0,787919	1,5
18.	Диф.-Изм.Щит.	0,762834	1,5
19.	Узловой зоб	0,685614	1,5
20.	Давление диаст.	0,519392	115 мм рт.ст.

Из таблицы 9 видно, что вклад перенесенной ТИА в развитие ИИ весьма значителен (вес равен 7,99). Однако отсутствие прямой связи между перенесенной ТИА и последующим развитием ИИ, вероятно, указывает на то, что факторы риска развития ИИ в группах пациентов, перенесших ТИА и без таковой, различаются.

Следующим этапом нашей работы явилось выявление различий между закономерностями, характеризующими взаимосвязь риска ишемического инсульта с различными клинико-лабораторными показателями в группах больных с перенесенной ТИА и без таковой. Был применен метод ОДР, а также методы распознавания, были применены для поиска закономерностей в каждой из групп. На последующих этапах была проведена оценка статистической достоверности различий между найденными закономерностями в двух группах с помощью новой разработанной технологии, основанной на использовании специального варианта перестановочного теста (см. приложение 2).

3.4. Определение клинико-лабораторных показателей, связанных с риском развития первичного ишемического инсульта у больных с ХИГМ (без ТИА)

Метод ОДР был использован для поиска закономерностей, характеризующих взаимосвязь риска ишемического инсульта с различными клинико-лабораторными показателями в группе больных без ТИА в анамнезе.

3.4.1. Одномерные закономерности в группе больных с хронической ишемией головного мозга без случаев транзиторных ишемических атак в анамнезе

Использование одномерной модели ОДР позволило выявить закономерности, представленные в таб. 10.

Таблица 10. Основные клинико-лабораторные и инструментальные показатели, связанные с развитием ишемического инсульта у пациентов с ХИГМ (выявлены при одномерных разбиениях методом ОДР) (Показатели расположены по убыванию коэффициента качества H -информативности)

Показатели	Ниже границы (число случаев, %)		Граница	Выше границы (число случаев, %)		H - коэфф.	p
	ХИГМ	ХИГМ, ИИ		ХИГМ	ХИГМ, ИИ		
ХИГМ – стадии	113, 87%	17, 13%	2,5	4, 14%	24, 86%	4,681	<0,0005
МРТ – ИИ	32, 100%	0	1,5	1, 4%	26, 96%	3,188	<0,0005
ХС-ЛПВП	10, 30%	23, 70%	0,55	84, 92%	7, 8%	2,928	<0,0005
ХС-ЛПВП-ч	33, 55%	27, 45%	1,5	61, 95%	3, 5%	2,149	<0,0005
ХС-ЛПОНП-ч	66, 87%	10, 13%	1,5	28, 58%	20, 42%	1,214	<0,0005
ХС-ЛПОНП	89, 83%	18, 17%	1,025	5, 29%	12, 71%	1,163	<0,0005
ТГ	93, 82%	20, 18%	2,4	6, 35%	11, 65%	1,083	<0,0005
ИБС-им	112, 78%	32, 22%	1,5	5, 36%	9, 64%	1,000	0,0015
ГЛП	79, 85%	14, 15%	3,5	20, 54%	17, 43%	0,847	0,001
ТГ-ч	74, 84%	14, 16%	1,5	25, 60%	17, 40%	0,765	0,0055
ЭКГ-им	109, 78%	30, 22%	1,5	8, 42%	11, 58%	0,737	0,003
Давление сист.	115, 78%	33, 22%	235	2, 20%	8, 80%	0,733	0,002
Курение	108, 78%	31, 22%	1,5	9, 47%	10, 53%	0,698	0,008
Фибриноген – ч	48, 87%	7, 13%	1,5	69, 67%	34, 33%	0,501	0,0095
Фибриноген	117, 76%	37, 24%	6,75	0, 0%	4, 100%	0,560	0,0175
ЭКГ-нп	80,81%	19, 19%	1,5	37, 63%	22, 37%	0,590	0,0195

Из таблицы видно, что, согласно перестановочному тесту, максимально значимой и достоверной ($H = 4,68$, $p < 0,0005$) оказывается закономерность, связывающая развитие ИИ с III-ей стадией ХИГМ, вероятно, как наиболее клинически выраженной.

Также наиболее значимыми факторами риска развития ИИ являются: низкий уровень ХС-ЛПВП ($< 0,55$ ммоль/л), уровень триглицеридов ($> 2,4$ ммоль/л), признаки гипертрофии левого желудочка на ЭКГ и перенесенный инфаркт миокарда (ИБС-им).

3.4.2. Двумерные закономерности в группе больных с хронической ишемией головного мозга без случаев транзиторных ишемических атак в анамнезе

Далее была использована модель двумерных разбиений с одной граничной точкой по каждой из релевантных переменных. С использованием методики, примененной в предыдущих случаях, было выявлено 57 двумерных закономерностей.

Из набора одномерных и двумерных закономерностей была выделена базовая система, включающая 3 одномерных и 45 двумерных закономерностей.

Вклад различных показателей в базовую систему представлен в таблице 11. Вклад наиболее значимых переменных вместе с границами разбиений представлен в таблице 9. Области повышенного риска ИИ, выделяемые границами, могут быть прослежены по таблице 10 и таблице 4 приложения 1.

Таблица 11. Основные показатели, вошедшие в закономерности базовой системы, для прогнозирования развития ИИ по группе пациентов с ХИГМ без ТИА

№	Показатель	Вес	Границы
1.	ХИГМ-стадии	12,80833	2,5
2.	ХС-ЛПВП-ч	11,38855	1,5
3.	Давление сист.	7,557786	235 мм рт. ст.
4.	ЭКГ-им	5,510342	1,5
5.	ХС-ЛПВП	5,500637	0,55 ммоль/л
6.	ХС-ЛПОНП-ч	4,826254	0,63 ммоль/л
7.	ИБС-им	3,953749	1,5
8.	Пол	3,391553	1,5
9.	Тип ГЛП	3,595654	3,5
10.	ТГ-ч	3,354387	2,3 ммоль/л
11.	п/ж ЕстьИзмен	3,336459	1,5
12.	Давление диаст.	2,867962	115 мм рт.ст.
13.	Аут.Тир-т	2,851754	1,5
14.	ХС-ЛПОНП	2,759948	1,025 ммоль/л
15.	Св.х-ч	2,589052	85 мг/дл
16.	Курение	2,557985	1,5
17.	УЗИ-щит. уз.	2,459768	1,5
18.	Фибриноген – ч	2,312142	1,5
19.	Возраст	2,024206	68 лет
20.	Группы по возр.	1,81426	2,5
21.	Геп ЕстьГепат	1,756795	1,5
22.	СОЭ	1,712503	14,0 мм/час
23.	МС	1,677072	1,5
24.	ИБС-ст	1,626825	1,5
25.	ЭКГ-нп	1,620169	1,5
26.	Степень ожирения	1,191773	2,5
27.	Бр	0,981785	6,75 мкмоль/л
28.	ТГ	0,902234	2,4 ммоль/л
29.	Св.ж.к-ч	0,751657	1,5
30.	Лимфоциты	0,713563	25,5%
31.	Св.х	0,610811	73,9 мг/дл

Были проведены эксперименты по оценке точности прогноза развития ИИ в группе пациентов с ХИГМ (табл.12). Прогноз производился по переменным, входящим в базовые закономерности (см. Приложение 2) и представленным в таблице 11, за исключением показателя стадии ХИГМ. Поскольку показатель «стадия ХИГМ» был очень информативным, важнейшим фактором, определяющим риск развития ИИ, для того, чтобы выявить роль остальных клинико-лабораторных параметров, он был исключён из набора прогностических показателей. Точность прогноза оценивалась с помощью метода скользящего контроля.

Среднее по всем методам значение чувствительности составило 55,7%, среднее значение специфичности – 13,4%. Наивысшую среднюю точность (76,4%) показал метод «Логические закономерности». Близкую точность (75,4%) продемонстрировал метод МСВС.

3.5. Определение клинико-лабораторных показателей, связанных с риском развития повторного ишемического инсульта у больных ХИГМ, перенесших транзиторную ишемическую атаку

Данные группы больных ХИГМ, перенесших транзиторную ишемическую атаку, анализировали по той же схеме, что и данные группы больных без ТИА в анамнезе. Поиск одномерных и двумерных закономерностей проводили методом ОДР.

Таблица 12. Прогнозирование развития ИИ у больных ХИГМ без ТИА с помощью различных методов машинного обучения (по всем переменным, за исключением стадии ХИГМ).

<i>Метод</i>	<i>Точность в группе с ИИ-чувствительность прогноза</i>	<i>Специфичность</i>	<i>Средняя точность</i>
Линейный дискриминант Фишера	48,8%	9,4%	69,7%
Метод опорных векторов	56,1%	8,5%	73,8%
МСВС	75,6%	24,8%	75,4%
Линейная машина	48,8%	19,7%	64,6%
Q-ближайших соседей	48,8%	13,7%	67,6%
Логические закономерности	56,1	4,4%	76,4%
Среднее по всем методам	55,7%	13,4%	71,2%

3.5.1. Одномерные закономерности для группы больных хронической ишемией головного мозга, перенесших транзиторную ишемическую атаку

Результаты применения одномерной модели ОДР приведены в табл. 13.

Таблица 13. Основные клинико-лабораторные и инструментальные показатели, связанные с развитием ишемического инсульта у пациентов с ХИГМ, перенесших ТИА (выявлены при одномерных разбиениях методом ОДР) (Показатели расположены по убыванию функционала качества *H*-информативности)

<i>Показатели</i>	<i>Ниже границы (число случаев, %)</i>		<i>Граница</i>	<i>Выше границы (число случаев, %)</i>		<i>H-коэфф.</i>	<i>p</i>
	<i>ХИГМ, ТИА</i>	<i>ХИГМ, ТИА, ИИ</i>		<i>ХИГМ, ТИА</i>	<i>ХИГМ, ТИА, ИИ</i>		
ХИГМ-стадии	50, 80%	12, 20%	2,5	3, 20%	12, 80%	1,892	<0,0005
Глюк-ч	52, 76%	16, 24%	2,5	1, 11%	8, 89%	1,533	<0,0005
ИБС-ма	53, 63%	20, 27%	1,5	0, 0%	4, 100%	1,0	0,006
Глюкоза	52, 76%	16, 24%	6,3	1, 11%	8, 89%	0,931	0,002
ОХС	20, 49%	21, 51%	7,95	18, 100%	0, 0%	0,921	0,0025
ХС-ЛПНП	19, 49%	20, 51%	6,425	19, 100%	0, 0%	0,902	0,001
Фибриноген	23, 53%	21, 47%	4,725	30, 90%	3, 10%	0,773	0,0045
Лимфоциты	6, 33%	12, 67%	27,5	47, 70%	12, 30%	0,704	0,005
ХС-ЛПВП	23, 92%	2, 8%	0,65	15, 46%	18, 54%	0,674	0,0005
ХС-ЛПНП-ч	3, 27%	8, 73%	1,5	35, 74%	12, 26%	0,653	0,0075
ХС-ЛПВП-ч	32, 78%	9, 22%	1,5	7, 37%	12, 63%	0,529	0,004
ГЛП	20, 51%	19, 49%	3,5	18, 90%	2, 10%	0,516	0,0045
ЭКГ-нр	47, 75%	16, 25%	1,5	6, 43%	8, 57%	0,41	0,027

Из таблицы видно, что, наиболее значимым факторы риска ($H = 1,533$) развития ИИ у больных с ХИГМ, перенесших ТИА, является гипергликемия ($> 6,3$ ммоль/л).

Повышение уровня общего холестерина, ХС-ЛПОНП, триглицеридов и фибриногена, тип дислипидемии, перенесенный инфаркт миокарда у больных с ХИГМ, старших возрастных групп, не влияют на риск развития повторного ИИ. Уровень ХС-ЛПВП < 0,65 ммоль/л связан с риском развития ТИА, но не ИИ. Возможно, у пациентов, перенесших ТИА и имеющих такие низкие уровни ХС-ЛПВП, крайне высока летальность. Относительно связи уровней ХС-ЛПВП (> 0,65 ммоль/л) с риском ИИ у больных с ХИГМ, перенесших ТИА – вопрос требует дальнейшего изучения.

Данные в отношении связи высокого уровня общего холестерина или низкой концентрации ХС-ЛПВП с ишемическим инсультом не однозначны. Так, например, во Фрамингемском исследовании не было отмечено значимого влияния уровня ХС-ЛПВП на общий риск инсульта у мужчин и женщин и ишемического инсульта у мужчин [18]. Однако в исследовании SPARCL показано предиктивное значение низкого уровня ХС-ЛПВП для риска развития повторного инсульта [16]. Кроме того, ишемический инсульт, в основном, заболевание пожилых людей [19], а липидный профиль может меняться с возрастом [20], и это может влиять на связь между уровнями липидов и риском развития инсульта. Также возможно, что связи между уровнями липидов и риском развития ИИ являются более сложными [21], чем между уровнями липидов и ИБС [22].

По результатам проведенного нами исследования, основными причинами, приводящими к повторным ишемическим инсультам (ИИ), являются: ИБС с неклапанной мерцательной аритмией, нарушения ритма, гипергликемия > 6,3 ммоль/л. Они значительно увеличивает риск развития ишемического инсульта: в 3,7, 2,5 или 3,8 раза, соответственно (OR = 3,65 ДИ95% [2,5–5,3]; OR = 2,50 ДИ95% [1,2–4,2]; OR = 3,78 ДИ95% [2,3–6,2]).

3.5.2. Двумерные закономерности в группе больных ХИГМ, перенесших ТИА.

С использованием методики, применённой в предыдущих случаях, было выявлено 22 двумерные закономерности. Из набора одномерных и двумерных закономерностей была выделена базовая система, включающая 3 одномерных и 15 двумерных закономерностей. Вклад различных показателей в базовую систему представлен в таблице 14. Области повышенного риска ИИ, выделяемые границами, могут быть прослежены по таблице 13 и таблице 3 приложения 1.

Таблица 14. Основные показатели, вошедшие в закономерности базовой системы, для прогнозирования развития ИИ в группе пациентов, перенёсших ТИА

№	Показатель	Вес	Границы
1.	Фибриноген	5,273006	4,7 г/л
2.	АСТ	3,362968	25,5 ммоль/л
3.	ХС-ЛПВП	3,05232	0,65 ммоль/л
4.	Глюкоза - ч	2,606823	1,5
5.	ХС-ЛПНП	2,205635	6,43 ммоль/л
6.	Лимфоциты	1,756272	27,5%
7.	ХС-ЛПНП-ч	1,748902	1,5
8.	ХИГМ-стадии	1,744115	2,5
9.	ОХС	1,120775	7,95
10.	Глюкоза	0,904835	6,3 ммоль/л
11.	Бр	0,848245	12,2 мкмоль/л
12.	ХС-ЛПВП-ч	0,844878	1,5
14.	Алкоголь	0,756432	1,5
15.	УЗИ-холецистит	0,741038	1,5
16.	ЭКГ-глж	0,679956	1,5
17.	Давление сист.	0,630671	197,5 мм рт.ст.

Были проведены эксперименты по оценке точности прогноза развития ИИ по группе пациентов, перенёсших ТИА (табл.15). Прогноз производился по переменным, входящим в базовые закономерности и представленным в таблице 14, за исключением показателя стадии ХИГМ.

Точность прогноза оценивалась с помощью метода скользящего контроля.

Таблица 15. Прогнозирование развития ИИ у больных ХИГМ, перенесших ТИА, с помощью различных методов машинного обучения (по всем переменным за исключением стадии ХИГМ)

<i>Метод</i>	<i>Точность в группе с ИИ – чувствительность прогноза</i>	<i>Специфичность</i>	<i>Средняя точность</i>
Линейный дискриминант Фишера	66,7	13,2%	76,8%
Метод опорных векторов	62,5%	5,7%	78,4%
МСВС	66,7	13,2%	76,8%
Линейная машина	62,5%	17%	72,8%
Q-ближайших соседей	41,7%	5,7%	68%
Логические закономерности	70,3	13,2%	78,6%
Среднее по всем методам	61,7%	11,3%	75,2%

Среднее по всем методам значение чувствительности составило 61,7%, среднее значение специфичности – 11,3%. Наивысшую среднюю точность (78,6%) показал метод «Логические закономерности». Близкую точность (78,4%) продемонстрировал «Метод опорных векторов», а метод МСВС 76,8%.

3.6. Выявление статистически достоверных различий между закономерностями, связывающими возникновение ИИ с различными клинико-лабораторными показателями в группах больных с ХИГМ, перенесших ТИА и без ТИА

Задачей исследования явилось выявление комплексов показателей, по-разному связанных с развитием ишемического инсульта в двух группах больных с ХИГМ – перенёсших ТИА и тех, у кого ТИА в анамнезе не было.

Для определения статистически достоверных различий между закономерностями, связывающими возникновение ИИ с различными клинико-лабораторными показателями в группах больных ХИГМ с ТИА и без ТИА, была использована специальная технология, представленная в приложении 2.

Метод основан на построении оптимальных разбиений по одной из сравниваемых групп. Для оценки статистической достоверности используется статистика, оценивающая суммарное различие между группами по обоим квадрантам разбиения. Оценка достоверности основывается на использовании перестановочного теста.

Пример различий между закономерностями, связанных с риском развития ИИ, в группах пациентов с ТИА и без ТИА представлен на рисунках 6 и 7. Оптимальное разбиение строилось по группе без ТИА.

На рис. 6 показано, что при уровне глюкозы менее 6,3 ммоль/л, у больных с ХИГМ, ИИ встречается в 26%, а при уровне глюкозы более 6,3 ммоль/л – в 25%. Таким образом, у больных ХИГМ, уровень глюкозы не влияет на риск развития первичного ИИ.

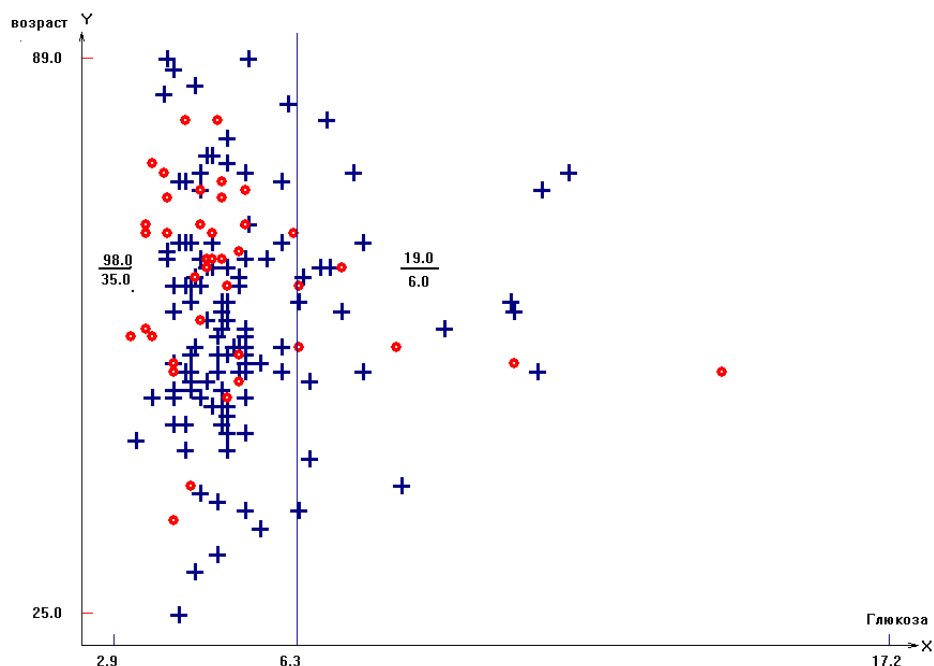


Рис. 6. Связь развития ИИ с уровнем глюкозы крови у больных с ХИГМ. По осям X и Y , соответственно, показаны уровень глюкозы (ммоль/л) и возраст (лет). Случаи с ИИ обозначены « \circ », случаи без ИИ обозначены « $+$ », $p = 0,001$.

Напротив, на рис. 7 у больных с ХИГМ, перенесших ТИА, при уровне глюкозы менее 6,3 ммоль/л – ИИ встречается в 23%, а при уровне глюкозы более 6,3 ммоль/л – в 88% случаев ($p < 0,0001$)

Таким образом, у больных с ХИГМ, перенесших ТИА, гипергликемия является важным фактором риска повторных ИИ.

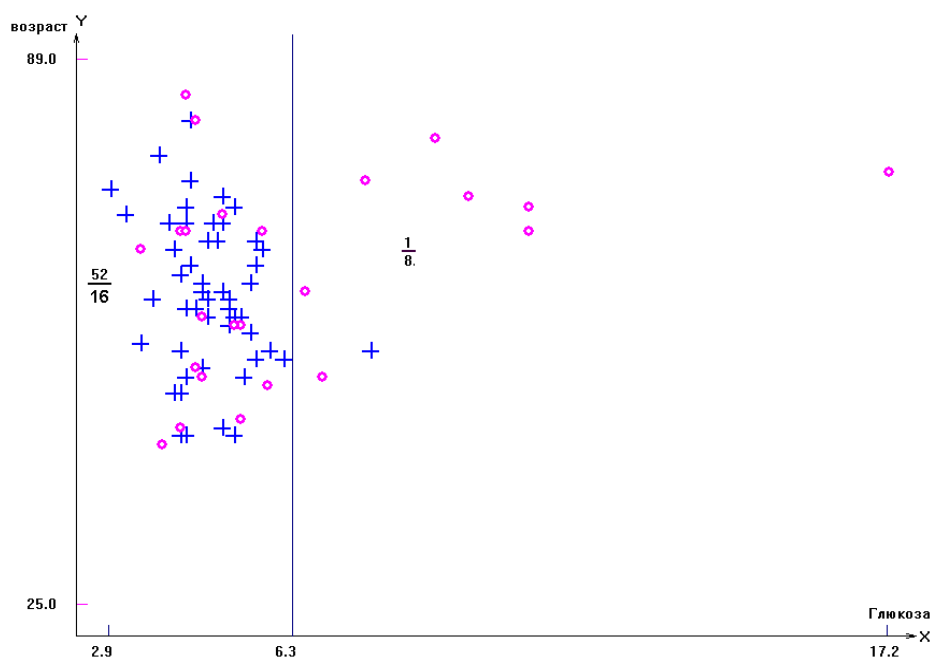


Рис. 7. Связь развития ИИ с уровнем глюкозы крови у больных с ХИГМ, перенесших ТИА. По осям X и Y , соответственно, показаны уровень глюкозы (ммоль/л) и возраст (лет). Случаи с ИИ обозначены « \circ », случаи без ИИ обозначены « $+$ », $p = 0,001$.

В таблице 16 приведены все показатели, для которых достоверность различий между соответствующими одномерными закономерностями в группах больных с ХИГМ, перенесших ТИА и без ТИА, была оценена на уровне не хуже $p < 0,01$. Оптимальное разбиение везде строилось по группе без ТИА.

В таблице используются обозначения:

$n11$ – число случаев без ИИ в левом квадранте (квадрант I);

$n21$ – число случаев с ИИ в левом квадранте (квадрант I);

$n12$ – число случаев без ИИ в правом квадранте (квадрант II);

$n22$ – число случаев с ИИ в правом квадранте (квадрант II).

Нами показано, что в одной группе может быть выраженная достоверная связь, а в другой группе такая связь отсутствует (пункты 3–6,9 таблицы 16). Связь с прогнозируемой величиной может также быть противоположной в двух группах (пункты 1, 2, 7, 8 таблицы 16).

Таблица 16. Сравнение характера взаимосвязи клинико-лабораторных показателей с развитием ИИ у больных с ХИГМ без ТИА и перенесших ТИА

Показатели	р-значение	Граница	ХИГМ		ХИГМ и ТИА	
			($n11/n21$)	($n12/n22$)	($n11/n12$)	($n12/n22$)
1. ГЛП (ниже/выше) 1 – нет ГЛП ... 2 – тип I 3 – тип IIa 4 – тип IIb	$p < 0,0005$	3,5	78/14 85% / 15%	20/17 54% / 46%	20/19 51% / 49%	18/2 90% / 10%
2. ЛПВП (ммоль/л, ниже/выше границы ОДР)	$p < 0,0005$	0,55	10/23 30% / 70%	84/7 92% / 8%	23/2 92% / 8%	15/18 45% / 55%
3. ЛПОНП моль/л (ниже/выше границы ОДР)	$p < 0,0005$	1,025	89/18 83% / 17%	5/12 29% / 71%	26/18 59% / 41%	12/2 86% / 14%
4. ЭКГ-им (1 – нет/ 2 – есть)	0,002	1,5	109/30 78% / 22%	8/11 42% / 58%	40/21 66% / 34%	16/3 84% / 16%
5. ТГ (ммоль/л, ниже/выше границы ОДР)	0,0025	2,4	93/20 82% / 18%	6/11 35% / 65%	31/20 61% / 39%	7/1 88% / 12%
6. Фибриноген (ммоль/л, ниже/выше границы ОДР)	0,004	6,75	117/37 76% / 24%	0/4 0% / 100%	49/24 65% / 35%	4/0 100% / 0%
7. ИБС-ма (1 – нет/ 2 – есть)	0,0055	1,5	105/38 73% / 27%	12/3 80% / 20%	53/20 73% / 27%	0/4 0% / 100%
8. ЭКГ-нр (нет – 1/есть – 2)	0,0075	1,5	96/37 72% / 28%	21/4 84% / 16%	47/16 75% / 25%	6/8 43% / 57%
9. Глюкоза (ммоль/л, ниже/выше границы ОДР)	0,001	6,3	98/35 74% / 26%	19/6 76% / 24%	52/16 76% / 24%	1/8 11% / 89%

Из таблицы 16 видно, что среди пациентов с ХИГМ, имеющих гиперлипидемию IIb типа, количество случаев развития ИИ значительно выше, чем у больных с другими типами ГЛП (46% против 15%). Напротив, в группе пациентов с ХИГМ, перенесших ТИА, наличие гиперлипидемии IIb типа не влияет на риск развития повторного ИИ (10% против 49%).

Для группы пациентов с ХИГМ существует статистически достоверная связь между наличием ЭКГ-признаков инфаркта миокарда и возникновением ИИ. Для группы пациентов с ХИГМ и перенесенной ТИА подобной достоверной связи не обнаружено. То же самое можно сказать при анализе таких показателей, как ХС-ЛПОНП, триглицериды, фибриноген.

Таким образом, можно говорить о том, что риск развития повторного ишемического инсульта у больных с ХИГМ старших возрастных групп, перенесших ТИА, связан с иными факторами риска, чем при развитии ТИА или первичного ИИ.

4. ВЫВОДЫ

1. Уровень ХС-ЛПВП $< 0,55$ ммоль/л, возможно, один из основных параметров, влияющих на риск развития ТИА. Прогностические факторы риска развития ТИА: показатели липидного спектра (ХС-ЛПВП, ХС-ЛПНП, ХС-ЛПОНП, общий холестерин, триглицериды), показатели клеточного состава крови (лимфоциты, СОЭ), биохимические показатели (фибриноген, АСТ, АЛТ), II тип гиперлипидемии, II-III стадии ХИГМ, ЭКГ-признаки гипертрофии левого желудочка, УЗИ-признаки диффузных изменений щитовидной железы.

2. Прогностическая точность вышеперечисленных данных факторов составляет 81,5% по методу МСВС, основанному на голосовании по системам закономерностей, вычисляемых с помощью оптимальных разбиений.

3. Риск развития повторного инсульта после перенесенной ТИА у больных с ХИГМ старших возрастных групп связан с иными факторами риска, чем при развитии ТИА или первичного ишемического инсульта.

4. У больных ХИГМ, перенесших ТИА, факторами риска последующего развития ишемического инсульта являются: ИБС с мерцательной аритмией неклапанного генеза, нарушения ритма, гипергликемия ($> 6,3$ ммоль/л). Точность прогноза развития ишемического инсульта, с использованием примененных методов распознавания, составляет 76,8%.

5. Наши исследования не выявили достоверной связи ишемического инсульта у больных ХИГМ, старших возрастных групп, перенесших ТИА, с такими показателями как повышенные уровни общего холестерина, ХС-ЛПОНП, триглицеридов, фибриногена, тип гиперлипидемии, перенесенный инфаркт миокарда.

5. ЗАКЛЮЧЕНИЕ

Сосудистые заболевания головного мозга – одна из ведущих причин заболеваемости, смертности и инвалидности в Российской Федерации. В среднем, их в России ежегодно регистрируется до 400 тысяч. Ишемический инсульт, как острая сосудистая патология, наиболее неблагоприятный вариант течения данного заболевания. Около 50% ишемических инсультов заканчиваются летальным исходом, а из выживших пациентов большинство остаются инвалидами.

Учитывая медицинскую и социальную значимость проблемы инсульта, было крайне важным выявить его прогностические критерии. Проведенные исследования с использованием новых математических методов (логико-статистические методы интеллектуального анализа данных (DataMining), система «Распознавание») позволили выявить наиболее информативные показатели для прогнозирования риска развития нарушений мозгового кровообращения.

Нами было показано, что для больных с ХИГМ пожилого возраста факторы риска практически едины, как для впервые возникшей ТИА, так и для первичного ишемического инсульта, вероятно, имеющих один механизм этиопатогенеза – атеротромботический. Показано, что риск развития повторного инсульта у больных с ХИГМ, перенесших ТИА, связан с иными факторами риска, чем при развитии ТИА или

первичного ишемического инсульта. Возможно, это обусловлено уже существующим морфологическим и неврологическим дефектом и кардиоэмболическим патогенетическим вариантом ИИ. Значимыми факторами риска повторного инсульта являются: сама перенесенная ТИА, ИБС с мерцательной аритмией неклапанного генеза, нарушения ритма, гипергликемия ($> 6,3$ ммоль/л).

Выявленные различия по факторам риска имеют важное значение для определения основных точек воздействия при проведении первичной и вторичной профилактики острых цереброваскулярных событий у пациентов старших возрастных групп.

ПРИЛОЖЕНИЕ 1

В приложении представлены закономерности, характеризующие связь ТИА и ишемического инсульта с клиническими и биохимическими показателями, и образующие соответствующие базовые системы.

В таблицах, соответствующих двумерным закономерностям, строящимся в рамках модели III с одной граничной точкой для каждой из переменных, описанию одной закономерности отведено две строки. В двух левых ячейках верхней строки, находящихся в столбцах Факт.-1 и Факт.-2, даны названия релевантных (см. приложение 2) переменных (факторов). В ячейках верхней строки из столбцов Гран.-1 и Гран.-2 даны граничные точки. В ячейках верхней строки из столбцов, обозначенных P1, N1, P2, N2, даны значения показателей, характеризующие достоверность описываемых закономерностей. Смысл данных показателей описан в приложении 2. В последних четырёх столбцах, обозначенных римскими цифрами I, II, III, IV даны количества случаев ТИА (верхняя строка) и с ТИА (нижняя строка) в квадрантах двумерного разбиения. Смысл квадрантов разъяснён в таблице 1.

Квадрант I -соответствует ситуации Факт.-1 < Гран.-1, Факт.-2 > Гран.-2	
Квадрант II -соответствует ситуации Факт.-1 > Гран.-1, Факт.-2 > Гран.-2	
Квадрант III -соответствует ситуации Факт.-1 > Гран.-1, Факт.-2 < Гран.-2	
Квадрант IV -соответствует ситуации Факт.-1 < Гран.-1, Факт.-2 < Гран.-2	

Таблица 1. Двумерные закономерности, характеризующие связь ТИА с клиническими и биохимическими показателями и образующие базовую систему

№ закономерности	Фактор -1	Фактор- 2	Гран.-1	Гран.-2	$P1$	$H1$	$P2$	$H2$	I	II	III	IV
1.	Группы по возраст	УЗДГ-изв	1,5	1,5	0,0055	0,79	0,002	0,83	15	26	4	19
	—	—	—	—	—	—	—	—	5	3	7	3
2.	Наследственность	ЛПВП-ч	1,5	1,5	< 0,0005	1,509	0,003	0,85	47	14	6	27
	—	—	—	—	—	—	—	—	7	0	17	15
3.	ИБС-атер	Диф.-Изм.Щит.	1,5	1,5	0,0055	0,71	0,0025	0,9	24	49	19	4
	—	—	—	—	—	—	—	—	1	26	1	0
4.	ХИГМ-стадии	Диф.-Изм.Щит.	2,5	1,5	0,005	0,908	0,0015	0,779	28	45	14	9
	—	—	—	—	—	—	—	—	1	26	0	1
5.	ХИГМ-стадии	Холестер.	2,5	5,75	0,0015	0,87	< 0,0005	1,04	25	38	37	17
	—	—	—	—	—	—	—	—	2	43	7	1
6.	ХИГМ-стадии	Фибр-ч	2,5	1,5	0,001	0,83	< 0,0005	1,29	32	37	38	10
	—	—	—	—	—	—	—	—	3	40	10	0
7.	ХИГМ-стадии	ЭКГ-ИМ	2,5	2,5	0,0005	0,95	< 0,0005	1,12	6	2	73	36
	—	—	—	—	—	—	—	—	1	12	38	2
8.	ХИГМ-стадии	ГЛП	2,5	3,5	0,003	0,907	< 0,0005	1,13	8	12	53	26
	—	—	—	—	—	—	—	—	0	18	19	1
9.	ХИГМ-стадии	ТГ	2,5	2,05	0,001	0,816	0,002	0,9	4	4	61	30
	—	—	—	—	—	—	—	—	0	13	24	1
10.	ХИГМ-стадии	ЛПВП	2,5	0,85	< 0,0005	1,35	0,004	0,907	21	51	9	13
	—	—	—	—	—	—	—	—	0	12	25	1
11.	ХИГМ-стадии	ЛПВП-ч	2,5	1,5	< 0,0005	1,39	< 0,0005	1,09	19	42	18	15
	—	—	—	—	—	—	—	—	0	7	31	1
12.	ХИГМ-стадии	ЛПОНП	1,5	0,89	< 0,0005	1,21	0,002	0,95	4	2	58	30
	—	—	—	—	—	—	—	—	0	15	22	1
13.	ХИГМ-стадии	ЛПОНП-ч	1,5	1,5	0,0065	0,797	< 0,0005	1,03	10	18	42	24
	—	—	—	—	—	—	—	—	0	22	15	1
14.	Диф.-Изм.Щит.	ЛПВП	1,5	0,65	< 0,0005	1,37	0,0055	0,96	17	53	6	5
	—	—	—	—	—	—	—	—	0	8	13	0
15.	Диф.-Изм.Щит.	ЛПВП-ч	1,5	1,5	< 0,0005	1,43	0,006	0,82	13	40	19	9
	—	—	—	—	—	—	—	—	0	4	18	0
16.	Лимфоцит	АСТ	25,5	25,5	< 0,0005	1,03	0,0035	0,77	17	30	61	9

ЛОГИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ СВЯЗИ КЛИНИКО-ЛАБОРАТОРНЫХ ПОКАЗАТЕЛЕЙ С ВОЗНИКНОВЕНИЕМ

	—	—	—	—	—	—	—	—	0	36	13	4
17.	СОЭ	МРТ-ЛО	12,7	1,5	0,006	0,89	0,0075	0,74	9	1	15	8
	—	—	—	—	—	—	—	—	2	7	3	6
18.	Холестер.	Фибрин-н	5,25	4,325	< 0,0005	1,00005	0,001	0,95	19	24	57	17
	—	—	—	—	—	—	—	—	1	35	15	2
19.	Холестер.	Фибр-ч	5,25	1,5	0,0065	0,816	0,002	0,74	26	43	38	10
	—	—	—	—	—	—	—	—	2	41	9	1
20.	Холестер.	ЭКГ-глж	5,25	1,5	< 0,0005	1,11	< 0,0005	1,089	27	53	28	9
	—	—	—	—	—	—	—	—	1	49	1	1
21.	Холестер.	ЛПВП-ч	5,4	1,5	< 0,0005	1,45	0,0055	0,906	19	42	20	13
	—	—	—	—	—	—	—	—	3	4	31	1
22.	обХЛ-ч	ЭКГ-глж	1,5	1,5	< 0,0005	1,32	0,0005	0,946	19	61	30	7
	—	—	—	—	—	—	—	—	1	49	1	1
23.	Фибрин-н	ЛПВП	4,875	0,45	< 0,0005	1,72	< 0,0005	1,05	76	11	4	3
	—	—	—	—	—	—	—	—	6	13	3	16
24.	АЛТ-ч	УЗДГ-изв	1,5	1,5	0,0015	0,85	0,0035	0,65	40	1	5	18
	—	—	—	—	—	—	—	—	5	3	0	10
25.	ЭКГ-им	ЛПОНП-ч	1,5	1,5	0,0045	0,81	0,003	0,833	28	0	4	62
	—	—	—	—	—	—	—	—	18	4	4	12
26.	ЭКГ-глж	ГЛП	1,5	3,5	0,0055	0,84	0,008	0,891	5	15	54	25
	—	—	—	—	—	—	—	—	0	18	17	2
27.	ЭКГ-глж	ЛПНП	1,5	6,15	< 0,0005	1,15	0,004	0,826	10	9	56	19
	—	—	—	—	—	—	—	—	1	20	15	1
28.	ГЛП	ЛПВП-ч	3,5	1,5	< 0,0005	1,07	0,004	0,961	48	13	5	28
	—	—	—	—	—	—	—	—	5	2	16	15
29.	ХС-ч	ЛПВП-ч	1,5	1,5	< 0,0005	2,31	0,005	0,799	15	46	22	10
	—	—	—	—	—	—	—	—	4	3	29	2
30.	ТГ	ЛПВП-ч	1,85	1,5	< 0,0005	1,15	0,003	0,88	52	9	2	31
	—	—	—	—	—	—	—	—	4	3	13	18
31.	ТГ-ч	ЛПВП-ч	1,5	1,5	< 0,0005	1,23	0,0085	0,84	44	17	6	27
	—	—	—	—	—	—	—	—	4	3	16	15
32.	ЛПНП	ЛПВП-ч	7,2	1,5	< 0,0005	1,29	0,003	0,88	52	9	1	32
	—	—	—	—	—	—	—	—	7	0	12	19
33.	ЛПВП-ч	ЛПОНП	1,5	0,84	0,002	0,95	0,0005	0,998	2	5	56	31
	—	—	—	—	—	—	—	—	14	2	5	17
34.	ЛПВП-ч	ЛПОНП-ч	1,5	1,5	0,0015	0,89	< 0,0005	1,41	9	19	42	24
	—	—	—	—	—	—	—	—	20	2	5	11
35.	Вазилип	УЗДГ-изв	0,5	1,5	0,011	0,84	0,0005	0,847	32	9	1	22
	—	—	—	—	—	—	—	—	6	2	6	4

Например, для закономерности, характеризующей связь возникновения ТИА с переменными «Наследственность» и «ЛПВП-ч»:

- в квадрант I (нет наследственной предрасположенности к сердечно-сосудистым заболеваниям и ЛПВП-ч = 2, т.е. ЛПВП > 1,0 ммоль/л) попало 47 случаев без ТИА и 7 случаев с ТИА;
- в квадрант II (есть наследственная предрасположенность к сердечно-сосудистым заболеваниям и ЛПВП-ч = 2, т.е. ЛПВП > 1,0 ммоль/л) попало 14 случаев без ПНМК и 0 случаев с ПНМК;
- в квадрант III (есть наследственная предрасположенность к сердечно-сосудистым заболеваниям и ЛПВП-ч = 1, т.е. ЛПВП < 1,0 ммоль/л) попало 6 случаев без ТИА и 17 случаев с ТИА;
- в квадрант IV (нет наследственной предрасположенности к сердечно-сосудистым заболеваниям и ЛПВП-ч = 1, т.е. ЛПВП < 1,0 ммоль/л) попало 27 случаев без ТИА и 15 случаев с ТИА.

Таким образом, низкий уровень ЛПВП (< 1,0 ммоль/л), в сочетании с наследственной предрасположенностью к сердечно-сосудистым заболеваниям, значительно усиливает риск возникновения ТИА.

Таблица 2. Двумерные закономерности, характеризующие связь ишемического инсульта с клиническими и биохимическими показателями и образующие базовую систему (для всей исследуемой выборки)

№ закономерности	Фактор - 1	Фактор -2	Гран.-1	Гран.-2	p1	H1	p2	H2	I	II	III	IV
1.	Рост	ЛПВП	1,775	0,35	< 0,0005	1,23	0,003	0,816	114	3	5	10
	–	–	–	–	–	–	–	–	26	6	0	18
2.	Стадии-ожирен.	Св.х	2,5	78,05	0,0095	0,712	0,0075	0,683	16	3	32	37
	–	–	–	–	–	–	–	–	1	6	3	17
3.	МС	Св.х	1,5	78,05	0,0025	0,828	0,005	0,788	16	3	26	43
	–	–	–	–	–	–	–	–	1	6	2	18
4.	Дсис	Св.х-ч	223	2,5	0,003	0,849	< 0,0005	1,119	19	0	5	64
	–	–	–	–	–	–	–	–	1	5	0	21
5.	Ддиас	Узлов. Зоб	> 85,0	1,5	0,0155	0,686	0,008	0,519	12	45	42	25
	–	–	–	–	–	–	–	–	9	6	16	2
6.	ХИГМ-стадии	Билирубин	2,5	6,75	0,0005	0,988	< 0,0005	4,116	139	3	4	24
	–	–	–	–	–	–	–	–	21	36	0	8
7.	ХИГМ-стадии	УЗИ-щит.уз.	2,5	1,5	0,000500	0,998	< 0,0005	2,465	66	0	3	46
	–	–	–	–	–	–	–	–	14	13	2	3

ЛОГИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ СВЯЗИ КЛИНИКО-ЛАБОРАТОРНЫХ ПОКАЗАТЕЛЕЙ С ВОЗНИКНОВЕНИЕМ

8.	ТИА	Диф.-Изм.Щит.	1,5	1,5	0,0190	0,763	< 0,0005	1,14	73	27	1	23
	–	–	–	–	–	–	–	–	14	14	4	1
9.	ТИА	ГЛП	1,5	3,5	< 0,0005	1,12	< 0,0005	1,16	20	18	20	79
	–	–	–	–	–	–	–	–	17	2	19	14
10.	ТИА	ЛПВП	1,5	0,55	< 0,0005	2,65	0,0005	1,73	84	16	22	10
	–	–	–	–	–	–	–	–	7	18	2	23
11.	ТИА	ЛПВП-ч	1,5	1,5	< 0,0005	2,46	< 0,0005	1,57	61	7	32	33
	–	–	–	–	–	–	–	–	3	12	9	27
12.	ТИА	ЛПОНП	1,5	1,07	< 0,0005	1,23	< 0,0005	1,18	4	11	27	90
	–	–	–	–	–	–	–	–	11	0	20	19
13.	ТИА	Св.х-ч	1,5	2,5	0,007000	0,724	< 0,0005	1,21	12	7	15	54
	–	–	–	–	–	–	–	–	6	0	15	6
14.	Лимфоцит	ЛПВП-ч	23,5	1,5	0,000000	1,022	0,0005	0,983	14	54	64	1
	–	–	–	–	–	–	–	–	3	12	26	10
15.	Лимфоцит	геп НетИзмен.	25,5	1,2	0,013000	0,709	0,0035	0,79	16	59	79	12
	–	–	–	–	–	–	–	–	2	20	23	18
16.	Триглицер	МРТ-ИИ	0,85	1,5	0,000000	3,68	0,008	0,99	2	2	45	2
	–	–	–	–	–	–	–	–	0	43	0	0
17.	Св.х	п/ж ЕстьИзмен	71,7	1,5	0,007000	0,787	0,006	0,658	22	4	23	38
	–	–	–	–	–	–	–	–	2	7	2	15
18.	УЗДГ-Сз	геп НетИзмен.	1,5	1,5	0,012500	0,597	0,003000	0,942	27	9	3	42
	–	–	–	–	–	–	–	–	3	2	7	11
19.	Рг-аорты	УЗИ-п/ж	1,5	1,5	0,009000	0,804	0,004000	0,744	17	48	28	14
	–	–	–	–	–	–	–	–	0	35	6	4
20.	Рг-аорты	п/ж НетИзмен	1,5	0,5	0,005000	0,596	0,001000	0,927	14	28	48	17
	–	–	–	–	–	–	–	–	4	6	35	0

Таблица 3. Двумерные закономерности, характеризующие связь ишемического инсульта с клиническими и биохимическими показателями в группе пациентов с ХИГМ, перенёсших ТИА и входящих в базовую систему

№ закономерности	Фактор- 1	Фактор - 2	Граница-1	Граница-2	H1	p1	H2	p2	I	II	III	IV
1.	СЕТР	ЛПНП-ч	1,5	1,5	0,87	0,0005	0,74	0,0075	7	24	0	3
	—	—	—	—	—	—	—	—	6	5	5	1
2.	СЕТР_V1 V1	ЛПНП-ч	1,5	1,5	0,87	0,0005	0,82	0,0095	24	7	3	0
	—	—	—	—	—	—	—	—	5	6	1	5
3.	СЕТР V1V2	УЗИ-хол- т	1,5	1,5	0,74	0,006	0,62	0,0085	12	19	2	12
	—	—	—	—	—	—	—	—	7	5	5	1
4.	Дсис	ЭКГ-глж	197,5	1,5	0,68	0,018	0,63	0,0060	42	8	1	1
	—	—	—	—	—	—	—	—	9	11	0	4
5.	Алкоголь	АСТ	1,5	31,5	0,76	0,009	0,76	0,0050	19	0	2	32
	—	—	—	—	—	—	—	—	1	3	0	20
6.	ХИГМ- стадии	ЛПВП	2,5	0,60	0,83	0,002	0,74	0,0080	15	0	2	21
	—	—	—	—	—	—	—	—	9	9	0	2
7.	Лимфоцит	АСТ	27,5	23,5	0,91	0,0115	0,90	0,0030	1	39	8	5
	—	—	—	—	—	—	—	—	7	3	9	5
8.	Фибрин-н	Бр	4,375	12,2	0,85	0,008	0,95	0,0005	7	3	33	10
	—	—	—	—	—	—	—	—	0	5	4	15
9.	Фибрин-н	АСТ	4,6	25,5	0,79	0,005	0,93	0,0025	17	19	12	5
	—	—	—	—	—	—	—	—	4	3	1	16
10.	Фибрин-н	Глюкоза	4,875	5,65	0,90	0,0030	0,82	0,0055	0	4	19	30
	—	—	—	—	—	—	—	—	9	1	0	14
11.	Фибрин-н	Глюк-ч	4,875	2,5	0,86	0,0035	0,80	0,0090	0	1	22	30
	—	—	—	—	—	—	—	—	7	1	0	16
12.	Фибрин-н	ЛПНП	4,875	6,425	1,09	< 0,0005	0,86	0,0025	14	5	11	8
	—	—	—	—	—	—	—	—	0	0	0	20
13.	Фибрин-н	ЛПВП	4,875	0,65	1,14	< 0,0005	0,91	0,0010	5	10	6	17
	—	—	—	—	—	—	—	—	18	0	0	2
14.	Глюк-ч	ЛПВП-ч	2,5	1,5	0,84	0,0060	0,75	0,0015	7	0	1	31
	—	—	—	—	—	—	—	—	8	4	4	5

Таблица 4. Двумерные закономерности, характеризующие связь ишемического инсульта с клиническими и биохимическими показателями в группе пациентов с хронической ишемией головного мозга (ХИГМ), без транзиторных ишемических атак (ТИА), и входящих в базовую систему

№ закономерности	Фактор -1	Фактор- 2	Граница-1	Граница-2	p1	H1	p2	H2	I	II	III	IV
1.	Пол	ИБС-им	1,5	1,5	0,94	0,0005	0,82	0,0080	5	0	90	22
									3	6	21	11
2.	Пол	ЭКГ-ИМ	1,5	1,5	1,36	< 0,0005	0,82	0,0085	4	4	86	23
									1	10	17	13
3.	Возраст	Группы-повозр	68,5	2,5	0,82	0,0020	0,81	0,0005	0	19	3	95
									0	4	12	25
4.	Стадии-ожирен.	Аут.Тир-т	2,5	1,5	0,84	0,0020	0,54	0,0015	11	4	31	50
									0	4	1	10
5.	МС	Аут.Тир-т	1,5	1,5	0,64	0,0035	0,76	0,0030	11	4	26	55
									0	4	1	10
6.	Дсис	Фибр-ч	235,0	1,5	0,91	0,0045	0,96	0,0005	67	2	0	48
									30	4	4	3
7.	Дсис	ГЛП	235,0	3,5	0,76	0,0075	0,83	0,0080	20	0	2	77
									14	3	4	10
8.	Дсис	ЛПОНП	235,0	1,025	1,02	< 0,0005	0,90	0,0060	5	0	2	87
									11	1	6	12
9.	Дсис	Св.х-ч	205,0	2,5	0,86	0,0055	1,00	0,0005	12	0	6	48
									1	5	0	6
11.	Ддиас	ИБС-ст	115,0	1,5	0,73	0,0105	0,82	0,0030	28	1	9	79
									8	8	3	22
12.	Ддиас	Св.х-ч	115,0	2,5	0,79	0,0080	0,94	0,0075	12	0	7	47
									2	4	1	5
13.	Ддиас	УЗИ-хол-т	107,5	1,5	0,78	0,0070	1,11	0,0000	65	10	14	25
									14	15	2	10
14.	Курение	СОЭ	1,5	14,5	0,90	0,0010	0,90	0,0015	29	6	3	79
									19	2	8	12
15.	Курение	ЭКГ-ИМ	1,5	1,5	1,00	0,0005	0,66	0,0080	8	0	9	100
									9	2	8	22
16.	Курение	п/ж ЕстьИзмен	1,5	0,5	0,84	0,0065	1,00	0,0015	27	0	9	78
									14	4	6	17
17.	ИБС-ст	ЛПВП-ч	1,5	1,5	1,40	< 0,0005	0,90	0,0055	47	14	6	27
									1	2	13	14

18.	ИБС-им	ГЛП	1,5	3,5	0,79	0,0025	0,79	0,0010	19	1	2	77
									15	2	4	10
19.	ИБС-им	ЛПОНП-ч	1,5	1,5	0,89	0,0005	0,97	0,0005	28	0	2	64
									17	3	3	7
20.	ХИГМ-стадии	Фибрин-н	2,5	6,75	0,95	0,0085	2,47	< 0,0005	0	0	4	113
									3	1	23	14
21.	ХИГМ-стадии	Бр	2,5	6,75	0,98	0,0010	2,97	< 0,0005	98	1	3	15
									12	24	0	5
22.	ХИГМ-стадии	ЛПВП	2,5	0,55	1,04	< 0,0005	0,95	0,0005	80	4	0	10
									4	3	15	8
23.	ХИГМ-стадии	ЛПВП-ч	2,5	1,5	1,16	< 0,0005	1,53	< 0,0005	57	4	0	33
									2	1	17	10
24.	ХИГМ-стадии	ЛПОНП	2,5	1,025	0,96	0,0045	1,48	< 0,0005	5	0	4	85
									6	6	12	6
25.	ХИГМ-стадии	УЗИ-щит.уз.	2,5	1,5	0,93	0,0025	1,44	< 0,0005	50	0	3	36
									7	7	1	1
26.	Лимфоцит	Св.ж.к-ч	25,5	1,5	0,75	0,0165	0,71	0,0025	3	32	23	8
									7	2	3	0
27.	СОЭ	УЗДГ-Сз	14,5	1,5	0,67	0,0145	0,81	0,0030	9	1	16	38
									0	5	3	2
28.	Фибрин-н	ЛПОНП-ч	6,375	1,5	0,94	0,0005	0,76	0,0095	26	2	0	66
									18	2	3	7
29.	Фибр-ч	ЭКГ-нп	1,5	1,5	0,70	0,0125	0,67	0,0155	18	19	50	30
									4	18	16	3
30.	ЭКГ-ИМ	ЛПОНП-ч	1,5	1,5	1,11	< 0,0005	0,94	0,0055	28	0	4	62
									17	3	3	7
31.	ЭКГ-нп	п/ж ЕстьИзмен	1,5	0,5	0,84	0,0030	0,92	0,0020	21	6	31	56
									6	12	10	13
32.	ГЛП	ЛПВП	3,5	0,75	1,74	< 0,0005	1,02	< 0,0005	63	16	2	13
									6	0	17	7
33.	ГЛП	ЛПВП-ч	3,5	1,5	1,27	< 0,0005	1,02	< 0,0005	48	13	5	28
									3	0	17	10
34.	ТГ	ЛПВП-ч	1,15	1,5	1,32	< 0,0005	0,90	0,0010	33	28	11	22
									2	1	22	5
35.	ТГ-ч	ЛПВП	1,5	0,75	1,89	< 0,0005	0,96	0,0020	58	21	2	13
									6	0	17	7
36.	ТГ-ч	ЛПВП-ч	1,5	1,5	1,48	< 0,0005	0,99	0,0005	44	17	6	27
									3	0	17	10
37.	ТГ-ч	геп ЕстьГепат.	1,5	0,5	0,76	0,0110	0,83	0,0025	23	7	18	50

ЛОГИКО-СТАТИСТИЧЕСКИЙ АНАЛИЗ СВЯЗИ КЛИНИКО-ЛАБОРАТОРНЫХ ПОКАЗАТЕЛЕЙ С ВОЗНИКНОВЕНИЕМ

	—	—	—	—	—	—	—	—	5	12	5	9
38.	ТГ-ч	УЗИ-щит.уз.	1,5	1,5	0,75	0,0040	0,58	0,0145	33	9	11	25
	—	—	—	—	—	—	—	—	5	8	0	1
39.	ЛПВП-ч	ЛПОНП	1,5	0,525	0,77	0,0035	1,13	< 0,0005	11	24	37	22
	—	—	—	—	—	—	—	—	21	1	2	6
40.	ЛПВП-ч	ЛПОНП-ч	1,5	1,5	0,97	0,0010	1,24	< 0,0005	9	19	42	24
	—	—	—	—	—	—	—	—	19	1	2	8
41.	ЛПВП-ч	УЗИ-щит.уз.	1,5	1,5	0,78	0,0095	1,31	< 0,0005	12	28	20	14
	—	—	—	—	—	—	—	—	11	1	0	1
42.	Св.х	п/ж ЕстьИзмен	73,9	0,5	0,72	0,0135	0,61	0,0115	18	0	15	32
	—	—	—	—	—	—	—	—	1	3	2	6
43.	Св.ж.к-ты	геп НетИзмен.	33,45	0,5	0,69	0,0185	0,60	0,0050	27	0	7	31
	—	—	—	—	—	—	—	—	1	3	1	7
44.	УЗДГ-Сз	геп ЕстьГепат.	1,5	0,5	1,00	0,0030	1,00	0,0015	19	0	9	35
	—	—	—	—	—	—	—	—	4	3	2	1
45.	УЗДГ-Сз	п/ж ЕстьИзмен	1,5	0,5	0,93	0,0025	0,96	0,0010	15	0	9	39
	—	—	—	—	—	—	—	—	1	4	1	4

ПРИЛОЖЕНИЕ 2

1. Метод оптимальных достоверных разбиений

Метод оптимальных достоверных разбиений (ОДР) является методом интеллектуального анализа данных (Data Mining), который был ранее подробно описан в ряде статей [8, 23–25]. В настоящей работе использована модификация ОДР, использующая новый вариант перестановочного теста для верификации двумерных разбиений с одной граничной точкой по каждой из координатных осей, а также дополненная рядом инструментов, включая метод выделения базовой подсистемы закономерностей и метод оценки достоверных различий между закономерностями, существующими в двух разных группах объектов.

Метод ОДР основан на поиске разбиений признакового пространства, соответствующего набору признаков $\{X_1, \dots, X_n\}$, в рамках заранее фиксированных семейств (моделей), позволяющих наилучшим образом разделить объекты с различными уровнями значений прогнозируемой величины Y . До сих пор использовались одномерные и двумерные модели различного типа. Поиск оптимальных разбиений производится отдельно для каждого признака в случае использования одномерных моделей и для всевозможных пар признаков в случае использования двумерных моделей. Признаки, используемые при построении некоторой закономерности, называются релевантными ей.

Найденное оптимальное разбиение считается статистически достоверной закономерностью после проведения процедуры верификации. Верификация производится с помощью перестановочных тестов. Предполагается, что обучающая выборка представима в виде $\tilde{S}_0 = \{(y_1, \mathbf{x}_1), \dots, (y_m, \mathbf{x}_m)\}$, где y_j – часть описания, соответствующая прогнозируемой переменной Y , \mathbf{x}_j – вектор прогностических переменных.

Первоначальный вариант перестановочного теста предназначен для проверки нулевой гипотезы H_0 о независимости прогнозируемой переменной Y от потенциальных прогностических переменных X_1, \dots, X_n . Пусть $T(\tilde{S})$ – статистика, характеризующая степень зависимости Y от переменных X_1, \dots, X_n . В качестве статистики обычно используется максимальное значение функционала, характеризующего качество разбиения [8].

В качестве p -значения рассматривается условная вероятность $\Pr\{T(\tilde{S}) > T(\tilde{S}_0) \mid \mathbf{H}_0, W^p(\tilde{S}_0)\}$, где $W^p(\tilde{S}_0)$ – множество всех возможных выборок, совпадающих по числу объектов с \tilde{S}_0 , для которых эмпирические распределения частей описания Y и частей описания \mathbf{x} совпадают с соответствующими распределениями для \tilde{S}_0 . В работе [8] доказано справедливость равенства

$$\Pr\{T(\tilde{S}) > T(\tilde{S}_0) \mid \mathbf{H}_0, W^p(\tilde{S}_0)\} = \frac{N_1^p[T(\tilde{S}_0)]}{m!},$$

где $N_1^p[T(\tilde{S}_0)]$ является числом перестановок, генерирующих выборки, для которых справедливо неравенство $T(\tilde{S}) > T(\tilde{S}_0)$. Прямой расчёт отношения $N_1^p[T(\tilde{S}_0)]/m!$ чрезвычайно трудоёмок, что делает его невозможным для подавляющего числа прикладных задач. Однако достаточно легко осуществима Монте-Карло оценка, основанная на генерации представительного множества перестановок с помощью датчика случайных чисел.

Одной из проблем, связанных с использованием метода ОДР, является проблема избыточности закономерностей, найденных с использованием более сложных моделей

разбиений. Ранее для решения данной проблемы использовался метод, основанный на использовании перестановочных тестов, генерирующих случайные выборки с помощью случайных перестановок позиций относительно фиксированных векторов прогностических признаков только внутри подвыборок исходной обучающей выборки, индуцированных одномерными разбиениями, соответствующими простейшим закономерностям с одной граничной точкой [8, 23]. Однако данный подход не может быть использован для верификации двумерных закономерностей, если простейших достоверных одномерных закономерностей для признаков, релевантных верифицируемой двумерной закономерности, не существует.

Кроме того, эксперименты с реальными выборками показали, что при значительном отклонении границы двумерного разбиения для некоторого признака $X_{i'}$ от границы простейшей одномерной закономерности для $X_{i'}$, описанная выше процедура иногда приводит к ошибочной верификации как статистически достоверного вклада второго признака $X_{i'}$ в двумерную закономерность.

В связи с вышесказанным более эффективным при верификации двумерных закономерностей с граничными линиями, параллельными координатным осям (модель III, см.[8]), оказывается способ, использующий случайные перестановки внутри подобластей одномерных разбиений, задаваемых граничными точками, описывающими оптимальное двумерное разбиение.

Предположим, что по обучающей выборке $\tilde{S}_0 = \{(y_1, \mathbf{x}_1), \dots, (y_m, \mathbf{x}_m)\}$ найдено оптимальное разбиение R совместной области допустимых значений признаков $X_{i'}$ и $X_{i''}$ с максимальным значением интегрального функционала качества $F_I(R, \tilde{S}_0)$ (см. [8]). Пусть R задаётся с помощью границы $b_{i'}$ для признака $X_{i'}$ и $b_{i''}$ для признака $X_{i''}$. Для верификации необходимости использования в закономерности признака $X_{i'}$ обучающая выборка \tilde{S}_0 разбивается на две подвыборки: $\tilde{S}_i^{i''} = \{(y_1^i, \mathbf{x}_1^i), \dots, (y_{m_i}^i, \mathbf{x}_{m_i}^i)\}$, включающую объекты \tilde{S}_0 с $X_{i'} < b_{i'}$, и $\tilde{S}_r^{i''} = \{(y_1^r, \mathbf{x}_1^r), \dots, (y_{m_r}^r, \mathbf{x}_{m_r}^r)\}$, включающую объекты \tilde{S}_0 с $X_{i'} > b_{i'}$.

Сгенерируем N пар независимых случайных перестановок наборов натуральных чисел $\{1, \dots, m_i\}$ и $\{1, \dots, m_r\}$: $\{\tilde{f}_i^t = \{f_{i1}^t, \dots, f_{im_i}^t\}, \tilde{f}_r^t = \{f_{r1}^t, \dots, f_{rm_r}^t\} | t \in \{1, \dots, N\}\}$. Для каждой пары перестановок по объединению выборок $\tilde{S}_i^{ii''} = \{(y_{f_{i1}^t}^i, \mathbf{x}_1^i), \dots, (y_{f_{im_i}^t}^i, \mathbf{x}_{m_i}^i)\}$ и $\tilde{S}_r^{ii''} = \{(y_{f_{r1}^t}^r, \mathbf{x}_1^r), \dots, (y_{f_{rm_r}^t}^r, \mathbf{x}_{m_r}^r)\}$, которое обозначим как $\tilde{S}_o^{ii''}$, найдём оптимальное разбиение совместной области допустимых значений признаков $X_{i'}$ и $X_{i''}$ с максимальным значением функционала качества $F_I(R^t, \tilde{S}_o^{ii''})$. Иными словами, поиск оптимальных разбиений производится на случайной выборке, полученной из исходной обучающей выборки с помощью случайных перестановок позиций прогнозируемой переменной Y относительно фиксированных позиций векторов переменных X . Однако перестановки разрешены только внутри одномерных областей, задаваемых граничной точкой оптимального двумерного разбиения для признака $X_{i'}$. В качестве меры статистической достоверности вклада признака $X_{i'}$ может быть использована доля перестановок $p_{i'}$, для которых $F_*(R^t, \tilde{S}_o^{ii''}) > F_*(R, \tilde{S}_0)$. Как и в предыдущих вариантах метода ОДР, данная доля перестановок может рассматриваться в качестве аналога понятия p -значения, используемого в стандартных статистических тестах. При этом в качестве нулевой гипотезы используется предположение о независимости

прогнозируемой величины от признаков X внутри одномерных подобластей, задаваемых граничной точкой $b_{i'}$ для признака $X_{i'}$.

Двумерная закономерность из модели III считается достоверной, если достоверным является вклад обоих релевантных ей признаков.

Дополнительным способом оценки достоверности вклада в закономерность R релевантного ей признака $X_{i'}$ является отношение

$$H_{i'} = \frac{F_*(R, \tilde{S}_0)}{\max_{t \in \{1, \dots, N\}} [F_*(R^t, \tilde{S}_0^{t''})]}, \quad (1)$$

которое позволяет сравнить вклады признака $X_{i'}$ в закономерности, для которых оценки p -значений равны 0. Иными словами, $H_{i'}$ является отношением значения функционала качества достигнутого на исходной обучающей выборке отношению к максимальному значению функционала качества, достигнутому на случайных выборках, полученных с помощью перестановок. Очевидно, что величина $H_{i'}$ снижается при увеличении числа перестановок. Поэтому сравнивать H -коэффициенты, рассчитанные по формуле (1), имеет смысл только при соблюдении равенства числа перестановок, используемых для их расчёта. Оценка вклада признака $X_{i'}$ в некоторую систему закономерностей \tilde{R} , является сумма $\sum_{R \in \tilde{R}} H_{i'}$, где суммирование производится по всем закономерностям, которым признак $X_{i'}$ релевантен. **Данная сумма в статье называется «весом» признака.**

2. Базовые системы закономерностей

Наборы закономерностей, найденных по массивам клинико-биологических данных нередко бывают весьма обширны, что вызывает необходимость автоматического выделения основных эффектов и отсеивания второстепенных. В настоящем разделе описывается подход к выделению основных закономерностей, использующий понятие доминирования, основанное на возможности поставить каждой закономерности R_i , найденной с помощью метода ОДР, алгоритм, прогнозирующий величине Y .

Действительно, пусть закономерность R_i задаёт разбиение признакового пространства с элементами $\{q_1, \dots, q_l\}$, которое индуцирует разбиение обучающей выборки на подвыборки $\{\tilde{s}_1, \dots, \tilde{s}_l\}$. В качестве прогностического алгоритма, соответствующего закономерности R_i , может быть взят алгоритм, вычисляющий прогноз $\hat{y}(s)$ для произвольного объекта s с описанием $\mathbf{x}(s) \in q_i$ как среднее по всем объектам обучающей выборки из элемента q_i : $\hat{y}(s) = \frac{1}{m_i} \sum_{\mathbf{x}_j \in q_i} y_j$. Подобный прогностический алгоритм, а также вычисляемую им прогнозирующую функцию, мы будем называть предиктором и обозначать $Z(R_i)$ или Z_i .

Определение 1. Мы будем говорить, что предиктор Z_1 доминирует над предиктором Z_2 , если выполняются два условия:

- а) точность прогнозирования для Z_1 выше точности предиктора Z_2 ;
- б) никакая взвешенная (выпуклая) комбинация предикторов Z_1 и Z_2 , имеющая вид $c_1 Z_1 + c_2 Z_2$ при $c_2, c_1 \geq 0, c_1 + c_2 = 1$, не позволяет обеспечить точность прогноза выше точности прогноза предиктора Z_1 .

Определение 2. Мы будем говорить, что закономерность R_1 доминирует над закономерностью R_2 , если предиктор $Z(R_1)$ доминирует над предиктором $Z(R_2)$. В качестве системы основных (базовых) закономерностей предполагается использовать все закономерности, над которыми не доминируют другие закономерности.

3. Оценка достоверных различий между закономерностями, существующими в двух разных группах

В работах [8, 23] рассматривались методы построения и верификации отдельных закономерностей. Однако нередко возникает вопрос о том, насколько достоверно отличаются закономерности в двух группах объектов. Например, при оценке влияния некоторого бинарного фактора на исследуемый биологический процесс полезно оценить различия между закономерностями, существующими в группах, индуцируемых данным бинарным фактором. Так безусловный интерес представляет поиск достоверных различий между закономерностями в группах с различными вариантами некоторого гена.

Для оценки достоверности различий между закономерностями был предложен новый вариант метода ОДР, использующий модификацию перестановочного теста [24, 25].

Предположим, что нам необходимо сравнить закономерности, существующие в группах объектов \tilde{S}_1 и \tilde{S}_2 . В данной работе рассматриваются одномерные закономерности, принадлежащие к простейшей модели с одной граничной точкой. Построим по группе \tilde{S}_1 оптимальное разбиение интервала значений некоторого признака X_i . Предположим, что нами получено разбиение $R = \{q_1, q_2\}$, которое в свою очередь индуцирует разбиение группы \tilde{S}_1 на подгруппы $\{\tilde{s}_1^1, \tilde{s}_2^1\}$ и группы \tilde{S}_2 на подгруппы $\{\tilde{s}_1^2, \tilde{s}_2^2\}$. Пусть $\hat{y}(\tilde{s}_*^*)$ – среднее значение прогнозируемой величины для подгруппы \tilde{s}_*^* . Для количественной оценки различий между характером взаимосвязи Y с признаком X_i в группах \tilde{S}_1 и \tilde{S}_2 может быть использован функционал сравнения

$$F_{comp}(R, \tilde{S}_1, \tilde{S}_2) = \sum_{i=1}^2 [\hat{y}(\tilde{s}_i^1) - \hat{y}(\tilde{s}_i^2)]^2 \sqrt{m_i^1 m_i^2},$$

где m_i^j – число объектов в группе \tilde{s}_i^j . Для оценки статистической достоверности различий, описываемых функционалом $F_{comp}(R, \tilde{S}_1, \tilde{S}_2)$, используется следующая модификация перестановочного теста.

Пусть $\tilde{S}_1 = \{(y_{f_1^1}, \mathbf{x}_1^1), \dots, (y_{f_{m_1}^1}, \mathbf{x}_{m_1}^1)\}$ и $\tilde{S}_2 = \{(y_{f_1^2}, \mathbf{x}_1^2), \dots, (y_{f_{m_2}^2}, \mathbf{x}_{m_2}^2)\}$. Сгенерируем N пар независимых случайных перестановок наборов натуральных чисел $\{1, \dots, m_1\}$ и $\{1, \dots, m_2\}$ соответственно: $\{\tilde{f}_i^t = \{f_{i1}^t, \dots, f_{im_i}^t\} \mid i \in \{1, 2\}, t \in \{1, \dots, N\}\}$.

С помощью пары перестановок $\{\tilde{f}_i^t \mid i \in \{1, 2\}\}$ из выборок \tilde{S}_1 и \tilde{S}_2 генерируем выборки $\tilde{S}_1^t = \{(y_{f_{i1}^t}, \mathbf{x}_1^1), \dots, (y_{f_{im_1}^t}, \mathbf{x}_{m_1}^1)\}$ и $\tilde{S}_2^t = \{(y_{f_{i1}^t}, \mathbf{x}_1^2), \dots, (y_{f_{im_2}^t}, \mathbf{x}_{m_2}^2)\}$. Данная процедура соответствует случайным перестановкам значений прогнозируемой величины относительно фиксированных позиций векторов переменных X . Вычислим по \tilde{S}_1^t оптимальное разбиение R^t интервала значений признака X_i . Вычислим значение функционала сравнения $F_{comp}(R^t, \tilde{S}_1^t, \tilde{S}_2^t)$. В качестве меры статистической значимости

различий (р-значения) между характером взаимозависимости между Y и X_i в группах \tilde{S}_1 и \tilde{S}_2 предполагается использовать долю перестановок, для которых выполняется неравенство $F_{comp}(R^t, \tilde{S}_1^t, \tilde{S}_2^t) > F_{comp}(R, \tilde{S}_1, \tilde{S}_2)$.

В основе данной процедуры лежит проверка достижимости различий между закономерностями, существующих на исходных данных, при справедливости предположения о полной независимости Y от X_i . Альтернативным способом оценки статистической достоверности различий между характером взаимозависимости как и в случаях статистической достоверности отдельных закономерностей является отношение

$$H_i = \frac{F(R, \tilde{S}_1, \tilde{S}_2)}{\max_{t \in \{1, \dots, N\}} [F(R^t, \tilde{S}_1^t, \tilde{S}_2^t)]},$$

для которого справедливы все положения, приведённые ранее в пункте 2.

ПРИЛОЖЕНИЕ 3.

ОБОЗНАЧЕНИЯ, ИСПОЛЬЗУЕМЫЕ В ТАБЛИЦАХ

Краткое название	Полное название со стандартными границами
АЛТ	Аланинаминотрансфераза, в цифровых показателях
АЛТ – ч	1 – < 40,0 норма, 2 – > 40 ед.
АСТ	Аспартатаминотрансфераза, в цифровых показателях
АСТ – ч	1 – < 40,0 норма, 2 – > 40 ед
Бр	Билирубин, в цифровых показателях
Вазилип	Прием препарата Вазилип (из группы статинов): 1 – нет, 2 – да
ГБ	Гипертоническая болезнь в анамнезе: 1 – нет, 2 – да
Геп_ЕстьГепат.	Наличие гепатоза
Группы по возрасту	Деление групп по возрасту: 1 – 25–59 лет, 2 – 60–74 года, 3 – 75–89 лет, 4 – старше 90 лет
ГЛП	Гиперлипидемия: 1 – нет, 2 – I тип, 3 – IIa тип, 4 – IIb тип
Глюкоза	Глюкоза, в цифровых показателях
Глюкоза – ч	1 – 3,5 – 6,4 ммоль/л – норма, 2 – > 6,4 ммоль/л
Ддиас	Давление диастолическое, в цифровых показателях
Диф. Изм. Щит.	Диффузные изменения щитовидной железы: 1 – нет, 2 – да
Дсис	Давление систолическое, в цифрах
Дс – из.	Давление систолическое: 1 – меньше 140 мм. рт.ст, 2 – больше 140 мм/ рт.ст.
ИИ	Ишемический инсульт в анамнезе: 1 – нет, 2 – да
ИБС – атер	Ишемическая болезнь сердца, атеросклероз: 1 – нет, 2 – да
ИБС – им	ИБС – инфаркт миокарда в анамнезе: 1 – нет, 2 – да
ИБС – ма	ИБС – неклапанная мерцательная аритмия: 1 – нет, 2 – да
ИБС – ст	ИБС – стабильная стенокардия: 1 – нет, 2 – да
Лимфоциты	Лимфоциты, в в цифровых показателях
МРТ – ЛО	Лейкоареоз (по результатам МРТ): 1 – нет, 2 – да
МРТ – ИИ	Ишемический инсульт (по результатам МРТ): 1 – нет, 2 – да
МС	Метаболический синдром: 1 – нет, 2 – да
Насл-ть	Наследственность (по сердечно-сосудистой патологии у родственников I степени родства: 1 – нет, 2 – да.
Рост	В цифровых показателях
Рг аорты	Атеросклероз аорты (по данным рентгенологического исследования): 0 – нет, 1 – есть
Св.х	Свободный холестерин, в цифровых показателях
Св.х – ч	Свободный холестерин: 1 – < 35,0 мг/дл, 2 – > 85,0 мг/дл, 3 – 35,0–85,0 мг/дл
СД 2 типа	Сахарный диабет 2 типа: 1 – нет, 2 – да
СОЭ	Скорость оседания эритроцитов, в цифровых показателях
Степень ожирения	1 – норма (Индекс массы тела 18,5–25,0), 2 – 1 ст. (25–29,9), 3 – 2 ст. (30,0–39,9), 4 – 3 ст. (> 40,0)
ТИА	Транзиторная ишемическая атака (преходящее нарушение мозгового кровообращения): 1 – нет, 2 – да

Триглицер	Триглицериды, в цифровых показателях
Тригл – ч	Триглицериды: 1 – < 0,6, 2 – > 2,3 ммоль/л, 3 – 0,6–2,3 ммоль/л
Узлов. Зоб	Узловой зоб: 1 – нет, 2 – да (gj данным УЗИ)
УЗДГ – изв	Извитость сосудов (по данным УЗДГ)
Узи гепатоз	Гепатоз (по данным УЗИ): 1 – нет изменений, 2 – есть гепатоз, 3 – фиброз печени
УЗИ п/ж	УЗИ поджелудочной железы: 1 – нет изменений, 2 – есть изменения, 3 – фиброз поджелудочной железы, 4 – не проводилось УЗИ
УЗИ – щит.уз.	УЗИ щитовидной железы: 1 – нет изменений, 2 – есть узлы
Фибрин-н	Фибриноген, в цифровых показателях
Фибрин-н – ч	1 – < 4,0 г/л, 2 – > 4,0 г/л
ХИГМ стадии	Хроническая ишемия головного мозга: 1 – I стадия, 2 – II стадия, 3 – III стадия
ХС	Общий холестерин, в цифровых показателях
ХС – ч	Общий холестерин: 1 – < 5,0 ммоль/л, 2 – > 5,0 ммоль/л
ХС – ЛПВП	Лipoproteины высокой плотности, в цифровых показателях
ХС – ЛПВП – ч	Лipoproteины высокой плотности 1 – < 1,0, ммоль/л 2 – > 1,0 ммоль/л
ХС – ЛПОНП	Лipoproteины очень низкой плотности, в цифровых показателях
ХС – ЛПОНП – ч	1 – < 0,63, ммоль/л, 2 – > 0,63 ммоль/л
ЭКГ – глж	Гипертрофия левого желудочка (по данным ЭКГ): 1 – нет, 2 – да
ЭКГ – ИМ	Признаки перенесенного инфаркта миокарда (по данным ЭКГ): 1 – нет, 2 – да
ЭКГ нп	Нарушения проводимости (по данным ЭКГ): 1 – нет, 2 – да
ЭКГ нр	Нарушения ритма (по данным ЭКГ): 1 – нет, 2 – да

СПИСОК ЛИТЕРАТУРЫ

1. Анисимов В.Н. Горячие точки современной геронтологии. *Природа*. 2007. Т. 2. С. 52–61.
2. Гусев Е.И., Скворцова В.И. *Ишемия головного мозга*. М.: Медицина, 2001. С. 327–328.
3. Easton J.D., Saver J.L., Albers G.W., Alberts M.J., Chaturvedi S., Feldmann E., Hatsukami T.S., Higashida R.T., Johnston S.C., Kidwell C.S., Lutsep H.L., Miller E., Sacco R.L. Definition and evaluation of transient ischemic attack: a scientific statement for healthcare professionals from the American Heart Association/ American Stroke Association Stroke Council. *Stroke*. 2009. V. 40. № 6. P. 2277–2278.
4. Семак А.Е., Карнацевич Ю.С., Борисов А.В. *Факторы риска возникновения сосудистых мозговых катастроф*. Минск: Здравоохранение, 2001.
5. Гафуров Б.Г., Аликулова Н.А. Клинико-патогенетические особенности и вопросы лечебной тактики при ишемическом инсульте в пожилом и старческом возрасте. *Инсульт (приложение к Журналу неврологии и психиатрии им. С.С. Корсакова)*. 2004. Т. 11. С. 44–46.
6. Ansell B.J., Navab M., Hama S., Kamranpour N., Fonarow G., Hough G., Rahmani Sh., Mottahedeh R., Dave R., Reddy S.T., Fogelman A.M. *Circulation*. 2003. V. 108. P. 2751–2756.
7. Кузнецова А.В., Сенько О.В. Возможности использования методов Data Mining примедико-лабораторных исследованиях для выявления закономерностей в массивах данных. *Врач и информационные технологии*. 2005. Т. 1. С. 49–56.
8. Senko O.V., Kuznetsova A.V. *The Optimal Valid Partitioning Procedures*, «InterStat», *Statistics in Internet*. URL: <http://interstat.statjournals.net/YEAR/2006/articles/0604002.pdf> (дата обращения: 18.02.2013).
9. Кадыков А.С., Манвелов Л.С., Шахпаронова Н.В. *Хронические сосудистые заболевания головного мозга (дисциркуляторная энцефалопатия)*. СПб: ГЭОТАР-Медиа, 2006. 236 с.
10. *Справочник по формулированию клинического диагноза болезней нервной системы*. Под ред. В.Н. Штока, О.С. Левина. М: ООО "Медицинское информационное агентство", 2006. 520 с.
11. Журавлёв Ю.И., Рязанов В.В., Сенько О.В. «Распознавание». *Математические методы. Программная система. Применение*. Москва: Фазис, 2006.
12. Рязанов В. В. Логические закономерности в задачах распознавания (параметрический подход). *Ж. вычисл. матем. и матем. физ.*, 2007, Т. 47, №10, С. 1793–1808.
13. Senko O., Kuznetsova A. A recognition method based on collective decision making using systems of regularities of various types. *Pattern Recognition and Image Analysis*. 2010. V. 20. № 2. P. 152–162.
14. Rizzo M., Corrado E., Coppola G., Muratori I., Novo G., Novo S. Prediction of cardio- and cerebro-vascular events in patients with subclinical carotid atherosclerosis and low HDL-cholesterol. *Atherosclerosis*. 2008. V. 200. № 2. P. 389–395.
15. Novo S., Carità P., Corrado E., D'Ambrosi A., Novo G. Low HDL-cholesterol concentration cause atherosclerotic disease to develop. *E-Journal of the ESC Council for Cardiology Practice*. 2009. V. 7. № 32.
16. Amareco P., Golstein L.B., Gallahan A. 3 rd, Sillesen H., Hennericic M.G., O'Neill B.J., Rudolphg A.E., Simunovicg L., Zivinh J.A., Welchi K.M.A. Baseline blood pressure, low- and high-density lipoproteins, and triglycerides and the risk of vascular events in the Stroke Prevention by Aggressive Reduction in Cholesterol Levels (SPARCL) trial. *Atherosclerosis*. 2009. V. 204. № 2. P. 515–520.

17. Simons L.A., Simons J., Fridlander Y., McCallum J. Cholesterol and other lipids predict coronary heart disease and ischaemic stroke in the elderly, but only in those below 70 years. *Atherosclerosis*. 2001. V. 159. № 1. P. 201–208.
18. Gordon T., Kannel W.B., Castelli W.P., Dawber T.R. Lipoproteins, cardiovascular disease and death: the Framingham study. *Arch. Intern. Med.* 1981. V. 141:P. 1128–1131.
19. Palmer A., Bulpitt C., Beevers G., Coles E., Fletcher A., Ledingham J., Petrie J., Webster J., Dollery C. Risk factors for ischaemic heart disease and stroke mortality in young and old hypertensive patients. *J. Hum. Hypertens.* 1995. V. 9. P. 695–697.
20. Nawawi H., Day L., Mikhailidis D., Hamilton G., Winder A.F. Secondary prevention and referral of elderly patients to a cardiovascular lipid service. *Arch. Gerontol. Geriatr.* 1999. V. 29. P. 1–11.
21. Demchuk A.M., Hess D.C., Brass L.M., Yatsu F.M. Is cholesterol a risk factor for stroke? Yes. *Arch. Neurol.* 1999. V. 56. P. 1518–1520.
22. Wannamethee S.G., Shaper G., Ebrahim S. HDL-Cholesterol, total Cholesterol and the risk of stroke in middle-aged British men. *Stroke*. 2000. V. 31. P. 1882–1892.
23. Кузнецова В., Костомарова И.В., Водолагина Н.Н., Малыгина Н.А., Сенько О.В. Изучение влияния клиничко-генетических факторов на течение дисциркуляторной энцефалопатии с использованием методов распознавания. *Матем. биолог. и биоинформ.* 2011. Т. 6. № 1. С. 115–146. URL: [http://www.matbio.org/2011/Senko2011\(6_115\).pdf](http://www.matbio.org/2011/Senko2011(6_115).pdf) (дата обращения: 25.03.2013).
24. Senko O.V., Kuznetsova A.V. Methods for discovering and analysis of regularities systems. Approach based on optimal partitioning of explanatory variables space. В: *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR 2011)*. 2011. P. 423–426.
25. Senko O.V., Kuznetsova A.V., Malygina N.A., Kostomarova I.V. Method for Evaluating of Discrepancy between Regularities Systems in Different Groups. *Information Technologies & Knowledge*. 2011. V. 5. № 1. P. 46–54.

Материал поступил в редакцию 05.03.2012, опубликован 06.05.2013.